

Workgroup: Routing Area Working Group
Internet-Draft: draft-retana-rtgwg-eacp-05
Published: 22 February 2023
Intended Status: Informational
Expires: 26 August 2023
Authors: A. Retana
Futurewei Technologies, Inc.
M. Paul
Deutsche Telekom AG
R. White
Akamai Technologies

A Framework for Energy Aware Control Planes

Abstract

Energy is a primary constraint in large-scale network design, particularly in cloud-scale data center fabrics. While compute and storage clearly consume the largest amounts of energy in large-scale networks, the optics and electronics used in transporting data also contribute to energy usage and heat generation.

This document provides an overview of various areas of concern in the interaction between network performance and efforts at energy aware control planes, as a guide for those working on modifying current control planes or designing new ones to improve the energy efficiency of high density, highly complex, network deployments.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on 26 August 2023.

Copyright Notice

Copyright (c) 2023 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents

(<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

- [1. Introduction](#)
- [2. Background](#)
 - [2.1. Scope](#)
 - [2.2. Business Drivers](#)
 - [2.3. Application Drivers](#)
- [3. Framework](#)
 - [3.1. Example Network](#)
 - [3.2. Modes of Reducing Energy Usage](#)
 - [3.3. Global Versus Local Decisions](#)
- [4. Considerations](#)
 - [4.1. Energy Efficiency and Bandwidth Reduction](#)
 - [4.2. Energy Efficiency and Stretch](#)
 - [4.3. Energy Efficiency and Fast Recovery](#)
 - [4.4. Energy Efficiency and Microsleeps](#)
 - [4.5. Other Operational Aspects](#)
- [5. Security Considerations](#)
- [6. IANA Considerations](#)
- [7. Acknowledgements](#)
- [8. References](#)
 - [8.1. Informative References](#)
- [Authors' Addresses](#)

1. Introduction

The availability of low-cost energy sources, provisioning energy sources, and handling the heat generation from processing and transporting data are determining factors in the siting, development, and operation of large-scale data centers. The rise of edge computing, 5G, and diversified compute is causing the importance of understanding and reducing energy usage in networks to become increasingly important.

As with all network and protocol design, however, reducing energy use represents a tradeoff. In the case of networks, increasing energy efficiency can result in a loss of optimization in network operations in other areas. These kinds of tradeoffs can be described in terms of the state/optimization/surface triad; increasing local optimization in one area, energy consumption, can result in global

sub-optimization through increased state, more complex interaction surfaces, or even suboptimal global energy usage.

This document provides background information and a framework for understanding the tradeoffs between modifications made to network control plane protocols to conserve energy and network performance. This document also makes suggestions to designers and implementers of modifications intended to enable energy conservation in networks. The intent of this document is to encourage work in the area of reducing network energy usage through protocol design, network design, and network operations.

The document is organized as follows. [Section 2](#) provides material the reader needs to understand to appreciate the challenges inherent in balancing energy reduction with effective network performance. This section includes subsections considering the application and business requirements that are the basis of the rest of the document. [Section 3](#) provides a framework for understanding common mechanisms in energy management schemes. [Section 4](#) provides an analysis of the areas highlighted, including an explanation of how the specific area interacts with energy management, an example of the interaction, and, finally, a set of considerations for protocol designers when proposing either new protocols or modifications to existing protocols to reduce energy usage.

2. Background

This section describes the underlying business and application drivers for the considerations sections.

2.1. Scope

Radio based networks designed for rapid deployment for highly mobile users (often called Mobile Ad Hoc Networks [[RFC2501](#)]), and sensor networks designed using devices with limited power, memory, and processing resources [[RFC7102](#)], are not the target of this document. Readers should refer to the groups working within those areas for energy management requirements based on those specialized environments. While protocol developers for those environments may draw useful information from this document, this work is not intended to address those specialized networks specifically. Mobile cellular networks however are similarly affected by excess energy consumption as wireline networks and seek to save energy by methods such as the ones described in [[SDO-3GPP.25.927](#)].

Inter-domain applications require more work in policy than in technical and business considerations, and therefore fall outside the scope of this document. Intra-domain control planes are (intuitively) where most energy savings will be attained, at any

rate. Most high concentrations of routers, such as data centers and campus networks, are under a single administrative domain. Therefore, placing inter-domain control planes outside the scope of this document does not limit its usefulness in any meaningful way.

Energy monitoring deals with the collection of information related to energy utilization and its characteristics, and energy control relates to directly influencing the optimization and/or efficiency of devices in the network [[RFC7326](#)]. The focus of this document is on understanding the tradeoffs between modifications made to network protocols to conserve energy and network performance metrics, rather than functions, steps or procedures required for energy monitoring or control.

2.2. Business Drivers

Networks are driven by organizational, application specific, and general connectivity requirements. Organizational requirements include capital and operational expense, and the restrictions the network architecture places on the growth and operation of the organization. The interaction between the network and organization is managed through change management, availability, and network agility (the ability to quickly support or shed application demands on the network).

Modifying control planes to support energy awareness impacts capital and operational expenses primarily through tradeoffs against availability, and potentially through network agility.

2.3. Application Drivers

Applications drivers provide the background for each of the technical sections below. While network operators and protocol designers need to pay attention to a wide array of factors when considering how best to support specific applications, this document focuses on factors with broad impact. The first two questions involve bandwidth: how much bandwidth will the application consume, and is this bandwidth consumption fairly steady, or highly variable? For instance, applications such as streaming video tend to have long lasting flows with high bandwidth requirements, file transfers tend to produce shorter flows requiring high bandwidth, and HTML traffic tends to be bursty, with much lower bandwidth requirements.

The next question a protocol or network designer might ask about a specific application is its tolerance to jitter. Real time applications, such as voice and video conferencing, have a very low toleration for jitter. File transfers and streaming video, on the other hand, can often handle large variations in packet arrival times. If packets are delayed long enough, the application may

actually time out, shutting down sessions. Users will often "hang up" after a short period of time, as well, causing loss of revenue and productivity.

Delay is another crucial factor in the performance of many applications. Many server virtualization protocols, for instance, have very low tolerance for delay, having been written with connectivity through a short wire local broadcast segment in mind. Applications such as stock and commodity trading, remote medical, and collaborative video editing also exhibit very little tolerance for delay. Applications built on a microservices model will often exhibit deep performance loss when running over a network with high or variable delay.

Variable delay, or jitter, is another factor which broadly impacts application performance. Networks with high jitter require longer flow and error control timeouts, reducing application performance.

Control plane convergence events can cause jitter in traffic flows through the network. Changing the number of hops through the network (an increase or reduction in stretch) will cause the delay through the network to vary, which is perceived by the application as jitter. The selection of a suboptimal route by the control plane, for instance through the introduction of aggregation or summarization of reachability information, or the selection of a more heavily loaded link over a more likely loaded one, can increase the number of hops traffic takes through the network, or select a path with more deeply filled queues. Either of these can cause traffic to pass through the network more slowly, which is perceived by the application as delay.

Jitter and delay can also be introduced directly into the packet stream by reducing the throughput of individual links, or putting devices and/or links into energy reduced modes for very short periods of time (microsleeps). If a link is asleep when the first and third packets from a flow arrive at the head end of the link, and not when the second packet from that same flow arrives, each packet is going to be processed differently, and hence will have a different delay across the path.

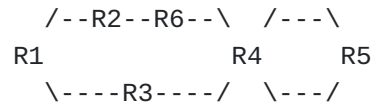
The following sections address bandwidth reduction, increasing stretch, network convergence, and introducing jitter through microsleeps, in more detail.

3. Framework

This section contains a sample network which is used throughout the rest of this document, considers some ways in which energy usage can be reduced, and provides some examples.

3.1. Example Network

To illustrate the impacts of link and device removal throughout the rest of this document, the following network is used.



This network is overly simplistic so the impact of removing various links and devices from the topology can be more clearly illustrated. More complex topologies will often exhibit these same effects in less obvious, and harder to understand, ways.

3.2. Modes of Reducing Energy Usage

There are four primary ways in which energy usage can be reduced:

- *Removing redundant links from the network topology; for instance, one of the two parallel links between R4 and R5 may be removed
- *Removing redundant network equipment from the network topology; for instance, R2 and R6, along with their associated links, may be removed
- *Reducing the amount of time equipment or links are operational; for instance, one of the two parallel links between R4 and R5 may be shut down during time periods when the traffic flow through the network is not large enough to justify operating both links
- *Reducing the link speed or processing rate of equipment; for instance, the speed of the two links between R4 and R5 may be reduced during time periods when the traffic through the network is not large enough to justify making some higher amount of bandwidth available

Completely removing nodes or links from the network topology has several impacts on the control plane which must be considered. In these cases, the control plane must:

- *Modify the network topology so removed links or devices are not used to forward traffic
- *Remember that such links exist, possibly including the neighbors and destinations reachable through them

3.3. Global Versus Local Decisions

It is often tempting to optimize for local conditions while ignoring system-level results, or to optimize system-level conditions while ignoring local results. Both of these, however, are often a mistake. The former extreme might be illustrated in a system where two nodes measure local link utilization, shutting down any interconnecting links when the utilization percentage drops below a certain level. In such a system, pairs of adjacent devices may decide to shut down a set of links which leaves no available path (or insufficient bandwidth) through the network as a system. An example of the latter might be a system where every node in the network must agree to shut down a link before the link can be disabled for energy conservation. In this case, a false perception of overall system health caused by timing issues could cause a lack of local optimizations to take place.

There are some considerations and tradeoffs which need to be outlined in considering the global versus local decisions in relation to energy efficiency. System designers should take note of the difficulties with preventing pathological conditions when purely localized decisions are made. For instance, in the example network, assume R1 determines to put the R1->R2 link into an energy saving mode, while R4 determines to put the R4->R3 link into an energy saving mode. In this case, no path will remain available through the network. It is also possible for the opposite to occur, that is for no links or devices to be placed into a reduced energy state because R1 and R4 don't agree through the control plane which links and devices should be removed from the topology.

Protocol designers should consider these tradeoffs in proposals for energy aware control planes.

4. Considerations

Each subsection considers a single energy saving mechanism in detail.

4.1. Energy Efficiency and Bandwidth Reduction

Bandwidth is an important consideration in high density networks; data center fabrics, for instance, are designed to provide a specific amount of bandwidth, often with relatively fixed delay and jitter, into and out of each server and to facilitate virtual server movement among physical devices. In campus and core networks bandwidth is finely coupled with quality of service guarantees for applications and services. It should be obvious that removing links or devices from a network topology will adversely affect the amount

of available bandwidth, which could, in turn, cause well thought out quality of service mechanisms to degrade or fail.

What might not be so obvious is the relationship between available bandwidth and jitter (or other network quality of service measures). If higher speed links are removed from the topology in order to continue using lower speed (and therefore presumably lower power) links, then serialization delays will have a larger impact on traffic flow. Longer serialization delays can cause input queues to back up, which impacts not only delay but jitter, and possibly even traffic delivery.

4.1.1. An Example of Lowering Bandwidth by Removing Parallel Links

In the network illustrated above, one of the two links between R4 and R5 could be an obvious candidate for removal from the network, especially if the network load can easily be transferred to the remaining link without failure. There can be multiple negative impacts from the perspective of optimal traffic flow, among which could be the mixing of different kinds of flows with multiple quality of service requirements.

If, for instance, a flow carrying voice data is mixed with a large file transfer, the mixed queueing of traffic with two different classes of service could cause variable delay (jitter), reducing application performance.

4.1.2. Considerations

Modifications to control plane protocols to achieve network energy efficiency should provide the ability to set the minimal bandwidth, jitter, and delay through the network, and not shut down links or devices that would violate those minimal requirements.

4.2. Energy Efficiency and Stretch

In any given network, there is a shortest path between any source and any destination. Network protocols discover these paths from the destination's perspective --routing draws traffic along a path, rather than driving along a path. Along with the shortest path, there are a number of paths that can also carry traffic from a given source to a given destination without the packets passing along the same logical link, or through the same logical device, more than once. These are considered [loop-free alternate](#) [[RFC5714](#)] paths.

The primary difference between the shortest path and the loop-free alternate paths is the total cost of using the path. In simple terms, this difference can be calculated as the number of links and devices a packet must pass through when being carried from the source to the destination, or the hop count. While most networks use

much more sophisticated metrics based on bandwidth, congestion, and other factors, the hop count of the path a flow takes through the network is a convenient measure of path efficiency.

When the control plane causes traffic to pass from the source to the destination along a path which is longer than the shortest path, the network is said to have stretch (see [[Krioukov](#)] for a more in depth explanation of network stretch). To measure stretch, simply subtract the metric of the shortest path from the metric of the longer path. For example, in hop count terms, if the best path is three hops, and the current path is four hops, the network exhibits a stretch of 1.

4.2.1. An Example of Stretch

In the network illustrated above, if a modification is made to the control plane to remove the link between R1 and R3 in order to save energy, all the destinations shown in the diagram remain reachable. However, from the perspective of R1, the best path available to reach R2 has increased in length by one hop. The original path is R1->R3->R4->R5, the new path is R1->R2->R6->R4->R5. This represents a stretch of 1.

Along with this increased stretch will most likely also come increased delay through the network; each hop in the network represents a measurable amount of delay. This increased stretch might also represent an increased amount of jitter, as there are more queues and more serialization events in the path of each packet carried. There will also be the modifications in jitter as the network switches between the optimal performance configuration and an energy efficient configuration.

4.2.2. Considerations

Designers who propose modifications to control plane protocols to achieve network energy efficiency should analyze the impact of their mechanisms on the stretch in typical network topologies, and should include such analysis when explaining the applicability of their proposals. This analysis may include an examination of the absolute, or maximum, stretch caused by the modifications to the control plane as well as analysis at the 95th percentile, the average stretch increase in a given set of topologies, and/or the mean increase in stretch.

Mechanisms that could impact the stretch of a network should provide the ability for the network administrator to limit the amount of stretch the network will encounter when moving into a more energy efficient mode.

4.3. Energy Efficiency and Fast Recovery

A final area where modifications to the control plane for energy efficiency is fast convergence or fast recovery. Many networks are now designed to recover from failures quickly enough to only result in a handful of traffic lost; recovery on the order of half a second is not an uncommon goal. It should be obvious that removing redundant links and devices from the network to reduce energy consumption could adversely affect these goals.

4.3.1. An Example of Impact on Fast Recovery

In the network shown, assume R2 and its associated links are shut down in order to save energy. The result is a one-connected network with no redundant link, impacting the resilience of the network to node and link failures. It is possible to craft a mechanism that will bring devices and links which have been powered down or taken into a low-energy mode back into service, but these will necessarily require some startup time, which will impact the Mean Time to Repair (MTTR) enabled through the control plane. This impact will appear to applications running over the network as extended jitter, and potentially the loss of packets.

For these reasons, it may be that only links and devices which are a "third point of failure" may be acceptable as removal candidates in order to conserve energy.

4.3.2. Considerations

Modifications to the control plane in order to remove links or nodes to conserve energy should entail the ability to choose the level of redundancy available after the network topology has been trimmed. For instance, it might be acceptable in some situations to move to single points of failure throughout the network, or in specific sections of the network, for certain periods of time. In other situations, it may only be acceptable to reduce the network to a double point of failure, and never to a single point of failure.

4.4. Energy Efficiency and Microsleeps

Another mechanism to reduce energy usage in a network is to sleep links or devices for very short periods of time, called microsleeps. For instance, if a particular link is only used at 50% of the actual available bandwidth, it should be possible to place the link in some lower power state for 50% of the time, thus reducing energy usage by some percentage. An example of one such mechanism is Energy-Efficient Ethernet [[IEEE 802.3az 2010](#)].

Such schemes can introduce delay and jitter into the network path directly; if a packet arrives while the link is in a reduced energy

state, it must wait until the link enters a normal operational mode before it can be forwarded. Most of the time the proposed sleep states are so small as to be presumably inconsequential on overall packet delay, but multiple packets crossing a series of links, each encountering different links in different states, could take very different amounts of time to pass along the path.

One possible way to resolve this somewhat random accrual of delays on a per packet basis is to coordinate these sleep states such that packets accepted at the entry of the network are consistently passed through the network when all links and devices are in a normal operating mode, and simply delaying all packets at the entry point into the network while the devices in the network are in an energy reduced state. This solution still introduces some amount of jitter; some packets will be delayed by the sleep state at the edge of the network, while others will not. This solution also requires coordinated timers at the speed of forwarding itself to effectively control the sleep and wake cycles of the network.

4.4.1. An Example of Microsleeps to Reduce Energy Usage

In the example network, assume the bandwidth utilization along the path R1->R3->R4->R5 is 50% of the actual available bandwidth. It is possible to consider a scheme where the R1->R3, R3->R4, and R4->R5 links are all put into an energy reduced operational mode 50% of the time, since packets are only available to send 50% of the time. A packet entering at R1 may encounter a short delay at the R1->R3, R3->R4, and R4->R5 links, or it might not. Even if these delays are small, say 200ms at each hop, the accumulated delay through the network due to sleep states may be 0ms (all links and devices awake) or 600ms (all links and devices asleep) as the packet passes through the network.

As network paths lengthen to more realistic path lengths in real deployments, the jitter introduced varies more widely, which could cause problems for the operation of a number of applications.

4.4.2. Considerations

Protocol designers should analyze the impact of accumulated jitter when proposing mechanisms that rely on microsleeps in either equipment or links. This analysis should include both worst case and best case scenarios, as well as an analysis of how coordinated clocks are to be handled in the case of coordinated sleep states.

4.5. Other Operational Aspects

Modification of the network topology in order to save energy needs to consider the operational needs of the network as well as application requirements. Change management, operational downtime,

and business usage of the network need to be considered when determining which links and nodes should be placed into a low energy state. Energy provisions have to be assigned and changed for nodes and links, optimally according to network usage profiles over the time of day.

Control plane protocol operation, in terms of operational efficiency on the wire, also needs to be considered when modifying protocol parameters. Any changes that negatively impact the operation of the protocol, in terms of the amount of traffic, the size of routing information transmitted over the network, and interaction with network management operations need to be carefully analyzed for scaling and operational implications.

4.5.1. An Example of Operational Impact

Time of day is an important consideration in business operations. During normal operational hours, the network needs to be fully available, including all available redundancy and bandwidth. During holidays, night hours, and other times when a campus might not be used, or when there are lower traffic and resiliency demands on the network, network elements can be removed to reduce energy usage.

4.5.2. Considerations

Protocol designers should analyze operational requirements, such as time of day and network traffic load considerations, and explain how proposed protocols or modifications to protocols will interact with them. Protocol designers should analyze increases in network traffic and the operational efficiency impact of proposed changes or protocols.

5. Security Considerations

This document provides an overview of various areas of concern in the interaction between network performance and the use of energy efficient control planes to improve the energy efficiency of a network. As such, it doesn't introduce any new security risk.

However, providing an API or other mechanism to dynamically modify available bandwidth, put devices in reduced energy states, and otherwise modify network behavior introduces surfaces along which attackers can use to deny effective service to critical applications. By reducing the amount of available bandwidth along a link by invoking energy saving mechanisms, for instance, an attacker could reduce the performance of an application, harming the interests of organizations relying on the network.

Protocol designers should carefully consider the introduction of any potential vulnerabilities as a result of the implementation of an energy aware control plane.

6. IANA Considerations

This document has no IANA actions.

7. Acknowledgements

The authors of this document would like to acknowledge the suggestions and ideas provided by Sujata Banerjee, Puneet Sharma, Dirk Von Hugo, and John Scudder.

8. References

8.1. Informative References

[IEEE_802.3az_2010] IEEE, "IEEE Standard for Information technology-- Local and metropolitan area networks-- Specific requirements-- Part 3: CSMA/CD Access Method and Physical Layer Specifications Amendment 5: Media Access Control Parameters, Physical Layers, and Management Parameters for Energy-Efficient Ethernet", IEEE 802-3az-2010, DOI 10.1109/IEEESTD.2010.5621025, 11 July 2013, <<https://ieeexplore.ieee.org/document/5621025>>.

[Krioukov] Krioukov, D., "On Compact Routing for the Internet", 2007, <http://www.caida.org/publications/papers/2007/compact_routing/>.

[RFC2501] Corson, S. and J. Macker, "Mobile Ad hoc Networking (MANET): Routing Protocol Performance Issues and Evaluation Considerations", RFC 2501, DOI 10.17487/RFC2501, January 1999, <<https://www.rfc-editor.org/info/rfc2501>>.

[RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, DOI 10.17487/RFC5714, January 2010, <<https://www.rfc-editor.org/info/rfc5714>>.

[RFC7102] Vasseur, JP., "Terms Used in Routing for Low-Power and Lossy Networks", RFC 7102, DOI 10.17487/RFC7102, January 2014, <<https://www.rfc-editor.org/info/rfc7102>>.

[RFC7326] Parello, J., Claise, B., Schoening, B., and J. Quittek, "Energy Management Framework", RFC 7326, DOI 10.17487/RFC7326, September 2014, <<https://www.rfc-editor.org/info/rfc7326>>.

[SDO-3GPP.25.927]

3GPP and B. Prakash, "Solutions for energy saving within UTRA Node B", 3GPP TR 25.927 14.0.0, 25 March 2017, <http://www.3gpp.org/ftp/Specs/archive/25_series/25.927/25927-e00.zip>.

Authors' Addresses

Alvaro Retana
Futurewei Technologies, Inc.

Email: aretana@futurewei.com

Russ White
Akamai Technologies

Email: ruessw@riw.us

Manuel Paul
Deutsche Telekom AG

Email: Manuel.Paul@telekom.de