Routing Area Working Group
Internet-Draft
Intended status: Informational
Expires: April 20, 2014

A. Retana
Cisco Systems, Inc.
R. White

M. Paul
Deutsche Telekom AG
October 17, 2013

A Framework and Requirements for Energy Aware Control Planes
draft-retana-rtgwg-eacp-02

Abstract

There has been, for several years, a rising concern over the energy usage of large scale networks. This concern is strongly focused on campus, data center, and other highly concentrated deployments of network infrastructure. Given the steadily increasing demand for higher network speeds, always-on service models, and ubiquitous network coverage, it is also of growing importance for telecommunication networks both local and wide area in scope. One of the issues in moving forward to reduce energy usage is to ensure that the network can still meet the performance specifications required to support the applications running over it.

This document provides an overview of the 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 control planes to improve the energy efficiency of high density, highly complex, network deployments.

Table of Contents

1. Introduction

As energy prices continue to increase, and energy awareness becomes a watchword for most large companies, places where the network infrastructure used a good deal of power have come under increased scrutiny for savings. There is a concern, however, in saving energy at the cost of network operations --to reduce performance along with energy consumption, negatively impacting the operation of a network and the applications reliant on that network. This concern is primarily focused on the network control plane, but will necessarily apply to network performance and energy usage overall.

This document provides a background, a framework for understanding and managing the tradeoffs between modifications made to network protocols to conserve energy and network performance metrics and requirements, and a set of requirements for protocol designers to consider in proposals for new control plane protocols or modifications to existing control plane protocols. It is intended to encourage work on mechanisms that will reduce network energy usage while providing perspective on balancing energy usage against performance. The ultimate goal is to provide the tools and knowledge necessary for protocol designers to modify network protocols to best balance efficiency against performance, and to provide the background information network operators will need to intelligently deploy and use protocol modifications to network protocols.

The document is organized as follows. Section 3 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 reset of the document. Section 4 provides a framework for understanding mechanisms common to all energy management schemes proposed to date in general terms. Section 5 provides an analysis of the areas highlighted, including an explanation of how the specific area interacts with energy management, and example of the interaction, and, finally, a set of requirements protocol designers should consider when proposing either new protocols or modifications to existing protocols to reduce energy usage.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

3. Background

The background covered here describes the underlying business and application drivers for the consideration and requirements sections below. This section also contains a small example network used throughout the remainder of this document for explaining various mechanisms and technical points.

3.1. Scope

The reader should differentiate between radio based and wireline (or rather, "plugged in"), networks. Radio based networks designed for rapid deployment for highly mobile users (often called Mobile Ad Hoc Networks, or MANETs [MANET]), and sensor networks designed for low power, processing, and memory (such as those described in [ROLL]), 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 environment. 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 as described in the following (see e.g. [3GPP]).

The reader should also differentiate between intradomain and interdomain applications. Interdomain applications require more work in policy than in technical and business considerations, and therefore fall outside the scope of this document. Intradomain 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 interdomain control planes outside the scope of this document does not limit its usefulness in any meaningful way.

The reader should further differentiate between the components of an energy management system, namely energy monitoring and energy control. Energy monitoring deals with the collection of information related to energy utilization and characteristics, as described in [EMAN]. Energy control relates to directly influencing the optimization and/or efficiency of devices in the network. The focus of this document is on understanding the tradeoffs between modifications made to network protocols to conserve energy and network performance metrics and requirements, not on the functions, steps or procedures required for energy monitoring.

3.2. Business Drivers

Networks are primarily built to support both broad and narrow business requirements. Broad business requirements might include general communication requirements, such as providing email service between internal and external personnel, or providing general access to the World Wide Web for research and business support. Narrow requirements would relate to specific applications, such as supporting a particular financial application in the case of a bank or other financial enterprise, or supporting customer traffic in the case of a service provider. Application requirements will be considered in greater detail in the next section.

Another class of requirements business place on networks can be called operational requirements. These include (but are not limited to), capital expense, operational expense, and the restrictions the network architecture places on the growth and operation of the business itself. These, in turn, drive requirements such as change management, total uptime (availability), and the ability of the network to be easily and quickly modified to meet new business demands, or to shed old business demands. Operational expense is the primary area this document covers in relation to business requirements, because this is where energy management most obviously overlaps with network performance.

3.3. Application Drivers

Applications drivers provide the background for each of the technical sections below. When approaching a specific application, there are only a small number of questions network and protocol designers need to fully understand to shape networks and protocols so a specific application can be supported. The first two questions revolve around 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 it's 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 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.

These last two application drivers, jitter and delay, are normally the result of two underlying causes within a network's control plane: stretch and convergence. Stretch (defined more fully in the section considering stretch below) causes longer paths to be taken through the network. Each hop in the network path adds serialization into and out of a set of queues in device memory, along with the delays of various queuing mechanisms implemented on that device. Each hop in the network increases delay directly, and has the potential to increase jitter as packets pass into and out of the additional devices.

Network convergence will also show up as jitter in an application's stream; if packets are held up or looped for hundreds of milliseconds during a network convergence event, applications running over the converging topology will see this convergence time as a massive jitter event, or a short term delay in the delivery of packets.

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 specific technical problems addressed in the following sections, then, are bandwidth reduction, increasing stretch, network convergence, and introducing jitter through microsleeps.

4. Framework

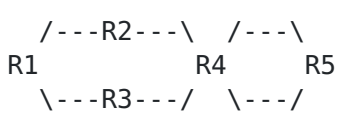
4.1. Modes of Reducing Energy Usage

Regardless of whether the control plane is centralized (such as some form of centrally computed traffic engineering or software defined network), or distributed (traditional routing protocols), there are four primary ways in which energy usage can be reduced:

- Removing redundant links from the network topology
- Removing redundant network equipment from the network topology
- Reducing the amount of time equipment or links are operational
- Reducing the link speed or processing rate of equipment

4.1.1. 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 impacts without being so obvious.

4.1.2. Examples of Energy Reduction

In the example network above, several different modes of energy reduction might be:

- Shutting down one of the two links between R4 and R5
- Shutting down one of the two links between R4 and R5, and shutting down any line cards (or part of the nodes themselves) associated with the removal of these links
- Shutting down R2 or R3, since these represent alternate paths to reach the same set of destinations
- Shutting down the link between R2 and R4, since similar connectivity is provided through R1->R3->R4
- Shutting down all links and devices for fractions of time in a coordinated fashion
- Shutting down individual links as traffic or the control plane permits for fractions of time (here the momentary shutdown of various links is not coordinated, but undertaken hop by hop)
- Reducing the speed of all links and devices for fractions of time in a coordinated fashion

- Reducing the speed of individual links as traffic or the control plane permits for fractions of time (here the momentary slowdown of various links is not coordinated, but undertaken hop by hop)

4.2. Global Verses Local Decisions

Independent of whether the control plane is centralized or distributed, the scope considered when making a decision about energy efficiency may affect the result and effectiveness of the system. There are clearly two extreme options when looking at the scope of the information used to make decisions. The first extreme is that of every device in the network considering only local conditions, and determining the proper local state from that information. An example of this mode of operation might be a local link where the devices on either side of that link measure the link utilization, and independently decide to automatically shut the link down when utilization reaches a specific threshold. An example of the other end of the spectrum might be a network control plane in which all the nodes involved agree before taking a specific action; in the case of two parallel links, the devices on each end not only would have similar configured policies, but would coordinate if one of the links was to be turned off. It is outside the scope of this document to determine which of these two options may be optimal or "best."

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.

5. Considerations and Requirements

5.1. Energy Efficiency and Bandwidth Reduction

Bandwidth is an important consideration in high density networks; most data centers are designed to provide a specific amount of bandwidth 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.

5.1.1. An Example of Lowered Bandwidth

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, and without serious consequences for delay or jitter in the network, there is a strong case to be made for doing so --particularly if the accompanying line cards could also be shut down to add to the energy savings.

5.1.2. Requirements

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.

5.2. 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 --the hop count. While most networks use much more sophisticated metrics based on bandwidth, congestion, and other factors, the hop count will stand in as the only metric used throughout this document.

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.

5.2.1. An Example of Stretch

In the network illustrated above, if a modification is made to the control plane in order to remove the link between R1 and R4 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 two hops. The original path is R1->R2, the new path is R1->R3->R4->R2. This represents a stretch of 2.

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.

5.2.2. Requirements

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.

5.3. 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 cause a handful of traffic to be 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.

5.3.1. An Example of Impact on Fast Recovery

In the network shown, assume R2 and its associated links are removed from the topology in order to save energy. Rather than this second path being available for immediate recovery on the failure of the R1->R3 link, some process must be followed to bring R2 and its associated links back up, reinject them into the topology, and finally begin routing traffic across this path.

In many situations, only links and devices which are a "third point of failure" may be acceptable as removal candidates in order to conserve energy.

5.3.2. Requirements

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.

The complete removal of 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 those links or devices

5.4. Introducing Jitter Through Microsleeps

One proposed 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 something percentage.

Such schemes introduce delay and jitter into the network path directly; if a packet arrives while the link to the next hop, or the next hop itself, is in a reduced energy state, the packet must wait until the link or next hop device enter 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 some 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.

5.4.1. An Example of Microsleeps to Reduce Energy Usage

In the example network, assume the bandwidth utilization along the path R1->R2->R4->R5 is 50% of the actual available bandwidth along this path. It is possible to consider a scheme where R1->R2, R2, R2->R4, and R4->R5 are all put into some 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 R1->R2, at R2->R4, and at R4->R5, or it might not. Even if these delays are very 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.

5.4.2. Requirements

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.

5.5. 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.

5.5.1. 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.

5.5.2. Requirements

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 these types of requirements. Protocols designers SHOULD analyze increases in network traffic and the operational efficiency impact of proposed changes or protocols.

6. Security Considerations

None.

7. Acknowledgements

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

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

8.2. Informative References

[3GPP] 3GPP, "3GPP TR 25-927 Solutions for energy saving within UTRA Node B", 2011, <http://3gpp.org/ftp/Specs/html-info/25927.htm>.

[EMAN] IETF, "Energy Management Working Group Charter", 2012, <http://datatracker.ietf.org/wg/eman/charter/>.

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

[MANET] IETF, "Mobile Ad Hoc Networks Charter", 2012, <http://datatracker.ietf.org/wg/manet/charter/>.

[RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework", RFC 5714, January 2010.

[ROLL] IETF, "Routing Over Lowe power and Lossy networks Charter", 2012, <http://datatracker.ietf.org/wg/roll/charter/>.

Appendix A. Change Log

A.1. Changes between the -00 and -01 versions.

- Updated authors' contact information.
- Modified some of the rfc2119 keywords.

A.2. Changes between the -01 and -02 versions.

- Updated authors' contact information.

Authors' Addresses

Alvaro Retana
Cisco Systems, Inc.
7025 Kit Creek Rd.
Raleigh, NC 27709
USA

Email: aretana@cisco.com

Russ White

Email: russw@riw.us

Manuel Paul
Deutsche Telekom AG
Winterfeldtstr. 21-27
Berlin 10781
Germany

Email: Manuel.Paul@telekom.de