

Network Working Group
Internet-Draft
Intended status: Informational
Expires: 13 November 2025

C. Pignataro, Ed.
Blue Fern Consulting
A. Rezaki
Nokia
S. Krishnan
Cisco
J. Arkko
Ericsson
A. Clemm
Sympotech
H. ElBakoury
Independent Consultant
S. Prabhu
Nokia
12 May 2025

Architectural Considerations for Environmental Sustainability
draft-pignataro-enviro-sustainability-architecture-02

Abstract

This document describes several of the design tradeoffs involved in optimizing for sustainability along with the other common networking metrics such as performance and availability.

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 13 November 2025.

Copyright Notice

Copyright (c) 2025 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.

Table of Contents

1. Introduction	2
1.1. Terminology	3
2. Architectural Considerations of Environmental Sustainability	3
2.1. Design Tradeoffs	3
2.2. Multi-Objective Optimization	4
2.3. How Much Resiliency is Really Needed?	5
2.3.1. Redundancy and Sustainability	6
2.4. How Much are Performance and Quality of Experience Compromised?	6
2.5. End-to-End Sustainability	7
2.6. Attributional and Consequential Models	7
2.7. The Role of Network Management and Orchestration	8
3. Sustainability Requirements and Phases	11
3.1. Phase 1: Visibility	11
3.2. Phase 2: Insights and Recommendations	12
3.3. Phase 3: Self-optimization and Automation	12
3.3.1. Cycle of Phases	12
4. Security Considerations	13
5. Acknowledgements	13
6. Informative References	13
Authors' Addresses	14

1. Introduction

Over the past decade, there has been increased awareness of the environmental sustainability impact produced by the widespread adoption of the Internet and internetworking technologies. The impact of Internet technologies has been overwhelmingly positive over the past years (e.g., providing alternatives to travel, enabling remote and hybrid work, enabling technology-based endangered species conservation, etc.), and there is still room for improvement.

At the same time, internetworking technologies themselves have a significant environmental footprint. Reducing this footprint and making network deployments environmentally more sustainable becomes a matter of increasing importance to network providers for various reasons, including the desire to reduce operational expenses associated with energy usage, regulatory pressures related to Net Zero mandates, and corporate citizenship demands from customers to become "greener".

Efforts to make internetworking more sustainable may in some cases conflict with other important goals, such as network availability, resilience against sudden spikes in traffic demand, and (at least in some cases) Quality of Experience. For example, a network operator might want to include excess capacity in their network to be able to absorb sudden spikes in demand and provision additional paths to increase resilience against failures. However, doing so may involve powering up additional ports and equipment, resulting in higher energy usage and increased greenhouse gas emissions, thus compromising sustainability goals. This indicates that there are certain choices that a network operator has to make, some of which may involve tradeoffs between goals.

This document describes some of the tradeoffs that could be involved while optimizing for sustainability in addition to or in lieu of traditional metrics such as performance or availability. Further, it discusses how Internet technologies can be used to help other fields become more sustainable.

Specifically, this document details environmental sustainability implications to Internet protocols, architectures, and technologies.

1.1. Terminology

This document leverages the terminology and concepts defined in [I-D.pignataro-green-enviro-sust-terminology], and readers are expected to be familiar with those.

2. Architectural Considerations of Environmental Sustainability

2.1. Design Tradeoffs

Traditionally, digital communication networks are optimized for a specific set of criteria that serve as proxies for business metrics. A network operator providing services to their customers intends to maximize profits, by increasing top-line revenue and decreasing bottom-line associated costs. This directly translates to goals of optimizing performance and availability, while reducing various costs.

Most recently, various forces elevate the need for sustainability in networking technologies and architectures, to quantify and minimize negative environmental impact.

Optimizing only network availability (e.g., by having excess capacity and backup paths) or optimizing only performance (e.g., by increasing speeds or selecting paths based solely on delays) can seemingly be in opposition to optimizing sustainability objectives. For a given application, use-case, or vertical realization of technology, a Pareto-efficient choice can potentially improve sustainability without sacrificing availability or performance beyond the application tolerance. That is, a win-win.

Consequently, network architects and designers are presented with a set of new design tradeoffs: a multi-objective optimization that satisfies broader requirements and global optima for availability, performance, and sustainability simultaneously. This is not unlike the doughnut economics model concept described in [Doughnut].

2.2. Multi-Objective Optimization

To understand this new model, we can analyze a simplified example. Assume the following topology, passing traffic from A to B:

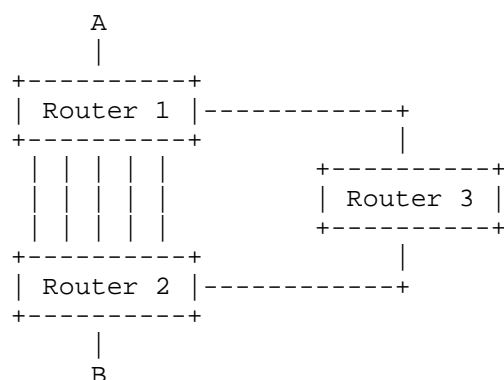


Figure 1: Simplified Network for Multi-Objective Optimization

Router 1 is directly connected to Router 2 through five parallel links, of 10 Gbps each. Router 1 can also reach Router 2 through Router 3 with 40 Gbps links between Router 1 and Router 3, and between Router 3 and Router 2. Let's assume that the capacity-planned traffic between A and B equals 15 Gbps.

In this scenario, a topology optimized for performance and availability/resiliency would have all links and routers on, and would likely forward traffic using two of the parallel links. Utilizing the path through Router 3 might lower performance, but it serves as a backup path.

On the other hand, when we add sustainability as a consideration, different options present themselves, each of which involves a tradeoff. One of the options is to remove from the topology Router 3 and associated links, and shutdown links and optics in two or three of the parallel links. This will allow the conservation of energy that will no longer be required to operate Router 3 and the affected links. Another option is to completely shutdown all the parallel links and route traffic through Router 3 (i.e., not maximizing performance alone, but maximizing at the time performance, availability and resiliency, and sustainability.) The choice between these two options, as well as the option to stick with the original topology, will depend on choices that the network operator will have to make, taking into account the aggregate sustainability metrics of network elements in each of the two topologies as well as the effect these choices will have on availability/resiliency which will be reduced as a result.

Another option is to use flexible Ethernet, where the five links combined are aggregated into one active virtual link which has 15 Gbps, and another inactive link of 35 Gbps of capacity -- although a physical link cannot be half-active and half-inactive as far as PHY and optics are concerned.

2.3. How Much Resiliency is Really Needed?

When we add sustainability considerations, resiliency is not the single objective to optimize.

There are many methods to improve network resiliency, including a design eliminating single-points-of-failure, performing software safe-release selections and upgrades, deploying network real-time testing systems (including operations, administration, and maintenance (OAM) tools, monitoring systems (e.g., [RFC8403]), chaos-based testing, and site reliability engineering (SRE) principles), and utilizing redundancy across network elements as well as across a topology. Each one of these methods incurs also a sustainability cost. Yet, the functions for resiliency improvement and sustainability cost for each of these methods are not the same. Considering sustainability means quantifying its impact in the decision of how to improve resiliency, and how much is needed.

2.3.1. Redundancy and Sustainability

Let's first explore redundancy. For example, consider the ratio of overall network capacity (in bandwidth, compute power, etc.) over the used network capacity, and let's call it "Redundancy Index". If this number is one, there's no redundancy; and as the ratio grows, so does the potentially unused capacity that could be utilized in a failure event. Similarly, consider the values of sustainability metrics for when the Redundancy Index is one and for when it is two. These border points might give an indication of the slope for each objective function.

Adequate Redundancy:

In order to determine how much redundancy needs to be built into the overall network capacity, which can be referred to as "adequate redundancy to avoid network outages", it will be important to (1) measure the bandwidth of attacks against the overall network capacity; and (2) understand how quickly "high bandwidth" attacks can be detected and diverted. Measuring these results over time may lead the "adequate redundancy" to become higher over time.

Justified Redundancy:

Traditionally, network operators would be much less worried about energy use than about the possibility that the network would have brownout or backout outages - thus the measuring will help better balance the "adequate redundancy" against the related energy use, resulting in turn in "justified redundancy": a balance between costs and benefits, with energy use as well as material use as a clear cost factor.

Please note that "justified redundancy" may be higher than "adequate redundancy" when we manage to organize the redundancy in a multi-layer fashion: (1) capacity that is "always on" and always uses energy; (2) capacity that can turn on quickly when needed; (and possibly (3) capacity that is "on the shelf" (even in the box) but ready to be deployed quickly when needed.)

2.4. How Much are Performance and Quality of Experience Compromised?

Network performance and Quality of Experience have always played an important role in the development of networking technology. Key parameters such as latency, jitter, and loss as well as their impact on the quality that the users of networked applications experience are well understood, the optimization of those parameters and the adherence to corresponding service level objectives being an important goal in most network deployments. However, the desire to improve sustainability and energy efficiency can conflict with those

goals. For example, in order to ensure minimal latency, a network operator may need to provision additional paths that require additional ports that need to be powered, instead of relying on a topology with fewer links and nodes. Such a topology might result in greater power efficiency as a result of more resources being shared, but it could also result in longer paths and an increased possibility for congestion, both of which would be detrimental to latency and associated Quality of Experience.

This implies a tradeoff between different goals. The challenge for operators lies in finding the sweet spot in which acceptable network performance is obtained and a point of diminishing returns is reached at which any incremental further performance improvement would come at the expense of significant deterioration in energy usage.

2.5. End-to-End Sustainability

The networking industry is in the starting phases of addressing this objective. We are seeing a sprinkling of sustainability features across the networking stack and components of devices, whether it is on forwarding chips, power supplies, optics, and compute. Many of those optimizations and features are typically local in nature, and widely scattered across different elements of a network architecture. An opportunity for maximizing the positive environmental impact of these technologies calls for a more cohesive and complementary view that spans the complete product lifecycle for hardware and software, as well as how some of these features work in unison.

For example, features that provide energy saving modes for devices can be dynamically managed when the network utilization is such that performance would not significantly suffer. A core router, instead of becoming obsolete due to the need for higher throughput in the core, could become a future edge/access router. That is an example of reuse and repurpose, before recycling. There are benefits of macro-optimizations by clustering in specific features, versus micro-optimizing locally without awareness of the network context.

2.6. Attributional and Consequential Models

Many of the subtleties and nuances of the measurement of GHG and environmental impacts stem from the very important distinction between attributional and consequential models. Detailed definitions can be found at [UNEP-LCA].

Attributional:

Also referred to as Allocational models, start by allocating or attributing quantities (e.g., GHG emissions) to entities (e.g., a router, a building, a town), and performing comparisons between the measurements (or estimates) of the quantity by the entities.

Consequential:

Perform the measurement of the quantity by establishing a baseline scenario (e.g., before feature introduction) and a modified scenario (e.g., after the feature introduction).

While both models are quite different, they do use the same terms and frames of references, measures, and language. Without explicit clarifications, they are prone to confusion.

For example, measuring the carbon footprint attributed to a batch process or a workload based on its energy efficiency would not consider that the hardware is still there running. When is it most effective to charge battery-powered devices, during the night when there's less load, or during the day when there's solar energy? In other words, if a person who commutes by train to their office five days a week starts working from home two days a week, there could be an attributional reduction of GHG emissions, yet the train still continues running equally. However, if that person commutes by combustion-engine car alone, the consequences are different.

Considering the attributional versus consequential distinction, there are some implications and a potential corollaries:

- * For an environmental-impact analysis, it is critical to explicitly cite the model used, as well as clearly define the boundary.
- * The activities that we embark upon as internetworking and protocol designers - including the ones targeting reduction of negative environmental impacts - have an energy footprint of themselves.
- * "Do no harm" in the context of improving sustainability of networks is to look beyond bounded attributions and consider (both intended and unintended) consequences.

2.7. The Role of Network Management and Orchestration

Deployment and operational aspects play a critical role in making networks more sustainable. A detailed explanation of that role, the associated challenges, as well as an outline of solution approaches is provided in [I-D.irtf-nmrg-green-ps]. Here are some areas in which network management can help make networks more sustainable; for a more extensive treatment, please refer to that document.

Dimensioning:

Networks should be deployed and configured with sufficient capacity to serve their intended purpose. At the same time, overprovisioning and providing too many resources should be avoided, as this results in waste and unnecessary environmental impact. Network management can facilitate proper dimensioning of networks by providing the methods and tools that allow to assess network usage, determine required capacities, identify trends to allow to proactively accommodate traffic growth and new services.

Network Optimization:

Network management applications can help solve difficult network optimization problems involving multiple parameters, multiple and sometimes conflicting objectives, and mitigation of tradeoffs. Network optimization examples include maximization of utilization or of aggregate QoE scores, minimization of the possibility of SLA violations with a given amount of network resources, or optimization of the cost of configured paths. Network metrics related to sustainability are another set of parameters that can be optimized.

Rapid Discovery and Provisioning Schemes:

One of the biggest potential opportunities in reducing environmental impact of networks concerns the ability to power resources such as equipment or line cards down when they are momentarily not needed due to swings in traffic demands. In general, this involves fully automated management control loops with very short time scales. Network management can enable such schemes, involving algorithms that determine and control the rapid de- and re-commissioning of networking resources, as well as the necessary control protocols that facilitate aspects such as rapid resource discovery, reprovisioning, or reconvergence of management state.

Policy-Driven Sustainability Enforcement:

Network management systems can play a pivotal role in enforcing explicit sustainability policies, much like how QoS, security, or routing policies are enforced today. These policies can express environmental objectives such as limiting power consumption within specific network domains, prioritizing traffic through paths powered by renewable energy, or dynamically adjusting service parameters to meet carbon footprint targets. Such policy-driven approaches allow sustainability intents to be specified at a high level and translated into actionable configurations using orchestration frameworks. By embedding sustainability into policy and intent-based networking models, operators gain precise control over how environmental goals are operationalized and maintained across diverse services and tenants.

Inter-Domain Sustainability Coordination:

Sustainability optimization must extend beyond individual administrative domains to realize its full potential at Internet scale. Network management and orchestration systems can be enhanced to support inter-domain coordination mechanisms that allow operators to share sustainability-related metadata, such as real-time carbon intensity of regional infrastructure, green routing preferences, or energy availability status. By enabling cooperative decision-making, networks can collectively route traffic in ways that reduce aggregate environmental impact. This requires the definition of interoperable data models, trust frameworks, and privacy-preserving methods for sharing sustainability metrics across organizational boundaries. As sustainability becomes a global imperative, inter-domain orchestration will be essential to align local optimizations with broader planetary goals.

To enable end-to-end sustainability, environmental objectives must persist across the full service path, including ingress, transit, and egress domains. This requires a mechanism for expressing sustainability intents (e.g., “carbon-sensitive” or “low-energy”) at the service origin and ensuring they are respected downstream. These intents may influence route selection, resource allocation, and power state decisions in intermediate networks. Failure to propagate such goals may result in sustainability regressions that cancel out upstream efforts.

Achieving this coordination demands standardized sustainability telemetry formats and semantic models. Exchanged data may include per-domain carbon intensity, real-time energy sourcing, or equipment-level energy efficiency indicators. Agreement on common ontologies and encoding formats will be essential to ensure compatibility across vendor and operator implementations.

Beyond metrics, cooperative orchestration protocols will be needed to act on this shared information. For example, inter-domain green routing agreements may optimize for end-to-end energy profiles in addition to latency or cost. Sustainability-aware SLAs could encode carbon or energy constraints alongside traditional service guarantees. Trust boundaries, policy asymmetries, and privacy concerns may necessitate abstraction layers, optional disclosure levels, or brokered negotiation intermediaries.

In addition to those aspects, perhaps the most important role of network management is to provide network operators with the necessary visibility into how and where power is used in their network. This is required in order to assess where the network stands in terms of

sustainability. It also allows to track progress over time, compare different alternatives for their effectiveness, and generally to facilitate network sustainability optimization. Providing this visibility requires the definition of metrics and corresponding instrumentation of the network so that those metrics can be monitored, assessed, compared, and improved.

3. Sustainability Requirements and Phases

The architectural considerations for environmental sustainability cannot always be achieved at the same time and we expect the following high level phases:

1. **Visibility:** In this phase we focus on the measurement and collection of metrics.
2. **Insights and Recommendations:** In this phase we focus on deriving insights and providing recommendations that can be acted upon manually over large time scales.
3. **Self-Optimization via Automation:** In this phase we build awareness into the systems to automatically recognize opportunities for improvement and implement them.

3.1. Phase 1: Visibility

Visibility represents collecting and organizing data in a standard vendor agnostic manner. The first step in improving our environmental impact is to actually measure it in a clear and consistent manner. The IETF, IRTF and the IAB have a long history of work in this field, and this has greatly helped with the instrumentation of network equipment in collecting metrics for network management, performance, and troubleshooting. On the environmental-impact side though, there has been a proliferation of a wide variety of vendor extensions based on these standards. Without a common definition of metrics across the industry and widespread adoption we will be left with ill-defined, potentially redundant, proprietary, or even contradicting metrics. Similarly, we also need to work on standard telemetry for collecting these metrics so that interoperability can be achieved in multi-vendor networks.

3.2. Phase 2: Insights and Recommendations

Once the metrics have been collected, categorized, and aggregated in a common format, it would be straightforward to visualize these metrics and allow consumers to draw insights into their GHG and energy impact. The visualizations could take the form of high-level dashboards that provide aggregate metrics and potentially some form of maturity continuum. We think this can be accomplished using reference implementations of the standards developed in "Phase 1: Visibility". We do expect vendors and other open projects to customize this and incorporate specific features. This will allow identifying sources of environmental impact and address any potential issues through operational changes, creation of best-practices, and changes towards a greener, more environmentally friendly equipment, software, platforms, applications, and protocols.

3.3. Phase 3: Self-optimization and Automation

Manually making changes as mentioned in "Phase 2: Insights and Recommendations" works for changes needed on large timescales but does not scale to improvements on smaller scales (i.e., it is impractical in many levels for an operator to be looking at a dashboard monitoring usage and making changes in real-time 24x7). There is a need to provision some amount of self-awareness into the network itself, at various layers, so that it can identify opportunities for improvement, implement the necessary changes, and measure the effects to complete the feedback loop. The goals of the consumers can be stated declaratively, and the networks can continually use mechanisms such as machine learning (ML), deep learning (DL), and artificial intelligence (AI) with an additional goal to optimize for improvements in the environmental impact. These include, for example:

- * Discovery and advertisement of networking characteristics that have either direct or indirect environmental impact,
- * greener networking protocols that can move traffic onto more energy efficient paths, directing topological graphs to optimize environmental impacts, and
- * protocols that can instruct equipment to move under-utilized links and devices into low-energy modes.

3.3.1. Cycle of Phases

The three phases run in an iterative fashion, such that after phases 1, 2, and 3 are completed for an iteration, there will be an added awareness of what (else) to collect back to phase 1.

Further, sustainability-aware self-optimization is something to explore in Autonomic Networking.

4. Security Considerations

Sustainable practices offer many environmental, economic, and social benefits, and security is a route to sustainability rather than a hurdle to clear.

The creation of sustainability features for an element or a system should not weaken or compromise their security posture, nor should it increase the surface of attack or create attack vectors.

- Sustainability metrics and data models ought to describe how to secure the sustainability data exposed and surfaced through telemetry.
- Sustainability control capabilities, as for example for power management, should consider potential attacks leveraging those controls. Setting a device on low-power or power-save modes during peak traffic can be a denial-of-service attack vector, negatively impacting end-to-end services.

The development of security features should, in turn, balance the environmental impact and sustainability considerations detailed in this document.

- Computational increase on cryptographic operations can result in higher power use. Since generally the increase of energy required is not linear with the increase of computational complexity, there's a desire to satisfy security requirements while minimizing environmental impact.
- Proof-of-Work schemes' and AI models' energy consumption also grows non-linearly as a function of the precision achieved. In these, perfect is the enemy of good, and bounding precision through specifications supports sustainable compute considerations.

5. Acknowledgements

This document is created greatly leveraging ideas and text from [I-D.cparsk-eimpact-sustainability-considerations], and consequently acknowledges all the many contributions that improved it.

6. Informative References

[Doughnut] Wikipedia, "Doughnut (economic model)", 13 October 2023, <[https://en.wikipedia.org/wiki/Doughnut_\(economic_model\)](https://en.wikipedia.org/wiki/Doughnut_(economic_model))>.

[I-D.cparsk-eimpact-sustainability-considerations]
Pignataro, C., Rezaki, A., Krishnan, S., ElBakoury, H., and A. Clemm, "Sustainability Considerations for Internetworking", Work in Progress, Internet-Draft, draft-cparsk-eimpact-sustainability-considerations-07, 24 January 2024, <<https://datatracker.ietf.org/doc/html/draft-cparsk-eimpact-sustainability-considerations-07>>.

[I-D.irtf-nmrg-green-ps]
Clemm, A., Pignataro, C., Westphal, C., Ciavaglia, L., Tantsura, J., and M. Oadini, "Challenges and Opportunities in Management for Green Networking", Work in Progress, Internet-Draft, draft-irtf-nmrg-green-ps-06, 15 March 2025, <<https://datatracker.ietf.org/doc/html/draft-irtf-nmrg-green-ps-06>>.

[I-D.pignataro-green-enviro-sust-terminology]
Pignataro, C., Rezaki, A., and H. ElBakoury, "Environmental Sustainability Terminology and Concepts", Work in Progress, Internet-Draft, draft-pignataro-green-enviro-sust-terminology-01, 27 December 2024, <<https://datatracker.ietf.org/doc/html/draft-pignataro-green-enviro-sust-terminology-01>>.

[RFC8403] Geib, R., Ed., Filsfils, C., Pignataro, C., Ed., and N. Kumar, "A Scalable and Topology-Aware MPLS Data-Plane Monitoring System", RFC 8403, DOI 10.17487/RFC8403, July 2018, <<https://www.rfc-editor.org/info/rfc8403>>.

[UNEP-LCA] "Global guidance principles for life cycle assessment databases : a basis for greener processes and products", 2011, <<https://www.lifecycleinitiative.org/library/global-guidance-principles-for-lca-databases-a-basis-for-greener-processes-and-products/>>.

Authors' Addresses

Carlos Pignataro (editor)
Blue Fern Consulting
United States of America
Email: cpignata@gmail.com, carlos@bluefern.consulting

Ali Rezaki
Nokia
Germany
Email: ali.rezaki@nokia.com

Suresh Krishnan
Cisco Systems, Inc.
United States of America
Email: sureshk@cisco.com

Jari Arkko
Ericsson
Email: jari.arkko@ericsson.com

Alexander Clemm
Sympotech
United States of America
Email: ludwig@clemm.org

Hesham ElBakoury
Independent Consultant
United States of America
Email: helbakoury@gmail.com

Shailesh Prabhu
Nokia
India
Email: shailesh.prabhu@nokia.com