Workgroup: Network Working Group
Internet-Draft: draft-cx-green-ps-01
Published: 20 October 2022
Intended Status: Informational
Expires: 23 April 2023
Authors: A. Clemm (Futurewei), C. Westphal (Futurewei), J. Tantsura (Microsoft), L. Ciavaglia (Nokia), M-P. Odini

Challenges and Opportunities in Green Networking

Abstract

Reducing technology's carbon footprint is one of the big challenges of our age. Networks are an enabler of applications that reduce this footprint, but also contribute to this footprint substantially themselves. The biggest opportunities to reduce the energy footprint may not be networking specific, for instance general power efficiency gains in hardware or hosting of equipment in more cooling-efficient buildings. Yet methods to make networking technology itself "greener" also need to be explored. This document outlines a corresponding set of opportunities, along with associated research challenges, for reducing this footprint and reducing network energy demand.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

Climate change and the need to curb greenhouse emissions have been recognized by the United Nations and by most governments as one of the big challenges of our time. As a result, improving energy efficiency and reducing power consumption are becoming of increasing importance for society and for many industries. The networking industry is no exception.

Arguably, networks can already be considered "green" technology in that networks enable many applications that allow users and whole industries to save energy and become more sustainable in a significant way. For example, it allows (at least to an extent) to replace travel with teleconferencing; it enables many employees to work from home and "telecommute," thus reducing the need for actual commute; IoT applications that facilitate automated monitoring and control from remote sites help make agriculture more sustainable by minimizing the application of resources such as water and fertilizer; networked smart buildings allow for greater energy optimization and sparser use of lighting and HVAC (heating, ventilation, air conditioning) than their non-networked not-so-smart counterparts.

The IETF has recently initiated a reflection on the energy cost of hosting meetings three times a year (see for instance https://www.ietf.org/blog/towards-a-net-zero-ietf/). It conducted a study of the carbon emissions of a typical meeting, and found out that 99% of the emissions were due to the air travel. In the same vein, [framework] compared an in-person with a virtual meeting and found a reduction in energy of 66% for a virtual meeting. These findings confirm that networking technology can reduce emissions when acting as virtual substitution for physical events.

That said, networks themselves consume significant amounts of energy. Therefore, the networking industry has an important role to play in meeting sustainability goals not just by enabling others to reduce their reliance on energy, but by also reducing its own. Future networking advances will increasingly need to focus on becoming more energy-efficient and reducing carbon footprint, both for economic reasons and for reasons of corporate responsibility. This shift has already begun and sustainability is already becoming an important concern for network providers. In some cases such as in the context of networked data centers, the ability to procure enough energy becomes a bottleneck prohibiting further growth and greater sustainability thus becomes a business necessity.

For example, in its annual report, Telefonica reports that in 2020, its network's energy consumption per PB of data amounted to 78MWh [telefonica2020]. This rate has has been dramatically decreasing (a five-fold factor over five years) although gains in efficiency are being offset by simultaneous growth in data volume. In the same report, it is stated as an important corporate goal to continue on that trajectory and reduce overall carbon emissions by 70% over the next 5 years.

Perhaps the most obvious gains in sustainability can be made with regards to improving the efficiency with which networks utilize power, reducing the amount of energy that is required to provide communication services. However, for a holistic approach other aspects need to be considered as well. For one, the sustainability of power sources need to be taken into account. A deployment that includes devices that are less energy-efficient but that are powered by a sustainable energy source can arguably be considered "greener" than a deployment that includes highly-efficient device powered by Diesel generators. In fact, in the same Telefonica report, extensive reliance on renewable energy sources is emphasized. Similar, deployments can take other environmental factors into account that affect carbon footprint. For example, deployments in which factors such as the need for cooling are reduced will be considered greener than deployments where this is not the case. Examples include deployments in cooler natural surroundings (e.g. in colder climates) where that is an option. Finally, manufacturing and recycling of networking equipment are also part of the sustainability equation. Extending the lifetime of equipment may in many cases be preferable over replacing it earlier with slightly more energy-efficient.

From a technical perspective, multiple vectors along which networks can be made "greener" should be considered:

* At the equipment level. Perhaps the most promising vector for improving networking sustainability concerns the network equipment itself. At the most fundamental level, networks (even softwarized ones) involve appliances, i.e. equipment that relies on electrical power to perform its function. However, beyond making those appliances merely energy-efficient, there are other important ways in which equipment can help networks become greener. This includes aspects such as support for port power saving modes allowing to reduce power consumption for resources that are not fully utilized, but also management instrumentation that allows to precisely monitor power usage at different levels of granularity, enabling (for example) controllers applications that aim to optimize energy usage across the network. (As a side note, the term "device", as used in the context of this draft, is used to refer to networking equipment. We are not taking into consideration end-user devices and endpoints such as mobile phones or computing equipment.)

* At the protocol level. Energy-efficiency and greenness are aspects that are rarely considered when designing network protocols. This suggests that there may be plenty of untapped potential. Some aspects involve designing protocols in ways that reduce the need for redundant or wasteful transmission of data to allow not only for better network utilization, but greater goodput per unit of energy being consumed. Techniques include approaches that reduce the "header tax" incurred by payloads as well as methods resulting in the reduction of wasteful retransmissions. Likewise, aspects such as restructuring addresses in ways that allow to minimize the size of lookup tables and associated memory sizes and their energy use can play a role as well. Another role of protocols concerns the enabling of functionality to improve energy efficiency at the network level, such as discovery protocols that allow for quick adaptation to network components being taken dynamically into and out of service depending on network conditions.

* At the network level. Perhaps the greatest opportunities to realize power savings exist at the level of the network as whole. For example, optimizing energy efficiency may involve directing

traffic in such a way that it allows for isolation of equipment that may at the moment not be needed so that it could be powered down or brought into power-saving mode. By the same token, traffic should be directed in a way that requires bringing additional equipment online or out of power-saving mode in cases where alternative traffic paths are available for which the incremental energy cost would amount to zero. Likewise, some networking devices may be more power-intensive than others or powered by less-sustainable energy sources. Their use might be avoided unless required to meet peak capacity demands. Generally, incremental power consumption can be viewed as a cost metric that networks should strive to minimize and consider as part of routing and of network path optimization.

* At the architecture level. The current network architecture supports a wide range of applications, but does not take into account energy efficiency as one of its design parameters. One can argue that the most energy efficient shift of the last two decades has been the deployment of Content Delivery Network overlays: while these were set up to reduce latency and minimize bandwidth consumption, from a network perspective, retrieving the content from a local cache is also much greener. What other architectural shifts can produce energy consumption reduction?

We believe that network standardization organizations in general, and IETF in particular, can make important contributions to each of these vectors. In this document, we will therefore explore each of those vectors in further detail and for each point out specific challenges for IETF.

It should be noted that this document borrows to a fair extent material from a prior paper, [GreenNet22]. This material has been both expanded (for example, in terms of some of the opportunities) and pruned (for example, in terms of background on prior scholarly work). In addition, unlike the prior paper, this document focuses on and attempts to articulate specific challenges as related to work that could be championed by the IETF.

2. Definitions and Acronyms

TBD

3. Contributors to Network Energy Consumption

When exploring possibilities to improve energy efficiency, it is important to understand which aspects contribute to power consumption the most and hence where the greatest potential for power savings lies.

Power is ultimately drawn from devices. The power consumption of the device can be divided into the consumption of the core device - the backplane and CPU, if you will - as well as additional consumption incurred per port and line card. Furthermore it is important to understand the difference between power consumption when a resource is idling versus when it is under load. This helps to understand the incremental cost of additional transmission versus the initial cost of transmission.

In typical networking devices, only roughly half of the energy consumption is associated with the data plane [bolla2011energy]. An idle base system typically consumes more than half of the power over the same system running at full load [chabarek08], [cervero19]. Generally, the cost of the first bit is very high, as it requires powering up a device, port, etc. The cost of transmission of additional bits (beyond the first) is many orders of magnitude lower. Likewise, the incremental cost of incremental CPU and memory needed to process additional packets becomes fairly negligible. This means that a device's power consumption does not increase linearly with the volume of forwarded traffic. Instead, it resembles more of a step function in which power consumption stays roughly the same up to a certain volume of traffic, followed by a sudden jump when when additional resources need to be procured to support a higher volume of traffic. By the same token, generally speaking it is more energy-efficient to transmit a large volume of data in one burst (and turning off the interface when idling), instead of continuously transmitting at a lower rate. In that sense it can be the duration of the transmission that dominates the energy consumption, not the actual data rate.

The implications on green networking from an energy-savings standpoint are significant: Potentially the largest gains can be made when network resources can effectively be taken off the grid (i.e. isolated and removed from service so they can be powered down while not needed). Likewise, for applications where this is possible, it may be desirable to replace continuous traffic at low data rates with traffic that is sent in burst at high data rates, in order to potentially maximize the time during which resources can be idled.

At the same time, any non-idle resources should be utilized to the greatest extent possible as the incremental energy cost is negligible. Of course, this needs to occur while still taking other operational goals into consideration, such as protection against failures (allowing for readily-available redundancy and spare capacity in case of failure) and load balancing (for increased operational robustness). As data transmission needs tend to fluctuate wildly and occur in bursts, any optimization schemes need to be highly adaptable and allow for very short control loops.

As a result, emphasis needs to be given to technology that allows to (for example) (at the device level) exercise very efficient and rapid discovery, monitoring, and control of networking resources so that they can be dynamically be taken offline or back into service, without (at the network level) requiring extensive convergence of state across the network or recalculation of routes and other optimization problems, and (at the network equipment level) support rapid power cycle and initialization schemes.

4. Challenges and Opportunities - Equipment Level

Perhaps the most obvious opportunities to make networking technology more energy efficient exist at the equipment level. After all, networking involves physical equipment to receive and transmit data. Making such equipment more power efficient, have it dissipate less heat to consume less energy and reduce the need for cooling, making it eco-friendly to deploy, sourcing sustainable materials and facilitating recycling of equipment at the end of its life-cycle all contribute to making networks greener. More specific and unique to networking are schemes to reduce energy usage of transmission technology from wireless (antennas) to optical (lasers).

One critical aspect of the energy cost of networking is the cost to manufacture and deploy the networking equipment. This is outside of the scope of this document: we only consider the energy cost of running the network, as this is where the IETF can play a role. However, a holistic approach would include into this the embedded energy that is included in the networking equipment. One aspect for the IETF may be to consider impact of deploying new protocols on the rate of obsolescence of the equipment. For instance, incremental approaches that do not require to replace equipment right away - or even extend the lifetime of deployed equipment - would have a lower energy footprint. This is one important benefit also of technologies such as Software-Defined Networking and Network Function Virtualization, as they may allow support of new networking features through software updates without requiring hardware replacements.

An attempt compute not only the energy of running a network, but also the energy embedded into manufacturing the equipment is described in [emergy] . This is denoted by "emergy", a portmanteau for embedded energy. [junkyard] Likewise, an approach to recycling equipment and a proof of concept using old cell phones recycled into a "junkyard" data center are described in [emergy].

Beyond such "first-order" opportunities, network equipment just as importantly plays an important role to enable and support green networking at other levels. Of prime importance is the equipment's ability to provide visibility to management and control plane into its current energy usage. Such visibility enables control loops for energy optimization schemes, allowing applications to obtain feedback regarding the energy implications of their actions, from setting up paths across the network that require the least incremental amount of energy to quantifying metrics related to energy cost used to optimize forwarding decisions.

One prerequisite to such schemes is to have proper instrumentation in place that allows to monitor current power consumption at the level of networking devices as a whole, line cards, and individual ports. Such instrumentation should also allow to assess the energy efficiency and carbon footprint of the device as a whole. In addition, it would be desirable to relate this power consumption to data rates as well as to current traffic, for example, to indicate current energy consumption relative to interface speeds, as well as for incremental energy consumption that is expected for incremental traffic (to aid control schemes that aim to "shave" power off current services or to minimize the incremental use of power for additional traffic). This is an area where the current state of the art is sorely lacking and standardization lags behind; for example, as of today, no corresponding standardized YANG data models [RFC7950] for network energy consumption that can be used in conjunction with management and control protocols have been defined.

Instrumentation should also take into account the possibility of virtualization, introducing layers of indirection to assess the actual energy usage. For example, virtualized networking functions could be hosted on containers or virtual machines which are hosted on a CPU in a data center instead of a regular network appliance such as a router or a switch, leading to very different power consumption characteristics. For example, a data center CPU could be more power efficient and consume power more proportionally to actual CPU load. Instrumentation needs to reflect these facts and facilitate attributing power consumption in a correct manner.

Beyond monitoring and providing visibility into power consumption, control knobs are needed to configure energy saving policies. For instance, power saving modes are common in endpoints (such as mobile phones or notebook computers) but sorely lacking in networking equipment.

5. Challenges and Opportunities - Protocol Level

There are several opportunities for energy savings at the protocol level. We characterize them along three main categories: protocols designed to reduce the volume of data to be transmitted; protocols designed to optimize data transmission rates under energy considerations; and protocols that enable energy optimization schemes at the network level. A fourth category, "other", is used to capture any other aspects not easily categorized into the other three.

5.1. Data Volume Reduction

The first category involves designing protocols in such a way that they reduce the volume of data that needs to be transmitted for any given purpose. Loosely speaking, by reducing this volume, more traffic can be served by the same amount of networking infrastructure, hence reducing overall energy consumption. Possibilities here include protocols that avoid unnecessary retransmissions. At the application layer, protocols may also use coding mechanisms that encode information close to the Shannon limit. Currently, most of the traffic over the Internet consists of video streaming and encoders for video are already quite efficient and keep improving all the time, resulting in energy savings as one of many advantages (of course being offset by increasingly higher resolution). However, it is not clear that the extra work to achieve higher compression ratios for the payloads results in a net energy gain: what is saved over the network may be offset by the compression/decompression effort. Further research on this aspect is necessary.

At the transport protocol layer, TCP and to some extent QUIC react to congestion by dropping packets. This is a highly energy inefficient method to signal congestion, since the network has to wait one RTT to be aware that the congestion has occurred, and since the effort to transmit the packet from the source up until it is dropped ends up being wasted. This calls for new transport protocols that react to congestion without dropping packets. ECN[RFC2481] is a possible solution, however not widely deployed. DC-TCP [alizadeh2010DCTCP] is tuned for Data Centers, L4S is an attempt to port similar functionality to the Internet [I-D.ietf-tsvwg-l4s-arch]. Qualitative Communication [QUAL] [westphal2021qualitative] allows the nodes to react to congestion by dropping only some of the data in the packet, thereby only partially wasting the resource consumed by transmitted the packet up to this point. Novel transport protocols for the WAN can ensure that no energy is wasted transmitting packets that will be eventually dropped.

Another solution to reduce the bandwidth of network protocols by reducing their header tax, for example applying header compression. An example in IETF is [RFC3095]. Again, reducing protocol header size saves energy to forward packets, but at the cost of maintaining a state for compression/decompression, plus computing these operations. The gain from such protocol optimization further depends on the application and whether it sends packets with large payloads close to the MTU (the header tax and any savings here are very limited), or whether it sends packets with very small payload size (making the header tax more pronounced and savings more significant).

An alternative to reducing the amount of protocol data is to design routing protocols that are more efficient to process at each node. For instance, path based forwarding/labels such as MPLS [RFC3031] facilitate the next hop look-up, thereby reducing the energy consumption. It is unclear if some state at router to speed up look up is more energy efficient that "no state + lookup" that is more computationally intensive. Other methods to speed up a next-hop lookup include geographic routing (e.g. [herzen2011PIE]). Some network protocols could be designed to reduce the next hop look-up computation at a router. It is unclear if Longest Prefix Match (LPM) is efficient from an energy point of view or if constitutes a significant energy burden for the operation of a router.

5.2. Protocol Optimization

The second category involves designing protocols in such a way that the rate of transmission is chosen to maximize energy efficiency. For example, Traffic Engineering (TE) can be manipulated to impact the rate adaptation mechanism [ren2018jordan]. By choosing where to send the traffic, TE can artificially congest links so as to trigger rate adaptation and therefore reduce the total amount of traffic. Most TE systems attempt to minimize Maximal Link Utilization (MLU) but energy saving mechanisms could decide to do the opposite (maximize minimial link utilization) and attempt to turn off some resources to save power.

Another example is to set up the proper rate of transmission to minimize the flow completion time (FCT) so as to enable opportunities to turn off links. In a wireless context, [TradeOff] studies how setting the proper initial value for the congestion window can reduce the FCT and therefore allow the equipment to go faster into a low-energy mode. By sending the data faster, the energy cost can be significantly reduced. This is a simple proof of concept, but protocols that allow for turning links into a low-power mode by transmitting the data over shorter periods could be designed for other types of networks beyond WiFi access. This should be done carefully: in the limit, a high rate of transmission over a short period of time may create bursts that the network would need to accommodate, with all attendant complications of bursty traffic. We conjecture there is a sweet spot between trying to complete flows faster while controlling for burstiness in the network. It is probably advisable to attempt to send traffic paced yet in bulk rather than spread out over multiple round trips. This is an area of worthwhile exploration.

5.3. Enabling Network Energy Saving Mechanisms

Novel protocols are also needed in two dimensions: to discover what links are available and/or energy efficient. For instance, links may be turned off in order to save energy, and turned back on based upon the elasticity of the demand. Protocols should be devised to discover when this happens, and to have a view of the topology that is consistent with frequent topology updates due to power cycling of the network resources.

Also, protocols are required to quickly converge onto an energy-efficient path once a new topology is created by turning links on/off. Current routing protocols may provide for fast recovery in the case of failure. However, failures are hopefully relatively rare events, while we expect an energy efficient network to aggressively try to turn off links.

Some mechanism is needed to present to the management layer a view of the network that identifies opportunities to turn resources off (routers/links) while still providing an acceptable level of Quality of Experience (QoE) to the users. This gets more complex as the level of QoE shifts from the current Best Effort delivery model to more sophisticated mechanisms with, for instance, latency, bandwidth or reliability guarantees.

5.4. Network Addressing

There are other ways to shave off energy usage from networks. One example concerns network addressing. Address tables can get very large, resulting in large forwarding tables that require considerable amount of memory, in addition to large amounts of state needing to be maintained and synchronized. From an energy footprint perspective, both can be considered wasteful and offer opportunities for improvement. At the protocol level, rethinking how addresses are structured can allow for flexible addressing schemes that can be exploited in network deployments that are less energy-intensive by design. This can be complemented by supporting clever address allocation schemes that minimize the number of required forwarding entries as part of deployments.

6. Challenges and Opportunities - Network Level

Networks have been optimized for many years under many criteria, for example to optimize (maximize) network utilization and to optimize (minimize) cost. Hence, it is straighforward to add optimization for "greenness" (including energy efficiency, power consumption, carbon footprint) as important criteria.

This includes assessing the carbon footprints of paths and optimizing those paths so that overall footprint is minimized, then applying techniques such as path-aware networking or segment routing [RFC8402] to steer traffic along those paths. It also includes aspects such as considering the incremental energy usage in routing decisions. Optimizing cost has a long tradition in networking; many of the existing mechanisms can be leveraged for greener networking simply by introducing energy footprint as a cost factor. Low-hanging fruit include the inclusion of energy-related parameters as a cost parameter in control planes, whether distributed (e.g. IGP) or conceptually centralized via SDN controllers.

Other opportunities concern adding energy-awareness to dynamic path selection schemes, requiring corresponding instrumentation as mentioned earlier. Again, considerable energy savings can potentially be realized by taking resources offline (e.g. putting them into power-saving or hibernation mode) when they are not currently needed under current network demand and load conditions. Therefore, weaning such resources from traffic becomes an important consideration for energy-efficient traffic steering. This contrasts and indeed conflicts with existing schemes that typically aim to create redundancy and load-balance traffic across a network to achieve even resource utilization. This usually occurs for important reasons, such as making networks more resilient, optimizing service levels, and increasing fairness. One of the big challenges hence concerns how resource weaning schemes to realize energy savings can be accommodated while preventing the cannibalization of other important goals, counteracting other established mechanisms, and avoiding destabilization of the network.

As an important prerequisite to capture many of those opportunities, good abstractions (and corresponding instrumentation) that allow to easily assess energy cost and carbon footprint will be required. These abstractions need to account for not only for the energy cost associated with packet forwarding across a given path, but related cost for processing, for memory, for maintaining of state, to result in a holistic picture. Optimization of carbon footprint involves in many cases trade-offs that involve not only packet forwarding but also aspects such as keeping state, caching data, or running computations at the edge instead of elsewhere. (Note: there may be a differential in running a computation at an edge server vs. at an hyperscale DC. The latter is often better optimized than the latter.) Likewise, other aspects of carbon footprint beyond mere energy-intensity should be considered. For instance, some network segments may be powered by more sustainable energy sources than others, and some network equipment may be more environmentally-friendly to build, deploy and recycle, all of which can be reflected in abstractions to consider.

A related set of challenges concerns the fact that such schemes result in much greater dynamicity and continuous change in the network as resources may be getting steered away from (when possible) and then leveraged again (when necessary) in rapid succession. This imposes significant stress on convergence schemes that results in challenges to the scalability of solutions and their ability to perform in a fast-enough manner. Network-wide convergence imposes high cost and incurs significant delay and is hence not susceptible to such schemes. In order to mitigate this problem, mechanisms should be investigate that do not require convergence beyond the vicinity of the affected network device. Especially in cases where central network controllers are involved that are responsible for aspects such as configuration of paths and the positioning of network functions and that aim for global optimization, the impact of churn needs to be minimized. This means that, for example, extensive recalculation e.g. of routes and paths based on the current energy state of the network needs to be avoided.

An opportunity may lie in making a distinction between "energy modes" of different domains. For instance, in a highly trafficked core, the energy challenge is to transmit the traffic efficiently. The amount of traffic is relatively fluid (due to multiplexing of multiple sessions) and the traffic is predictable. In this case, there is no need to optimize on a per session basis nor even at a short time scale. In the access networks connecting to that core, though, there are opportunites for this fast convergence: traffic is much more bursty, less predictable and the network should be able to be more reactive. Other domains such as DCs may have also more variable workloads and different traffic patterns.

7. Challenges and Opportunities - Architecture Level

Another possibility to improve network energy efficiency is to organize networks in a way that they can best serve important applications so as to minimize energy consumption. Examples include retrieval of content or remote computation. This allows to minimize the amount of communication that needs to take place in the first place, although energy savings within the network may at least in part be offset by additional energy consumption elsewhere. The following are some examples that suggest that it may be worthwhile reconsidering the ways in which networks are architected to minimize their carbon footprint.

For example, Content Delivery Networks (CDNs) have reduced the energy expenditure of the Internet by downloading content near the users. The content is sent only a few times over the WAN, and then is served locally. This shifts the energy consumption from networking to storage. Further methods can reduce the energy usage even more [bianco2016energy][mathew2011energy][islam2012evaluating]. Whether overall energy savings are net positive depends on the actual deployment, but from the network operator's perspective, at least it shifts the energy bill away from the network to the CDN operator.

While CDNs operate as an overlay, another architecture has been proposed to provide the CDN features directly in the network, namely Information Centric Networks [ahlgren2012survey], studied as well in the IRTF ICNRG. This however shifts the energy consumption back to the network operator and requires some power-hungy hardware, such as chips for larger name look-ups and memory for the in-network cache. As a result, it is unclear if there is an actual energy gain from the dissemination and retrieval of content within in-network caches.

Fog computing and placing intelligence at the edge are other architectural directions for reducing the amount of energy that is spent on packet forwarding and in the network. There again, the trade-off is between performing computation in a an energy-optimized data center at very large scale, but requiring transmission of significant volumes of data across many nodes and long distances, versus performing computational tasks at the edge where the energy may not be used as efficiently (less multiplexing of resources, and smaller sites are inherently less efficient due to their smaller scale) but the amount of long-distance network traffic is significantly reduced. Softwarization, containers, microservices are direct enablers for such architectures, and the deployment of programmable network infrastructure (as for instance Infrastructure Processing Units - IPUs or smartNICs that offload some computations from the CPU onto the NIC) will help its realization. However, the power consumption characteristics of CPUs are different from those of NPUs, another aspect to be considered in conjunction with virtualization.

Other possibilities concern taking economic aspects into consideration impact, such as providing incentives to users of networking services in order to minimize energy consumption and emission impact. An example for this is given in [wolf2014choicenet], which could be expanded to include energy incentives.

Other approaches consider performing a late binding of data and functions to be performed on the data [krol2017NFaaS]. The COIN Research Group in IRTF focuses on similar issues. Jointly optimizing for the total energy cost, taking into account networking and computing (and the different energy cost of computing in an hyperscale DC vs an edge node) is still an area of open research.

In summary, rethinking of the overall network (and networked application) architecture can be an opportunity to significantly reduce the energy cost at the network layer, for example by performing tasks that involve massive communications closer to the user. To what extend these shifts result in a net reduction of carbon footprint is an important question that requires further analysis on a case-by-case basis.

8. Conclusions

How to make networks "greener" and reduce their carbon footprint is an important problem for the networking industry to address, both for societal and for economic reasons. This document has highlighted some of the technical challenges and opportunities in that regard, for example:

* Equipment instrumentation advances for improved energy-awareness, definition and standardization of granular management information;

* Protocol advances for improving the ratio of goodput to throughput and to reduce waste: reduction in header tax, in protocol verbosity, improvements in coding, etc.

* Protocol advances to enable rapidly taking down, bring back online, and discover availability and power saving status of networking resources while minimizing the need for reconvergence and propagation of state;

* Network advances to allow to dynamically take resources offline where feasible while minimizing churn;

* Energy footprint aware traffic steering and routing; carbon footprint as a traffic cost metric to optimize;

* Reorganization of networking architecture for important classes of applications (examples: content delivery, right-placing of computational intelligence) to optimize green foot print and holistic approaches to trade off carbon footprint between forwarding, storage, and computation;

* Security issues imposed by greater energy awareness, to minimize the new attack surfaces that would allow an adversary to turn off resources, or to waste energy;

* Reliability issues for a network that relies on fewer resource diversity, and with more operational complexity.

Of those, perhaps the key challenge to address right away concerns the ability to expose at a fine granularity the energy impact of any networking actions. Providing visibility into this will enable many approaches to come towards a solution. It will be key to implementing optimization via control loops that allow to assess the energy impact of decisiont taken. It will also help to answer questions such as: is caching - with the associated storage energy - better than retransmitting from a different server - with the associated networking cost? Is compression more energy-efficient once factoring the computation cost of compression vs transmitting uncompressed data? Which compression scheme is more energy efficient? Is energy saving of computing at an efficient hyperscale DC compensated by the networking cost to reach that DC? Is the overhead of gathering and transmitting fine-grained energy telemetry data offset by the total energy gain by ways of better decisions that this data enables? Is transmitting data to a LEO constellation compensated by the fact that once in the constellation, the networking is fueled on solar energy? Is the energy cost of sending rockets to place routers in Low Earth Orbit amortized over time?

Determining where the sweet spots are and optimizing networks along those lines will be a key towards making networks "greener". We expect to see significant advances across these areas and believe that IETF has an important role to play in facilitating this.

9. IANA Considerations

This document does not have any IANA requests.

10. Security Considerations

Security considerations may appear to be orthogonal to green networking considerations. However, there are a number of important caveats.

Security vulnerabilities of networks may manifest themselves in compromised energy efficiency. For example, attackers could aim at increasing energy consumption in order to drive up attack victims' energy bill. Specific vulnerabilities will depend on the particular mechanisms. For example, in the case of monitoring energy consumption data, tampering with such data might result in compromised energy optimization control loops. Hence any mechanisms to instrument and monitor the network for such data need to be properly secured to ensure authenticity.

In some cases there are inherent tradeoffs between security and maximal energy efficiency that might otherwise be achieved. An example is encryption, which requires additional computation for encryption and decyption activities and security handshakes, in addition to the need to send more traffic than necessitated by the entropy of the actual data stream. Likewise, mechanisms that allow to turn resources on or off could become a target for attackers.

Energy consumption can be used to create covert channels, which is a security risk for information leakage. For instance, the temperature of an element can be used to create a Thermal Covert Channel[TCC], or the reading/sharing of the measured energy consumption can be abused to create a covert channel (see for instance [DRAM] or [NewClass]). Power information may be used to create side-channel attacks. For instance, [SideChannel] provides a review of 20 years of study on this topic. Any new parameters to consider in protocol designs or in measurements is susceptible to create such covert or side channel and this should be taken into account while designing energy efficient protocols.

11. Contributors

Michael Welzl, University of Oslo, michawe@ifi.uio.no

12. Acknowledgments

We thank Dave Oran for providing the information regarding covert channels using energy measurements. Additional acknowledgments will be added at a later stage.

13. Informative References (TBD)

[ahlgren2012survey] Ahlgren, B., Dannewitz, C., Imbrenda, C., Kutscher, D., and B. Ohlman, "A survey of information-centric networking", IEEE Communications Magazine Vol.50 No.7, 2012.

[alizadeh2010DCTCP]
Alizadeh, M., Greenberg, A., Maltz, D., Padhye, J., Patel, P., Prabhakar, B., Sengupta, S., and M. Sridharan, "Data Center TCP (DCTCP)", ACM SIGCOMM pp.63-74, 2010.

[bianco2016energy] Bianco, A., Mashayekhi, R., and M. Meo, "Energy consumption for data distribution in content delivery networks", IEEE International Conference on Communications (ICC) pp.1-6, 2016.

[bolla2011energy] Bolla, R., Bruschi, R., Davoli, F., and F. Cucchietti, "Energy Efficiency in the Future Internet: A Survey of Existing Approaches and Trends in Energy-Aware Fixed Network Infrastructures", IEEE Communications Surveys and Tutorials Vol.13 No.2, pp.223-244, 2011.

[cervero19] Cervero, A. G., Chincoli, M., Dittmann, L., Fischer, A., and A. Garcia, "Green Wired Networks", Wiley Journal on

Large-Scale Distributed Systems and Energy Efficiency pp. 41-80, 2019.

[chabarek08] Chabarek, J., Sommers, J., Barford, P., Tsiang, D., and S. Wright, "Power awareness in network design and routing", IEEE Infocom pp.457-465, 2008.

[DRAM] Paiva, T. B., Navaridas, J., and R. Terada, "Robust Covert Channels Based on DRAM Power Consumption", In book: Information Security (pp.319-338) , 2019.

[emergy] Raghavan, B. and J. Ma, "The Energy and Emergy of the Internet", ACM HotNets , 2011.

[framework] Faber, G., "A framework to estimate emissions from virtual conferences", International Journal of Environmental Studies, 78:4, 608-623 , 2021.

[GreenNet22] Clemm, A. and C. Westphal, "Challenges and Opportunities in Green Networking", 1st International Workshop on Network Energy Efficiency in the Softwarization Era IEEE NetSoft 2022, June 2022.

[herzen2011PIE] Herzen, J., Westphal, C., and P. Thiran, "Scalable routing easy as PIE: A practical isometric embedding protocol", 19th IEEE International Conference on Network Protocols (ICNP) pp.49-58, 2011.

[I-D.ietf-tsvwg-l4s-arch] Briscoe, B., Schepper, K. D., Bagnulo, M., and G. White, "Low Latency, Low Loss, Scalable Throughput (L4S) Internet Service: Architecture", Work in Progress, Internet-Draft, draft-ietf-tsvwg-l4s-arch-20, 29 August 2022, <https://www.ietf.org/archive/id/draft-ietf-tsvwg-l4s-arch-20.txt>.

[islam2012evaluating] Islam, S. U. and J. Pierson, "Evaluating Energy Consumption in CDN Servers", Proceedings of the Second International Conference on ICT as Key Technology against Global Warming pp.64-78, 2012.

[junkyard] Switzer, J., Kastner, R., and P. Pannuto, "Architecture of a Junkyard Datacenter", arXiv:2110.06870v1, October 2021 , 2021.

[krol2017NFaaS] Krol, M. and I. Psaras, "NFaaS: Named Function as a Service", ACM SIGCOMM ICN Conference , 2017.

[mathew2011energy] Mathew, V., Sitaraman, R., and P. Shenoy, "Energy-Aware Load Balancing in Content Delivery Networks", CoRR http://arxiv.org/abs/1109.5641 , 2011.

[NewClass]
Khatamifard, S. K., Wang, L., Kose, S., and U. R. Karpuzcu, "A New Class of Covert Channels Exploiting Power Management Vulnerabilities", IEEE Computer Architecture Letters , 2018.

[QUAL] Li, R., Makhijani, K., Yousefi, H., Westphal, C., Xong, L., Wauters, T., and F. D. Turck, "A framework for Qualitative Communications using Big Packet Protocol", Proceedings ACM Sigcomm Workshop On Networking For Emerging Applications And Technologies pp.22-28, 2019.

[ren2018jordan] Ren, J., Ren, K., Westphal, C., Wang, J., Wang, J., Song, T., Liu, S., and J. Wang, "JORDAN: A Novel Traffic Engineering Algorithm for Dynamic Adaptive Streaming over HTTP", IEEE International Conference on Computing, Networking and Communications (ICNC) pp.581-587, 2018.

[RFC2481] Ramakrishnan, K. and S. Floyd, "A Proposal to add Explicit Congestion Notification (ECN) to IP", RFC 2481, DOI 10.17487/RFC2481, January 1999, <https://www.rfc-editor.org/info/rfc2481>.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <https://www.rfc-editor.org/info/rfc3031>.

[RFC3095]
Bormann, C., Burmeister, C., Degermark, M., Fukushima, H., Hannu, H., Jonsson, L-E., Hakenberg, R., Koren, T., Le, K., Liu, Z., Martensson, A., Miyazaki, A., Svanbro, K., Wiebke, T., Yoshimura, T., and H. Zheng, "RObust Header Compression (ROHC): Framework and four profiles: RTP, UDP, ESP, and uncompressed", RFC 3095, DOI 10.17487/RFC3095, July 2001, <https://www.rfc-editor.org/info/rfc3095>.

[RFC7950] Bjorklund, M., Ed., "The YANG 1.1 Data Modeling Language", RFC 7950, DOI 10.17487/RFC7950, August 2016, <https://www.rfc-editor.org/info/rfc7950>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[SideChannel] Randolph, M. and W. Diehl, "Power Side-Channel Attack Analysis: A Review of 20 Years of Study for the Layman", Cryptography 2020, 4, 15 , 2020.

[TCC]
Rahimi, P., Singh, A. K., and X. Wang, "Selective Noise Based Power Efficient and Effective Countermeasure Against Thermal Covert Channel Attacks in Multi-Core Systems", Journal on Low Power Electronics and Applications , 2022.

[telefonica2020] Telefonica, "Consolidated Management Report 2020", 2021.

[TradeOff] Welzl, M., "Not a Trade-Off: On the Wi-Fi Energy Efficiency of Effective Internet Congestion Control", IEEE/IFIP WONS , 2022.

[westphal2021qualitative] Westphal, C., He, D., Makhijani, K., and R. Li, "Qualitative Communications for Augmented Reality and Virtual Reality", 22nd IEEE International Conference on High Performance Switching and Routing (HPSR) pp.1-6, 2021.

[wolf2014choicenet]
Tilman, W., Griffioen, J., Calvert, L., Dutta, R., Rouskas, G., Baldin, I., and A. Nagurney, "ChoiceNet: Toward an Economy Plane for the Internet", SIGCOMM Computer Communciations Review Vol.44 No.3, July 2014.

Authors' Addresses

Alexander Clemm
Futurewei
2330 Central Expressway
Santa Clara,, CA 95050
United States of America

Email: ludwig@clemm.org

Cedric Westphal
Futurewei

Email: cedric.westphal@futurewei.com

Jeff Tantsura
Microsoft

Email: jefftant.ietf@gmail.com

Laurent Ciavaglia
Nokia

Email: laurent.ciavaglia@nokia.com

Marie-Paule Odini

Email: mp.odini@orange.fr