IEEE ICIN
Conference on Innovation in Clouds, Internet and Networks
Thurdsay March 13th, 2025

Tutorial #3, 15:00 – 18:00

Greening the Network: Challenges and Solution Approaches For a Net Zero Internet

Cedric Westphal
cedric@ucsc.edu
University of California Santa Cruz, USA

Alexander Clemm
ludwig@clemm.org
Sympotech, California/USA

Cedric Westphal

- Currently teaching at University of California, Santa Cruz, Computer Science & Engineering Department
  - Before that, Futurewei, Nokia, Docomo and UCSC
- Over 120 papers, 30 patents, 3 IETF RFCs + some green networking draft working its way through the process
- PhD UCLA, MSs from UCLA & Ecole Centrale Paris
- Recipient of 2018 Technical Achievement Award from IEEE Information Infrastructure and Networking Technical Committee to "*recognize a lifelong set of outstanding technical contributions in the area of information infrastructure and networking*"

Alexander Clemm

- Currently in the stealth phase of new project “Sympotech”
- Formerly Distinguished Engineer at Futurewei
- 70+ papers, 70+ patents
- 16 RFCs
- Longtime member of the network management community (including as SC and OC member of IEEE/IFIP NOMS and NetSoft)
- Recipient of Salah Aidarous Award in 2020
- Contributions include programmable networking, network telemetry (e.g. YANG-Push), measurement (e.g. IPSLA), smart monitoring (e.g. DNA), network instrumentation, service provisioning, high-precision networking, and more
- Author of several Internet Drafts related to Sustainable Networking
- MS Stanford Univ, Dr. rer. nat. Univ of Munich

Tutorial Outline

1. Motivation (Alex)
2. Network sustainability foundations (Alex)
   a. Foundational sustainability concepts
   b. Energy proportionality and its implications
   c. Structuring the opportunity space
3. On the road towards Net Zero (Cedric)
   a. Assessing and characterizing progress
   b. How far have we come - driving scale and cost at Internet speed
4. Network sustainability challenges and opportunities (Cedric)
   a. Deployment aspects: power sourcing, HVAC, warm networking
   b. Architecture level
      - Communication patterns and function placement
      - Edge, Core, Compute inside and outside the network
      - Lessons from CDN

Tutorial Outline (contd)

4. Network sustainability challenges and opportunities (contd.) (Alex)
   c. Equipment level
      - Instrumentation and metrics
   d. Network level
      - Overprovisioning and peak shaving
      - Tradeoffs
   e. Protocol level
5. Where are we going (Cedric)
   a. Current initiatives and solution approaches
   b. Pollution-aware and context-variant routing
   c. Research opportunities going forward
   d. Standardization & Community Forums
6. Conclusions (Cedric)

Motivation

What is "Sustainability"?

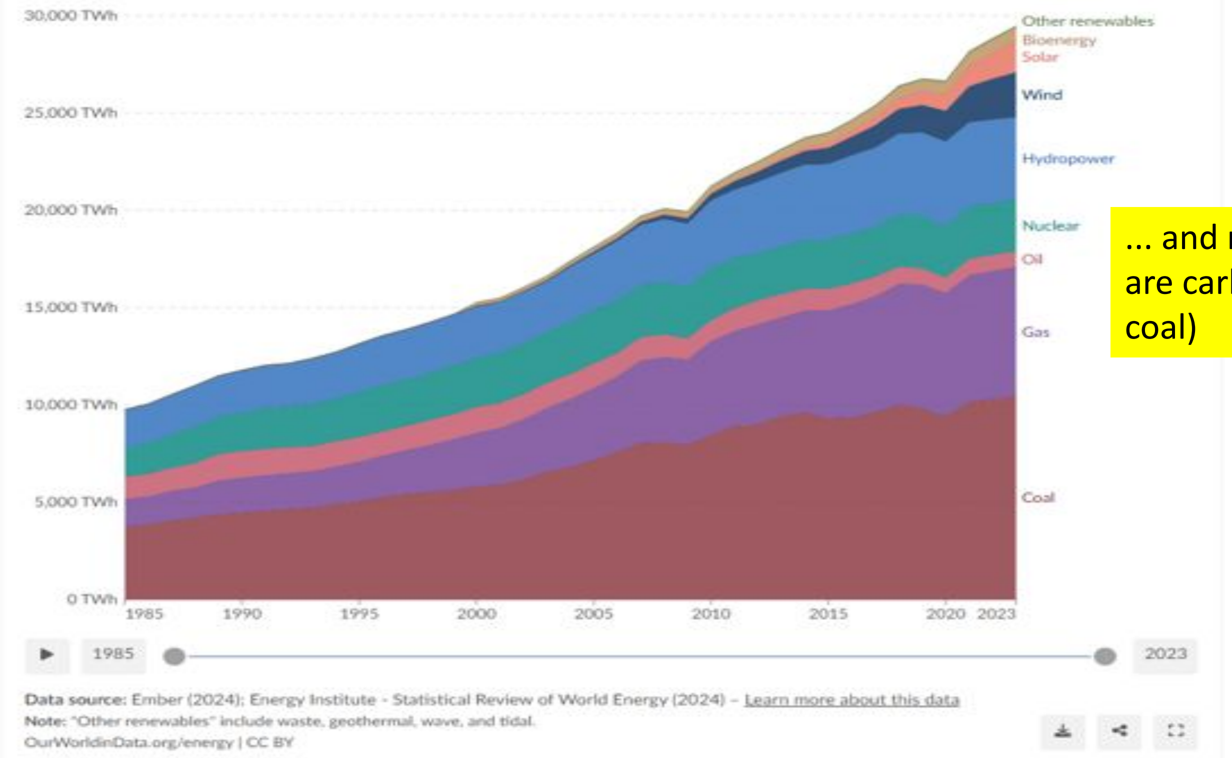
> Meeting the needs of the present
> without compromising the ability of future generations to meet their own needs
> [UN Brundtland Commission, 1987]

Specifically associated with environmental sustainability and "carbon footprint"

- The amount of carbon emitted from burning fossil fuels to generate power

Reducing carbon footprint to Net Zero to combat climate change is considered one of mankind's "Grand Challenges"

Energy demand continues to grow



... and majority of energy sources are carbon-intensive (oil, gas, coal)

Source: https://ourworldindata.org/grapher/electricity-prod-source-stacked

Net Zero

- Refers to the net amount of green house gas emissions
  - Green house gases: CO_2 and other gases (measured in CO_2 equivalents)
  - Net emissions: amount of CO_2 emissions minus amount of CO_2 removed (e.g. absorbed from the atmosphere)
- How to go to Net Zero
  - Use carbon-neutral energy sources (e.g. renewables)
  - Use less energy & become more efficient
    (as renewable sources may not always be available)
  - Avoid emissions in other places

Debatable -
What do you count?
How do you account for it?

Think of Net Zero as an aspiration

Networks as a solution enabler

- Substitute physical mail with e-mail
- Substitute travel with Web meetings
- Teleworking from home
- Replace truck rolls with remote maintenance
- Telemedicine
- IoT solutions
  - Remote surveillance (reducing need for patrols)
  - Smart cities (e.g. smart garbage cans)
  - Smart Agriculture (resource efficiency beyond energy: water, fertilizer, area)

This is great!
So what's the problem?

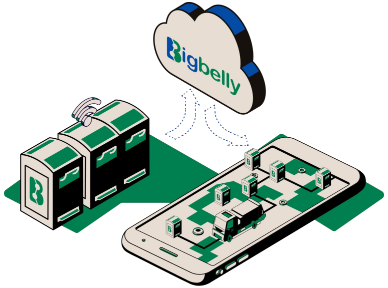
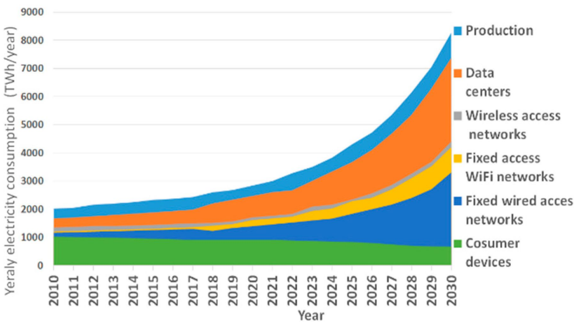


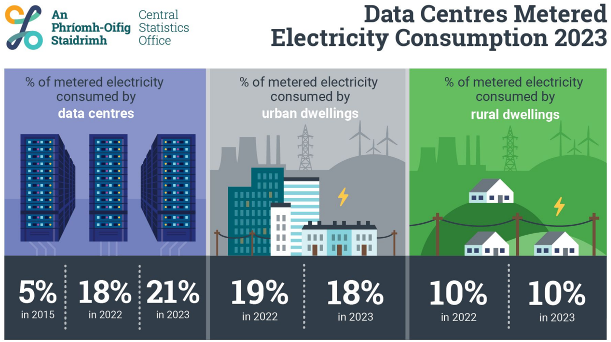
Image credits: John Deere, Bigbelly

ICT and networking share is considerable

- Precise numbers are surprisingly hard to come by
  - Hard to measure, hard to add up
- SMARTer 2030 report: ICT responsible for 2.7% of global emissions [https://unfccc.int/sites/default/files/smarter2030_executive_summary.pdf]
  - 1/3 from communications
- Other reports arrive at a range from 0.5%-1.2%
- Comparable to emissions of medium-sized country
  - e.g. UK: 0.86% [https://ourworldindata.org/co2-emissions]
- As a share this will continue to grow
  - Ireland 2023: DC energy demand surpassing that from urban housing [Reference / image credit: https://www.cso.ie/en/releasesandpublications/ep/p-dcmec/datacentresmeteredelectricityconsumption2023/]



Source: J. Lorincz et al, "Greener, energy-efficient and sustainable networks: state-of-the-art and new trends" Sensors, (2019): 4864.



Motivation (contd)

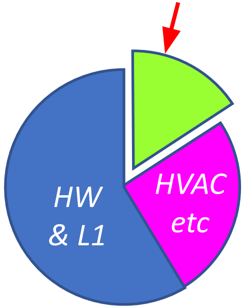
- Networks are a key solution enabler for greater sustainability…
- … but also a problem contributor themselves (even if to a lesser degree)
  - Overall contribution is noticeable, on the order of medium-sized countries
  - Higher power efficiency offset by bandwidth demand
  - Not all substitutions positive – and many inconclusive
    - Compare crypto mining
    - Traffic continues to grow (at a rate of currently 24%/year – roughly doubling every 3 years)
    - Everything everywhere at all times does have a cost
- Even efficiency gains are a double edged sword due to **Jevons paradox**
  - Greater efficiency in use of a resource (and associated falling cost) increases resource use overall
- It's not all bleak – very substantial progress is being made
  - Telefonica report: In 2023, Telefonica's network's energy consumption per PB of data was 41MWh, an 89% reduction since 2015
  - Telefonica achieved reduction of *total* carbon emissions by 51% since 2015, with goal of NetZero in 2040
    (Source: Telefonica Consolidated Annual Report 2023)

How can we do better still?

Advancing Network Sustainability

- Key contributors to network sustainability today
  - General hardware advances (e.g. Moore’s law – but slowing)
    - Greater feeds & speeds typically translates into less energy / bit
  - Deployment factors (e.g. Nordic locations for datacenters)
  - Antenna technology, transmission (e.g. physical layer stuff)
  - Shift towards sustainable power sources (where available)
- What about other network-specific factors (above L1)?
  - How could protocols help reduce carbon footprint?
  - What about control logic – can we build a “carbon-aware network”?
  - How much of an impact can management technology make?

Factors determined by upper networking layers



Not drawn to scale

The focus of our tutorial

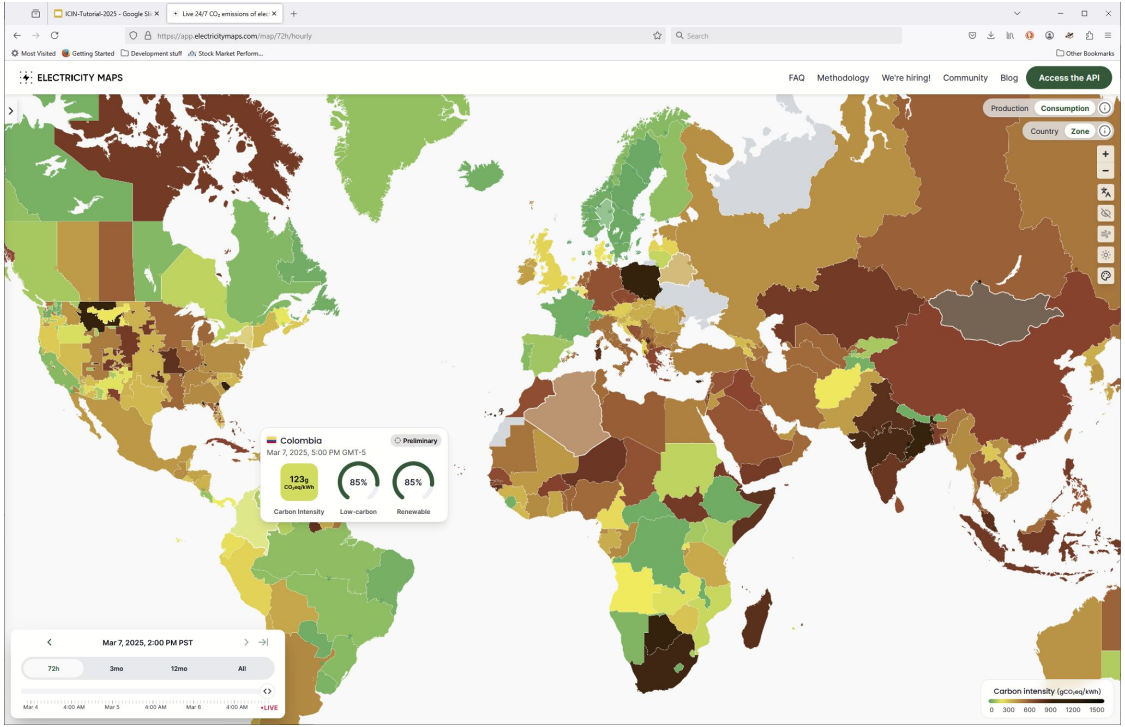
Tutorial Outline

1. Motivation (Alex)
2. Network sustainability foundations (Alex)
   a. Foundational sustainability concepts
   b. Energy proportionality and its implications
   c. Structuring the opportunity space
3. On the road towards Net Zero (Cedric)
   a. Assessing and characterizing progress
   b. How far have we come - driving scale and cost at Internet speed
4. Network sustainability challenges and opportunities (Cedric)
   a. Deployment aspects: power sourcing, HVAC, warm networking
   b. Architecture level
      - Communication patterns and function placement
      - Edge, Core, Compute inside and outside the network
      - Lessons from CDN

Energy-efficiency vs carbon footprint

- Environmental sustainability associated with climate neutrality
  - It’s about greenhouse gas emissions, including (but not limited to) CO^2
  - Carbon equivalence is a measurement term to make greenhouse gas emissions comparable
    - 1 ton of methane has same warming effect as 84 tons of CO^2 □ 84 tons CO^2eq
  - CO^2 is emitted when burning fuels to generate electricity
  - Fossil fuels especially problematic as they release “fossilized” carbon (which had been removed for ages from the atmosphere)
- IT equipment runs on electrical energy
  - If energy mix were 100% renewables, that would not be a problem
  - Generated (at least in part) from burning of fossil fuels

Energy mix



Reference: https://app.electricitymaps.com/zone/

- Varies by
  - Geography
  - Season
  - Time-of-day
- Carbon intensity not the whole story
  - Nuclear
  - Destruction of habitats

Energy vs power vs efficiency

- Power: rate of energy consumption (e.g., kW)
  - 1W = 1J/s
  - Actually, it is not consumed but “drawn”
- Energy: amount of energy (Joule)
  - 1 Wh: one Watt of power delivered for one hour – 3600 J
- Efficiency: amount of energy per unit of value
  - Higher amount of energy used may be acceptable with commensurate amount of value delivered

Energy consumption as proxy for carbon footprint

- Carbon footprint = f(energy consumption, energy mix)
- Energy mix usually a given*, but efficiency in energy use can be influenced...

  ... and a useful indicator of carbon footprint

  ** in some cases, fluctuations in energy mix can be taken advantage of*
  (e.g. TVR, smart workload placement, pollution-aware routing → later)
- Hence, a lot of emphasis given to energy use and energy efficiency
- Often this is associated with energy consumed while equipment is operated/in use

  ... but this is not the entire picture

Energy usage beyond equipment usage

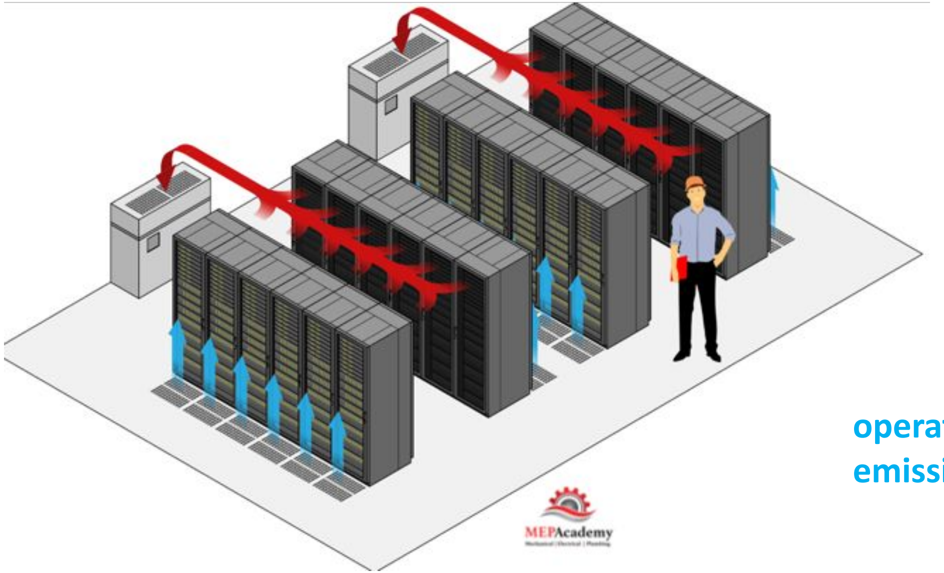


Image: https://mepacademy.com/data-center-hvac-systems/

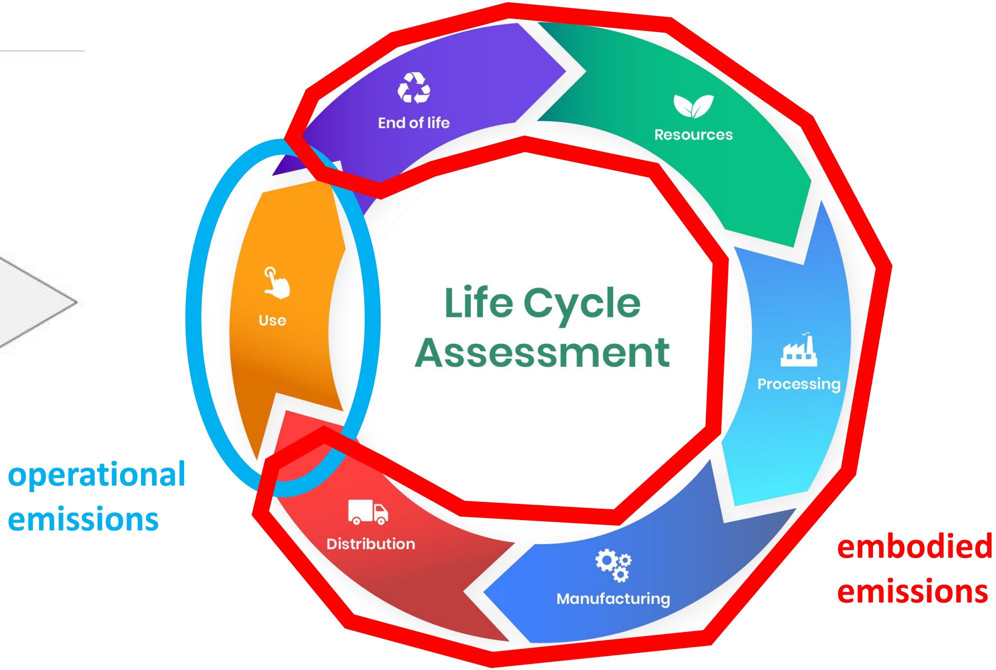
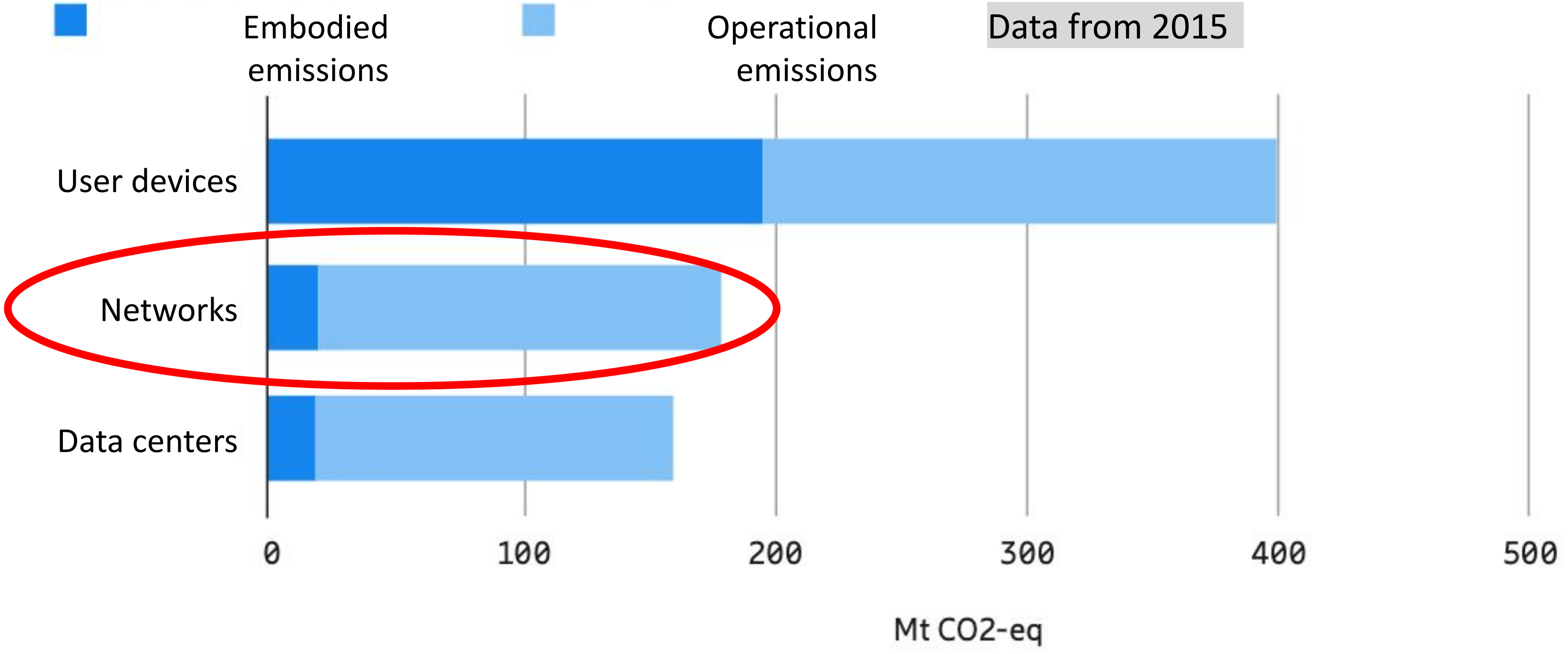


Image: https://www.oneclicklca.com/life-cycle-assessment-explained/

Embodied vs operational emissions



https://www.ericsson.com/en/reports-and-papers/industrylab/reports/a-quick-guide-to-your-digital-carbon-footprint

Power Usage Efficiency

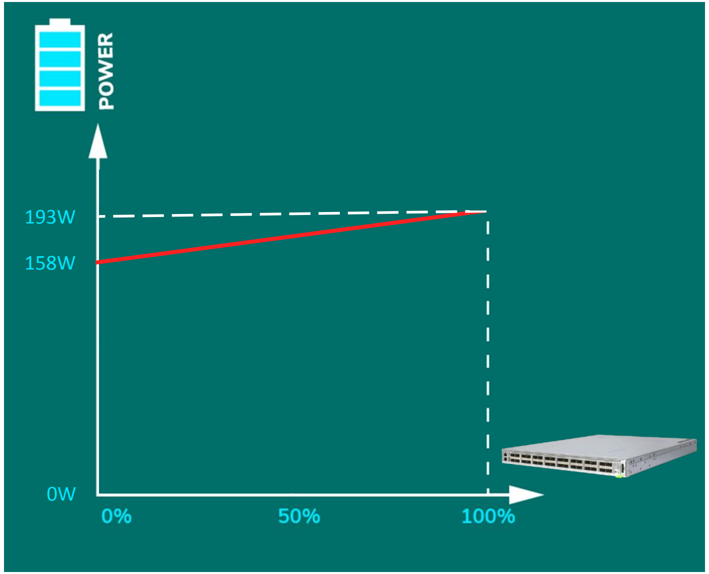
- Network sustainability is in many ways tied to energy usage
- Advancing network sustainability hence needs to ask, how can we minimize energy usage while still having the network deliver what is expected
  - Meeting traffic demands
  - Meeting service level expectations
  - Resilient, Secure, Elastic, ...
  - Cost effective (fortunately, minimizing energy usage helps with that)
- Power Usage Efficiency (PUE): How much energy is used to deliver the value that is ultimately derived from the network

Energy proportionality

- Ideally, energy usage would be proportional to the value delivered
  - Can directly attribute energy use to service demands
  - E.g. transmitting *n* TB of data results in *n* units of energy used
    - Wh/TB
- This would allow to map the problem of how to minimize energy usage to the problem of how to minimize the amount of traffic, e.g.
  - Data compression
  - Smart placement of data and functions (tradeoffs exist – eg compute, need to maintain state)
- These types of solutions have their role (later)
- Unfortunately, the reality is more complex

Idle power

- Power consumption is not linear but incremental
- Most of the power is drawn already when idle
  - CPU, Backplane
  - Transmission ports may be more proportional (eg optical lasers)
- Idle power dominates "productive" power



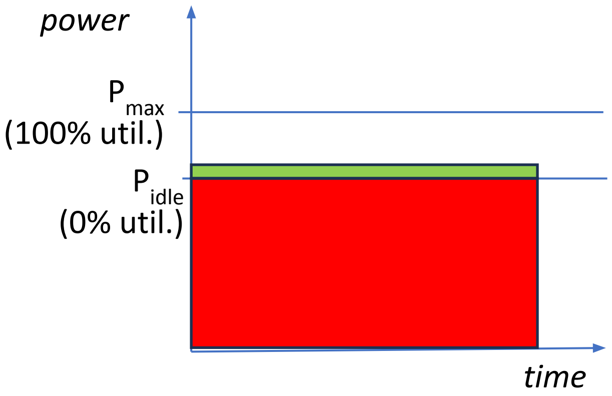
Credit: Romain Jacob, ETH Zurich

Analogy: public transportation

- Say a train or bus has capacity for *50* passengers
- Same energy usage regardless of whether the bus rides empty or at capacity
  - As long as you have capacity, you can take additional passengers at 0 incremental “cost”
- Implication
  - You do not really conserve energy by minimizing the number of passengers
    Caveat: your 51st passenger _will_ increase (double!) the amount of energy (for 2nd bus)
  - You conserve it by minimizing the number of busses
- How much energy usage do you attribute per passenger
  - 1/50 – share of available capacity that is used? (perhaps more “fair”?)
    Or 1/n – share of actually used capacity (perhaps more accurate?)

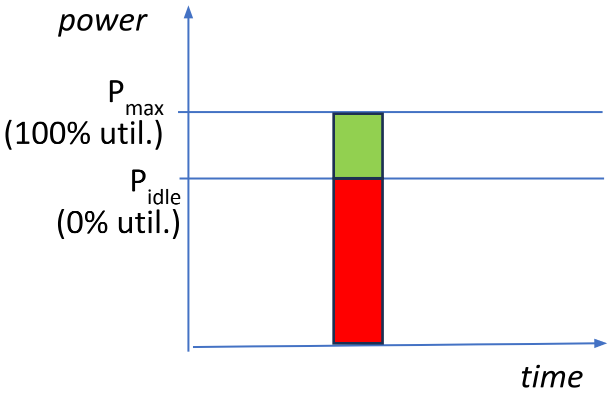
Key observations

- Reducing power that is wasted idling is key
- High resource utilization is good (as it maximizes PUE)
- From a sustainability perspective, there is much more gained in taking a resource offline than in, say, reducing resource utilization
  - Of course there are tradeoffs (e.g. resilience) & embodied carbon does not go away



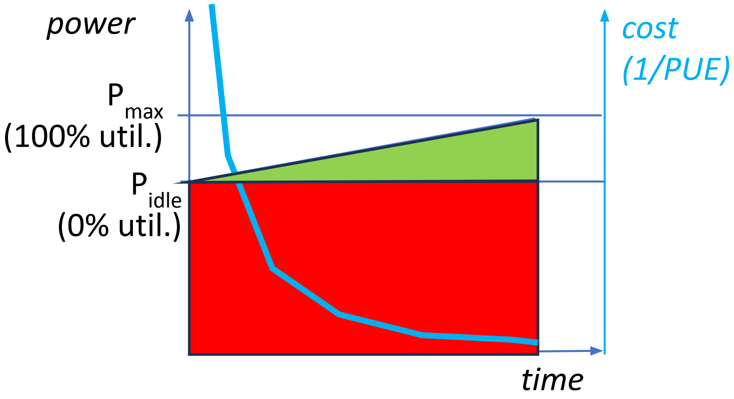
vs.

which is preferable?



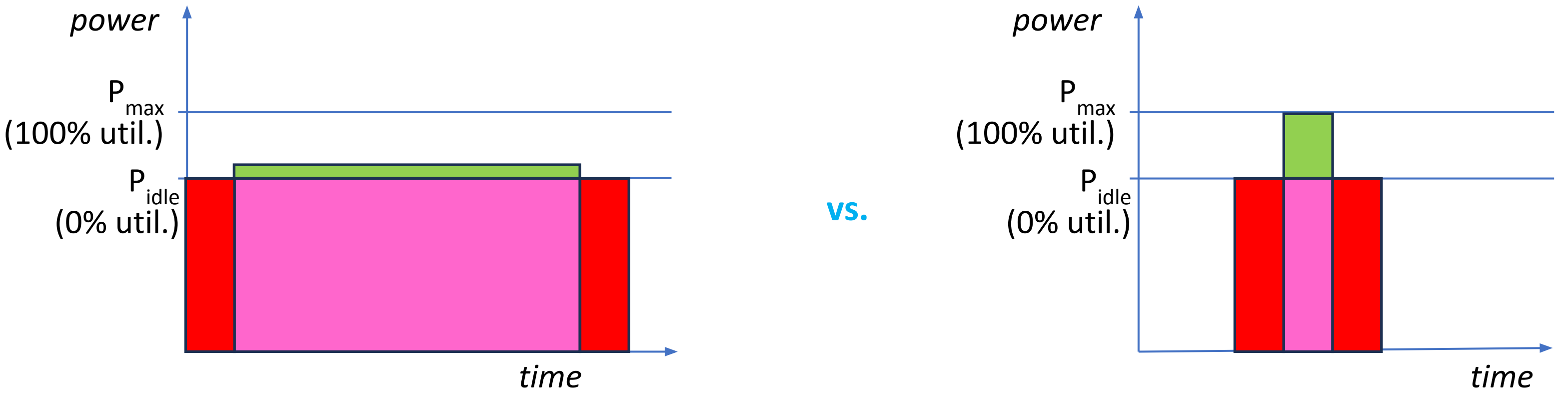
Key observations

- Reducing power that is wasted idling is key
- High resource utilization is good (as it maximizes PUE)
- From a sustainability perspective, there is much more gained in taking a resource offline than in, say, reducing resource utilization
  - Of course there are tradeoffs (e.g. resilience) & embodied carbon does not go away



If it only were so simple

- There is overhead and complexity in turning resources on/off
  - Bootup
  - Synchronization
  - Discovery
  - Convergence and propagation of state (including across the network)
- Worse, during that time the resource is not even available



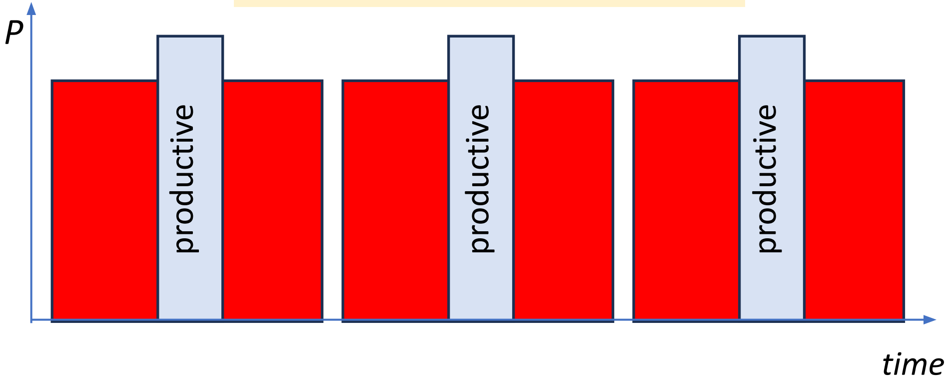
Time scales matter

- Cost/benefit implications

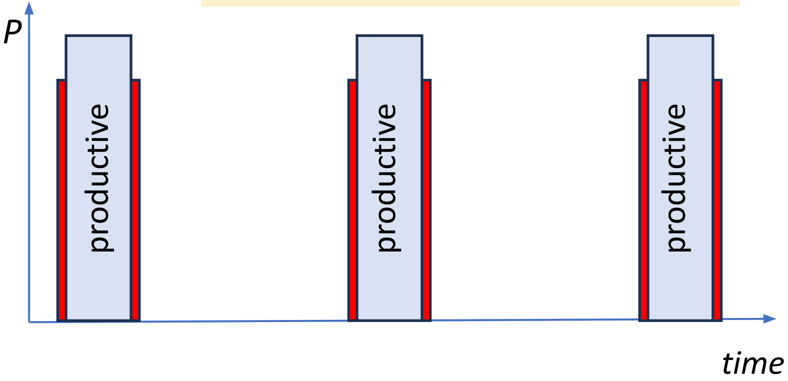
  Any complexity introduced by solutions must be justified by clear gains

- We will discuss this further later

little benefits in savings
large cost in reduced availability
requires long forecasting horizon



large benefits in savings
little cost in reduced availability
short forecasting horizon sufficient



Structuring the opportunity space

- Most obvious (& arguably largest) opportunities at the hardware layer
  - Ultimately, it is here that power is used
  - Hardware technology
  - Transmission technology
  - Power saving modes
    - e.g. WOL, means for downgrading/"rightgrading" of line speeds
  - Extending equipment life cycles
    - SDN plays a role here
    - Less HW upgrades less raw materials less land use greater bio diversity
  - Other sustainability aspects: recycling, warm computing, ...
- Very important, but we are not hardware engineers
  - A bit fatalistic to wait for others to save us
  - Also, Moore's law is coming to an end – it alone will not save us
- What about upper networking layers? Management technology?

Hardware

Sustainable networking opportunity space

Providing of visibility into power usage as a foundational opportunity!

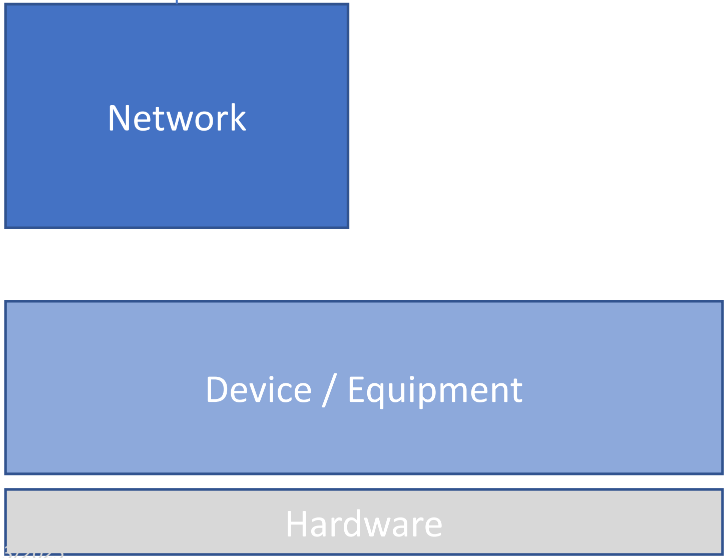
- **Assess usage, validate effectiveness**
- **Enable control loops** for energy/sustainability optimization schemes
- Requires **Instrumentation for energy metrics**
- Selected challenges+opportunities
  - Definition and instrumentation of management data models
  - Certification and compliance assessment methods
  - Virtualized energy and pollution metrics
  - Accounting for energy mix, energy sources
  - Fair carbon footprint attribution to flows & paths

Device / Equipment

Hardware

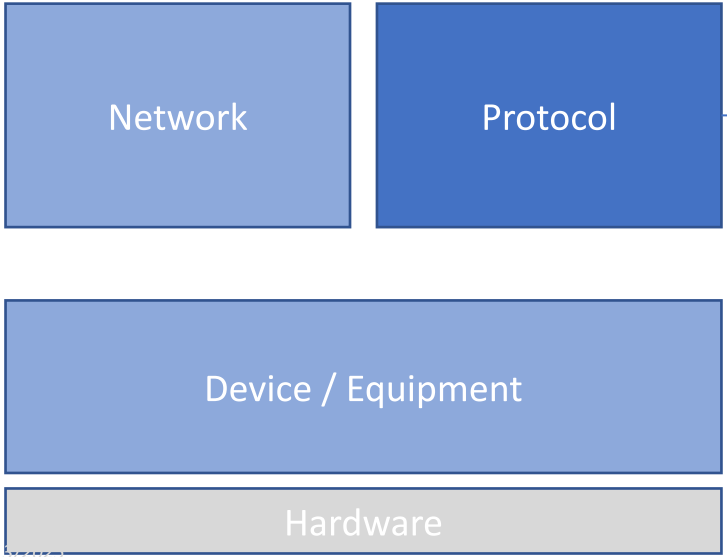
Challenges and Opportunities in Management for Green Networking

One device does not make a network
What about the bigger picture?



- **Network optimization**
  - Energy/carbon/pollution-aware routing & path configuration
  - Deployment / placement of VNFs
  - Optimize carbon footprint while maintaining other goals
  - AI and ML methods
  - Applicability of game-theoretic approaches
  - “Control knobs” for intent-based tradeoffs
- **Energy-related control protocol extensions**
  - Energy as a cost factor – in IGP, SDN controllers
  - Assess carbon intensity of paths,
    optimize networks to minimize overall footprint
- **Carbon-aware traffic steering, pollution-aware routing**
  to steer traffic along greener paths (might vary overtime)
- **Green abstractions**
  taking into account memory, processing, transmission
  can often trade off one versus another

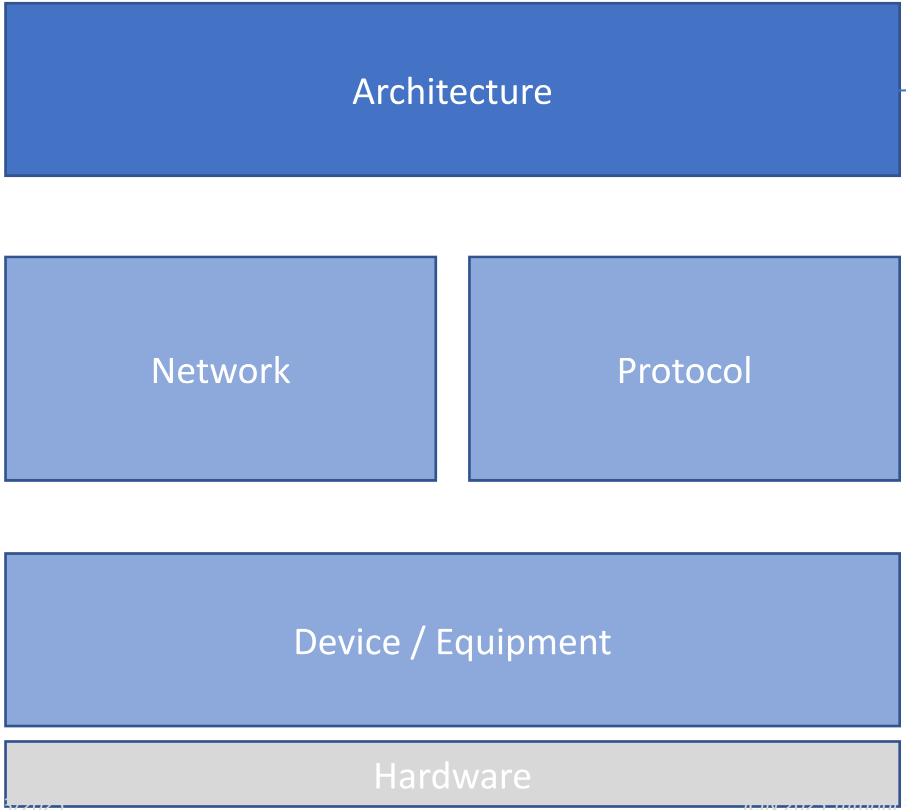
Challenges and Opportunities in Management for Green Networking



How can communication protocols contribute?

- **Protocol enablers for network energy saving mechanisms**
  - Blur mgmt. and control – taking resources on/offline on short time scales
    requires mechanisms for fast discovery, fast state reconvergence
  - Role of autonomics? of IBN?
- **Protocol optimization**
  - Traffic adaptation (e.g. bursty vs smoothened transmission to maximize efficiency;
    control knobs for carbon-aware traffic pacing)
  - Data volume reduction
    (e.g. codings, efficient retransmissions)
- **Instrumentation** (again)
  e.g. energy telemetry at flow & path level

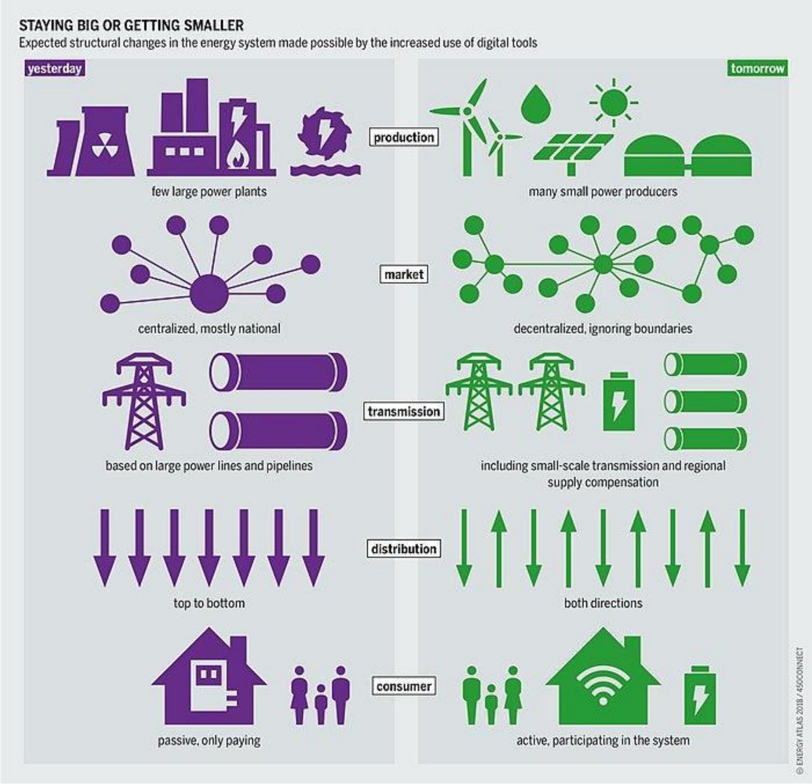
Challenges and Opportunities in Management for Green Networking



- **Facilitate organization of networking applications** to minimize energy consumption
- **Holistic carbon impact assessment methods** for alternative approaches
- **Examples**: retrieval of content, computation placement (compare CDN/ICN/COIN but from energy perspective)

Not considered here:
Networking for greater energy efficiency

SMART GRID



SYNCRO PHASOR NETWORKS

- Power generators need to feed power with correct phase into grid (50/60Hz)
  - Misfunction can cause severe blackouts
- Requires continuous adjustment using PMU (Phaser Measurement Units) over large geographically distributed area
- NASPI.org (North American Syncro Phasor Initiative): Transition to TCP/IP networking
  - more accurate clock and PMU distribution
  - lower latency
  - higher velocity (more PMUs per second)

Tutorial Outline

1. Motivation
2. Network sustainability foundations (Alex)
   a. Foundational sustainability concepts
   b. Energy proportionality and its implications
   c. Structuring the opportunity space
3. On the road towards Net Zero (Cedric)
   a. Assessing and characterizing progress
   b. How far have we come - driving scale and cost at Internet speed
4. Network sustainability challenges and opportunities (Cedric)
   a. Deployment aspects: power sourcing, HVAC, warm networking
   b. Architecture level
      - Communication patterns and function placement
      - Edge, Core, Compute inside and outside the network
      - Lessons from CDN

On the road towards Net Zero…

What are the mile markers?

How do we assess the progress?

How do we even account for consumption?

We organized a Dagstuhl Seminar on this topic and had to come up with a way to organize the discussion

- We picked
  1) applications, systems, and stakeholders;
  2) network technologies;
  3) lifecycle and control loops.

How do we progress towards Net Zero in these dimensions

A Taxonomy to Reason about Internet Energy Consumption (presented by R. Jacob (ETH Z) at that Dagstuhl seminar)

Some important dimensions to be aware of include:

- the particular portions of the overall network
  - which part of the network can, or needs to, be optimized, the core network, the access network, the metro area network, the datacenter.
- the network layer: from L1 to L7
- the class of protocols, inter-domain vs intra-domain

A Taxonomy to Reason about Internet Energy Consumption (inspired by R. Jacob (ETH Z) at that Dagstuhl seminar)

Some important dimensions to be aware of include:

- the target users: hyperscalers, ISPs, entreprises, or end-users
- the agent of change: technological, behavioral or regulatory
- the progress strategy: better proportionality, raw improvement, waste reduction, resource, use optimization, etc.
- the metrics
    - energy usage or carbon emissions,
    - in a relative or absolute measurement
    - primary, secondary, final, and useful energy

A Taxonomy to Reason about Internet Energy Consumption (inspired by R. Jacob (ETH Z) at that Dagstuhl seminar)

Some important dimensions to be aware of include:

- the life cycle phase: operational or including manufacturing and recycling as well
  - (that is, taking into account the embodied carbon)
- the time scale
  - per packet timescale to network planning and deployment timescale
  - bursts, daily, weekly, seasonal variations
- primary vs secondary impacts
- the class of reasoning: attributional or consequential

Attributional vs Consequential

- Attributional represents the emissions that can be attributed to one actor in particular, for instance by dividing the total amount of emissions by the number of participants.
- Consequential represent the activities that can be associated as a consequence for an activity.
- Consequential reasoning weighs the pros and cons of decisions often in terms of total carbon emissions rather than focusing on how to allocate the responsibility to each party.

- A bus has capacity for *50* passengers and uses roughly the same energy usage regardless of whether the bus rides empty or at capacity
  - As long as you have capacity, you can take additional passengers at 0 incremental "cost"
- Attributional: each passenger is attributed #bus/#passenger. If only one passenger at a given time, that passenger may be ascribe a huge energy cost.
- Consequential: how much more energy is consumed by adding one passenger.

Primary vs Secondary Impact

Challenge of quantifying the energy consumption of the Internet is that it has indirect effects on resource use in the overall economy.

- structural changes (where new institutional arrangements and technological capabilities become possible because of the Internet)
- substitution effects (where the new technologies substitute for established energy uses).

Substitution effect is the use of the network to replace an activity that would otherwise consume energy - Alex is doing this tutorial from California, saving on the emissions of a plane trip

Structural changes involve a decline in brick-and-mortar retail stores in favor of warehousing of retail goods for direct home delivery, or in movie theaters in favor of video streaming/Netflix.

Scope for a Net Zero Internet

We need a definition of the proper scope:

Here: Network "Host interface" to "Host interface" (actual network transport)

Does it include the "Cloud" –

- Administratively, not the purview of the network operator
  - somebody else's computers / data-center – compute & storage
- Except it is: integration of the application within the network
  - CDN for instance, but other as well
- Also, there are benefits from taking into account the application
  - Load balancing, but follow-the-sun policies, etc

It is usually estimated that DC use roughly as much energy as the network itself

NOT: End / User devices

- Still, most UE has become very power efficient (due to other incentive, like battery life…)

Scale & Speed

The faster networks become...

The lower the energy consumption/bit

300 bps modem ... 1 Tbps links

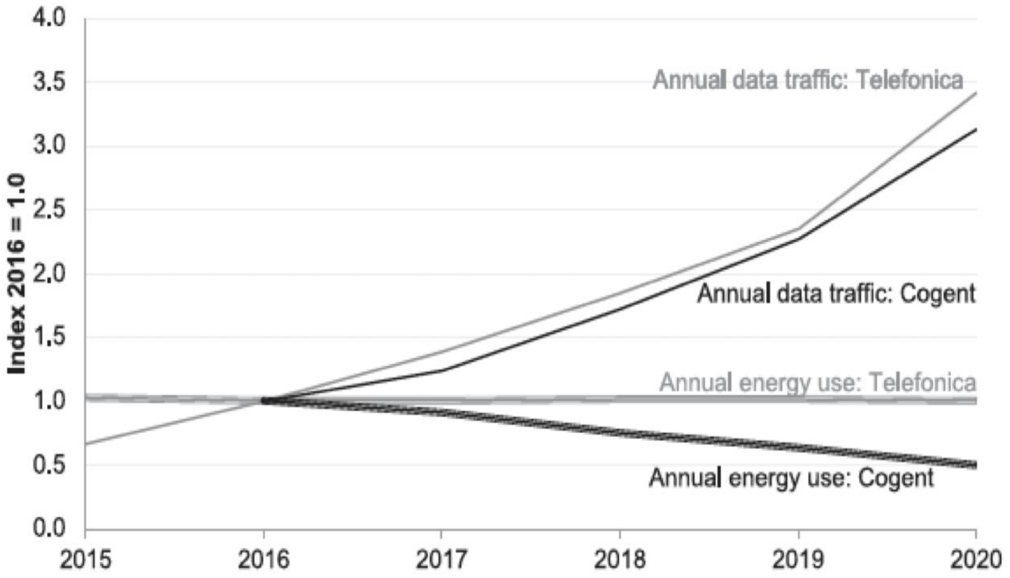


Figure 1. Annual energy use and network data flows for two large network providers, expressed as an index relative to 2016 = 1.0

Over time: Same type of packet processing device roughly maintains same degree of power consumption. Many device power consumption has gone down (PC at home!).

Distance matters:

Electrical / Optical transmission over longer distance dwarfs other power consumption in wide-area networks (packet processing)

Summary #2 How the network did/does save energy

Internet architecture core principles

to support fast growth with lower cost/energy

- Converged networks
  - Datagram multiplexing, IntServ, DiffServ, Re-use of caching
- Federated Networks
  - No central control
- End-to-end principle
  - Addressing, Transport
- End-to-end encryption

Internet Growth + Moore‘s law

allowed for never before seen reduction of energy/bit

- 300bps ... 1Gbps data connection at roughly same energy use: 3 Million times faster

Digitization

- Networked solutions are most often more energy efficient than non-networked Physical transport / production of physical products / people

Tutorial Outline

1. Motivation (Alex)
2. Network sustainability foundations (Alex)
   a. Foundational sustainability concepts
   b. Energy proportionality and its implications
   c. Structuring the opportunity space
3. On the road towards Net Zero (Cedric)
   a. Assessing and characterizing progress
   b. How far have we come - driving scale and cost at Internet speed
4. Network sustainability challenges and opportunities (Cedric)
   a. Deployment aspects: power sourcing, HVAC, warm networking
   b. Architecture level
      - Communication patterns and function placement
      - Edge, Core, Compute inside and outside the network
      - Lessons from CDN

Challenges and Opportunities in Management for Green Networking

Architecture

Network

Protocol

Device / Equipment

Hardware

- **Facilitate organization of networking applications** to minimize energy consumption
- **Holistic carbon impact assessment methods** for alternative approaches
- **Examples**: retrieval of content, computation placement (compare CDN/ICN/COIN but from energy perspective)

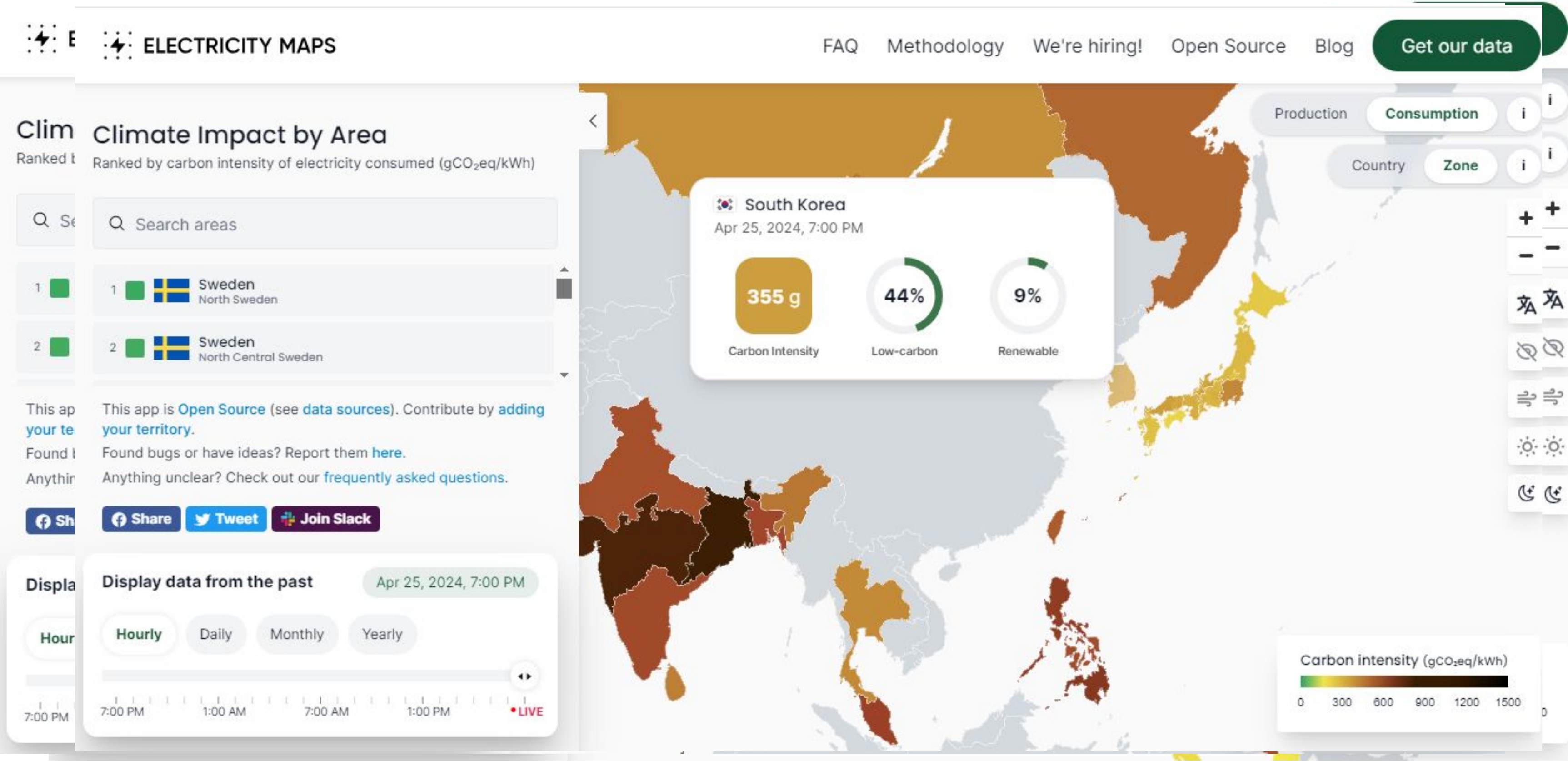
Deployment Aspects

Basic c

Energy

In part

- the
  -
  -
  -



Power source

Spacial/geographic dimension, but al

Energy may be temporarily efficient

- seasonal drivers to renewable (sn
  hydroelectric)
- diurnal patterns for solar/wind

electrek

Exclusives | Autos | Alt. Transport | Autonomy | Energy | Tesla Shop | Store

GREEN ENERGY | ELECTREK GREEN ENERGY BRIEF | EGEB | SOLAR POWER | WIND POWER

California exceeds 100% of energy demand with renewables over a record 30 days

Michelle Lewis | Apr 15 2024 - 10:29 am PT | 206 Comments

"Solar on Palm Desert Walmart" by Walmart Corporate is licensed under CC BY 2.0

In a major clean energy benchmark, wind, solar, and hydro exceeded 100% of demand on California's main grid for 30 of the past 38 days.

Stanford University professor of civil and environmental engineering Mark Z. Jacobson has been tracking California's renewables performance, and he shares his findings on Twitter (X) when the state breaks records. Yesterday he posted:

Deployment Aspects

Energy efficiency is a wh

- the energy produced
  also

DESIGN > POWER AND COOLING

4 Creative Ways to Reuse Excess Data Center Heat

Here are ways to repurpose excess data center heat to reduce the carbon footprint of the facilities.

Christopher Tozzi | Oct 31, 2023

General Terms

Human Factors, Experimentation, Measurement, Reliability

can solve social and political problems which have hindered placing user data away from local administrative boundaries. As the geographical distance decreases, so do the commu-

ENERGY HARVESTING COMMUNICATIONS

Energy Harvesting Small Cell Networks: Feasibility, Deployment, and Operation

Yuyi Mao, Yaming Luo, Jun Zhang, and Khaled B. Letaief

ABSTRACT

Small cell networks have attracted a great deal of attention in recent years due to their potential to meet the exponential growth of mobile data traffic, and the increasing demand for better quality of service and user experience in mobile applications. Nevertheless, wide deployment of small cell networks has not happened yet because of the complexity in the network planning and optimization, as well as the high expenditure involved in deployment and operation. In particular, it is difficult to provide grid power supply to all the small cell base stations in a cost-effective way. Moreover, a dense deployment of small cell base stations, which is needed to meet the capacity and coverage of next generation wireless networks, will increase operators' electricity bills and lead to significant carbon emission. Thus, it is crucial to exploit off-grid and green energy sources to power small cell networks, for which energy harvesting technology is a viable solution. In this article, we conduct a comprehensive study of energy harvesting small cell networks, and investigate

zation manner further helps save operating expenditures [1, 2].

However, as SCBSs are densely and irregularly located, some of them may be inaccessible to the power grid. Moreover, the network power consumption of the SCNs will be high despite the small power consumption of a single SCBS, which will produce a significant amount of carbon emissions. As a result, it is desirable to exploit off-grid and green energy sources to power the SCNs. Energy harvesting (EH) technology is a viable and promising solution, which can harvest ambient renewable energy (e.g., solar and wind energy) to power SCBSs [3]. It is estimated that applying EH techniques to SCNs can achieve a 20 percent CO_2 reduction in the information and communication technology (ICT) industry [4].

Communication networks with EH capability have been extensively studied in recent years, from point-to-point systems [5, 6], two-hop systems [7], multi-user systems [8], to EH heterogeneous networks [9]. However, so far, there has been no systematic study on how to effectively utilize the EH techniques in SCNs, that is, how to

$\epsilon > 0$, while ensuring that the network congestion and the required capacity of the energy storage devices are *deterministically* upper bounded by bounds of size $O(1/\epsilon)$. We then also develop the Modified-ESA algorithm (MESA) to

MDPI

sting Techniques for Low
:s (LPWANs)

atics, University of Siena, 53100 Siena, Italy;
l.: +39-0577-233-702

shed: 3 July 2020

check for updates

(IoT) architectures and applications has been the
gies for the Machine-to-Machine domain. In this
alled Low Power Wide Area Networks (LPWANs),
to satisfy three main system requirements: low cost,
mption. This last requirement is especially crucial
eriods on limited quantities of energy: to cope with
ed every day more frequently, and several different
The aim of this survey paper is to provide a detailed
ing on energy harvesting for their powering. In this
rotocols will be discussed and, for each technology,
e described as well as the architecture of the power

lems. It frees the network devices from having an "always on" energy source and provides a way of operating the network with a potentially infinite lifetime. These two advantages are particularly useful for networks that work autonomously, e.g., wireless sensor networks that perform monitoring tasks in dangerous fields [6], tactical networks [7], or wireless handheld devices that operate over a longer period [8], etc.

Energy Efficiency

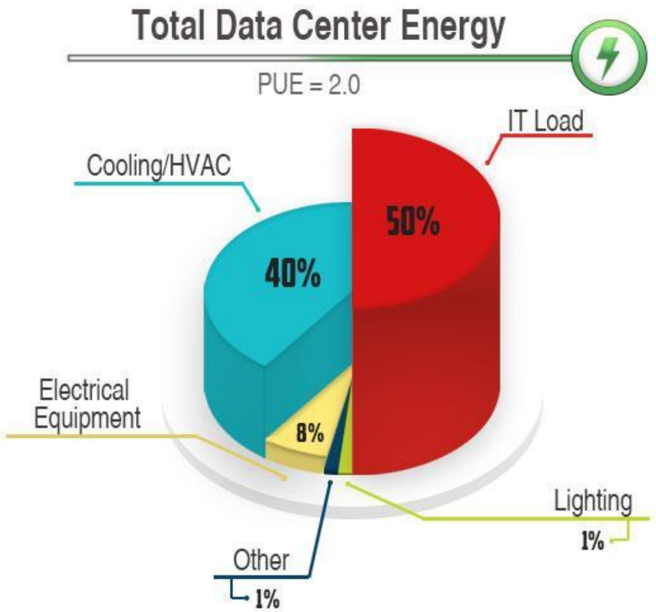
In the data center, energy consumption is a major OpEx.

Energy goes to: computing, but also cooling, lighting, etc.

Key performance indicator:

Power Usage Effectiveness - PUE

Namely, the ratio of *total power* into the data center divided by the *power that goes to the servers/IT equipment*



The

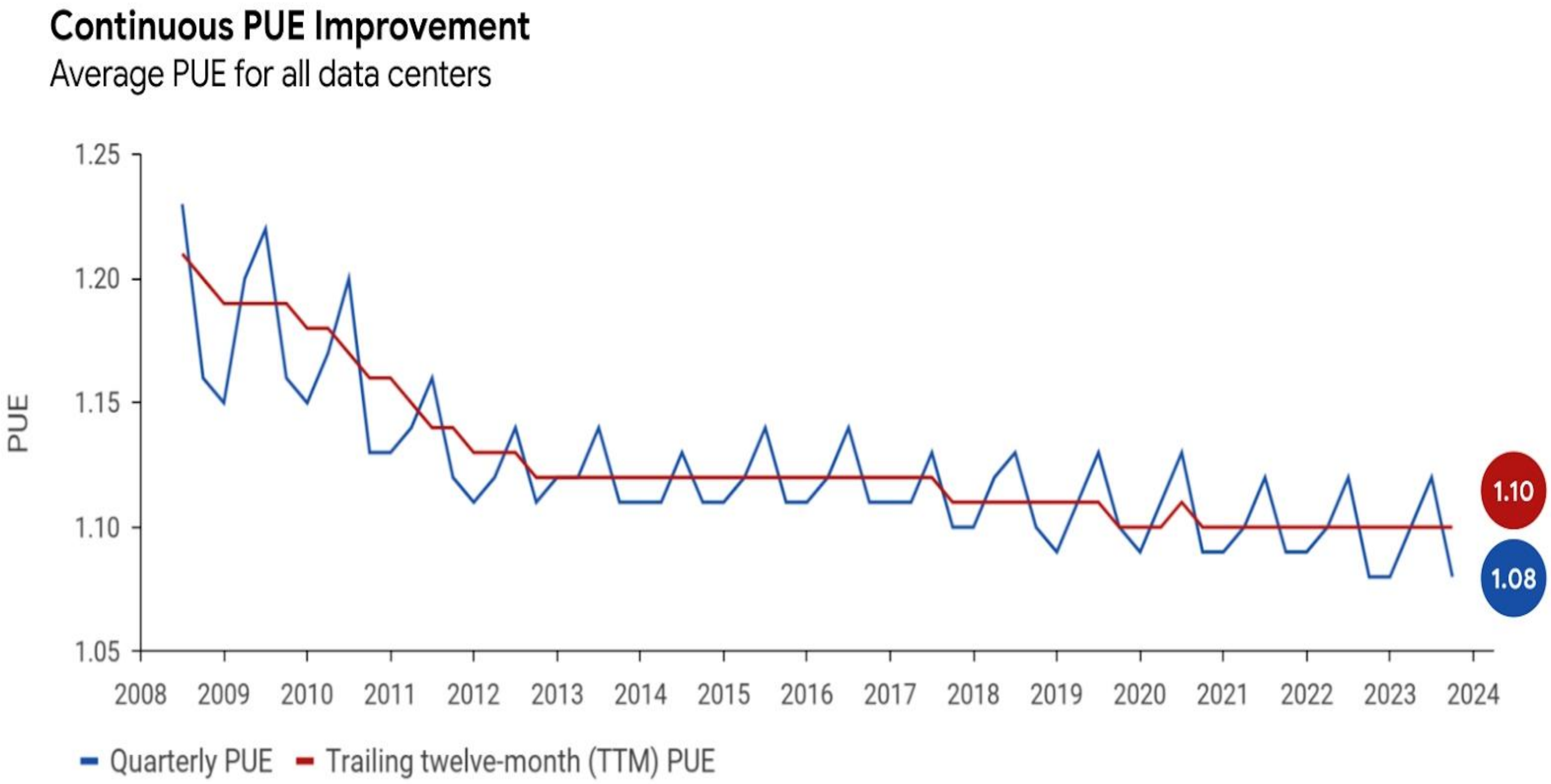
PUE of 1.

PUE very

- very e

What is t

What is t



Efficient Cooling of the DC Infrastructure

Techniques to cool DC:

- Air cooling
- Free cooling
- Hot/cold aisle containment
- Direct evaporative cooling
- Liquid immersion cooling
- Direct-to-chip liquid cooling (- lowest PUE)

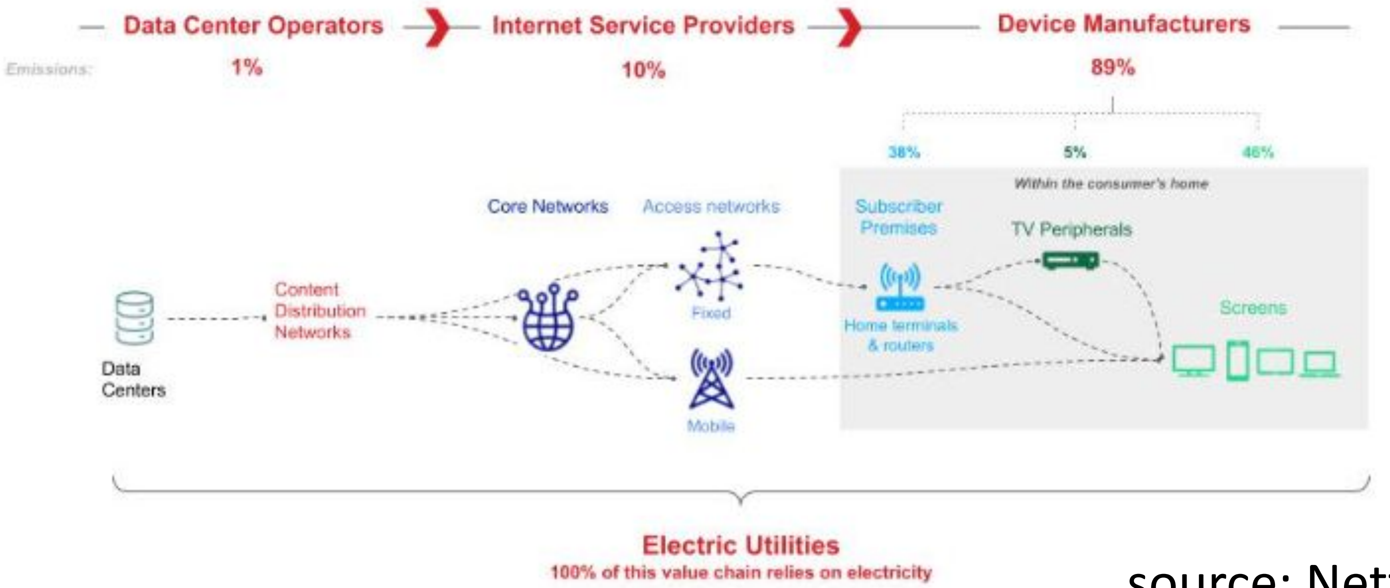
Can the network operate efficiently?

Key questions:

- how far do the similarities extend?
- what about the part of the infrastructure that can't be consolidated in one place?

Architecture Level

CDN, a case study



source: Netflix

Greening of Streaming

iea.org/commentaries/the-carbon-footprint-of-streaming-video-fact-checking-the-hea...

New Chrome available

Another claim estimates that watching a YouTube video uses over 1600W of electricity, equivalent to the consumption of 15 big screen TVs

comparable in scope to the Shift Project analysis – accounted for only 0.36kgCO2e per hour.

However, because the energy efficiency of data centres and networks is improving rapidly – doubling every couple of years – energy use and emissions from streaming today should be substantially lower.

Looking at electricity consumption alone, the original Shift Project figures imply that one hour of Netflix consumes 6.1 kilowatt hours (kWh) of electricity. This is enough to drive a Tesla Model S more than 30km, power an LED lightbulb constantly for a month, or boil a kettle once a day for nearly three months. The corrected figures imply that one hour of Netflix consumes 0.8 kWh.

With 167 million Netflix subscribers watching an average of two hours per day, the corrected Shift Project figures imply that Netflix streaming consumes around 94 terawatt hours (TWh) per year, which is 200 times larger than figures reported by Netflix (0.45TWh in 2019).

Another recent claim on Channel 4 Dispatches estimated that 7bn YouTube views of a 2017 hit song – "Despacito", by Luis Fonsi and Daddy Yankee, featuring Justin Beiber –

ɔtprint of Video Streaming

Keller*, Marco Dinuzzi*, Zhejiayu Ma†

e d'Azur, France
cast, France

r 2, 2023

parameters of the models, e.g. the energy efficiency per GB of the delivery network; (2) We analyse and compare the models under study in a number of scenarios; (3) We question some of the modeling assumptions made using a real deployment of a streaming server in a controlled environment with up to 2000 clients; (4) We propose a technique to reconcile the models and obtain a CO_2 estimate in between 60 and 140 grams when considering the average worldwide electricity efficiency.

2 Models Analysis

Borrowing the life cycle assessment (LCA) vocabulary,

Netflix

Some knobs are customer

help.netflix.com/en/node/87

Back to Help Home

How to control how much data Netflix uses

Watching TV shows or movies on the Netflix app uses varying amounts of data per hour, depending on the video quality. You can adjust your data usage settings by following the steps below.

Web browser

Netflix offers 4 data usage settings.

Data used per hour, per device:

1. **Low**: Basic video quality, up to 0.3 GB
2. **Medium:** Standard video quality, up to 0.7 GB
3. **High**: Best video quality:
   - Standard definition: up to 1 GB
   - High definition: up to 3 GB
   - Ultra high definition (4K): up to 7 GB
4. **Auto**: Adjusts automatically to deliver the highest possible quality, based on your current internet

Related Artic

Internet co recommen

Getting sta

TV show o keeps buff

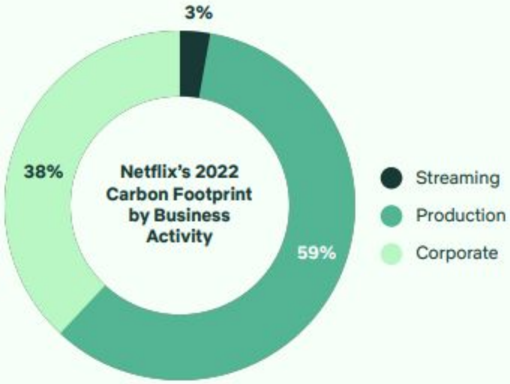
What is Ne

How to ge

Netflix ESG

CARBON FOOTPRINT COMPONENTS

Over half of our emissions (all scopes) are related to the production and licensing of films, series and games ("production"). The second largest source of emissions is in our corporate operations ("corporate"), followed by our data center providers ("streaming").



Netflix ESG Report 2022

Emissions Reductions

TRACKING PROGRESS AGAINST OUR 2030 SCIENCE-BASED TARGET

We track progress against our Scope 1 and 2 emissions target using target-based emissions figures noted in the carbon footprint table above. Target-based emissions only account for emissions reductions resulting from specific decarbonization actions and those related to direct renewable energy supply (e.g., onsite generation, utility and landlord supply, power purchase agreements or direct investments), but not for emissions reductions resulting from our annual purchase of renewable energy certificates (or environmental attribute certificates).

- 2022 Scope 1 and 2 emissions are **12% lower**[15] year-over-year, though they remain slightly higher than our 2019 baseline year due to growth in the business.
- The combined impact of our emissions reductions initiatives across our four priority levers in 2022 resulted in a **30% lower** Scope 1 and 2 footprint value than would have otherwise been reported.
- The vast majority of our emission reductions still come from renewable energy and renewable fuels. Transitioning from diesel to electric will take time to scale to a level that will have a measurable impact on emissions reductions.

[15] In addition to emissions reductions from our decarbonization efforts, external factors also influence the calculated emissions total, including data improvements (e.g., improved data quality, updated emissions factors) and changes in core business activities from year to year (e.g., number of self-managed vs. partner-managed productions which can shift the emissions representation between Scopes).

Netflix ESG

DATA CENTERS AND CONTENT DISTRIBUTION NETWORK (CDN)

Carbon Trust's white paper on the Carbon Impacts of Video Streaming determined that the use-phase emissions associated with data center and CDN operations are small (<1%) compared to the rest of the video streaming value chain.

Netflix partners with Amazon Web Services (AWS) for our data storage and cloud computation needs. AWS has a goal of using 100% renewable energy by 2025. In 2022, AWS reported via their customer carbon footprint tool that our carbon footprint was reduced by approximately **98%** compared to last year. This is due to AWS's renewable energy usage and purchasing, and energy efficiency initiatives.

This trend is consistent with research findings, which show that the energy intensity of global data centers has decreased by 20% annually since 2010. This is a notable improvement compared with recent annual efficiency gains in other major demand sectors (e.g., aviation and industry), which are an order of magnitude lower[20].

When it comes to distributing our content to our members, Netflix invests heavily in making this process as efficient as possible with our Open Connect program. We make 18,000 servers available for free to Internet Service Providers who operate them in their data centers across 6,000 locations in over 175 countries[21]. So when our members press play, instead of the film or series being streamed from halfway around the world, it's streamed from around the corner — increasing efficiency for operators while also ensuring a high-quality, no-lag experience for consumers. On top of that, we've developed encoding technology to reduce file sizes and optimize bandwidth use while maintaining high video quality for consumers. Over the past five years, we have been able to cut our bit rates by roughly **50%**. This is in addition to the significant efficiency gains the telecom industry has achieved within their own networks[21].

We are encouraged by the strong energy efficiency and renewable energy usage trends of both AWS and our own Open Connect program, and are committed to continually decarbonizing these processes over time. Refer to our recent blog post Energy Efficiency in Streaming: Innovation Reaping Rewards to learn more.

Netflix ESG Report 2022

[20] E. Masanet, A. Shehabi, N. Lei, S. Smith, J. Koomey, "Recalibrating global data center energy-use estimates," Science Magazine, 2020
[21] Co-CEO Greg Peters' Keynote Address to the 2023 Mobile World Congress

Architecture Level

From Netflix data:

- CDN is an energy efficient way to deliver the service
- No baseline comparison to “without CDN” scenario
    - Cost on the network would be prohibitive
        - 15% of Internet traffic as it is
    - Who remembers the throttling of Netflix?
- Design of CDN architecture most likely a net energy reduction
- Can it be generalized?

Softwarization/Containerization

Function placement:

- NVF moves functions into containers
- SDN softwarize the control plane

Shift from appliance to software

- increase data center energy use
- reduce client use

But what about the edge?

- is there a scaling effect if servers at edge?

Architecture Level

Key methods to reduce consumption:

- consolidate usage on a platform with high utilization
  - potentially turn off other resources
  - no down time when the idle power consumption is drawn: always churning
- optimize that platform to deliver the service efficiently
  - dedicated hardware acceleration for this platform/service
  - economically feasible to get a more efficient platform since high utilization
- CDN -> reduce network consumption by placing function at the edge and reduce overall network utilization

At the network architecture level, can we use this line of thinking beyond content distribution?

Compute in the Network

- COIN: multiple effor
  - COINRG in the IETF
  - From the charter: "
    comprehend the re
    compute, and stora
- Other efforts in aca
  - For instance, ClickI
    longer considered c
    infrastructure whic
    and throughput) or

ClickINC: In-network Computing as a Service in Heterogeneous Programmable Data-center Networks

Wenquan Xu†, Zijian Zhang†, Yong Feng†, Haoyu Song⋆, Zhikang Chen†,
Wenfei Wu§, Guyue Liu‡, Yinchao Zhang†, Shuxin Liu†, Zerui Tian†, Bin Liu†⋆
†Tsinghua University, ⋆Futurewei, §Peking University, ‡New York University Shanghai

ABSTRACT

In-Network Computing (INC) has found many applications for performance boosts or cost reduction. However, given heterogeneous devices, diverse applications, and multi-path network typologies, it is cumbersome and error-prone for application developers to effectively utilize the available network resources and gain predictable benefits without impeding normal network functions. Previous work is oriented to network operators more than application developers. We develop ClickINC to streamline the INC programming and deployment using a unified and automated workflow. ClickINC provides INC developers a modular programming abstractions, without concerning to the states of the devices and the network topology. We describe the ClickINC framework, model, language, workflow, and corresponding algorithms. Experiments on both an emulator and a prototype system demonstrate its feasibility and benefits.

CCS CONCEPTS

• **Networks** → **In-network processing**; **Programmable networks**; *Programming interfaces.*

servers). Such a paradigm shift, dubbed as *In-Network Computing (INC)*, has benefited many applications (e.g., key-value store [17, 22], machine learning (ML) aggregation [20, 29, 30], consensus [5, 6], coordination [16], and streaming [15]). These applications are typically enabled by the programmable switches (e.g., Tofino [12]) which however is limited by hardware capability and capacity [19], arising a trend to extend on *heterogeneous programmable network devices* [2, 4, 13, 35] (e.g., Tiara [35] achieves a layer-4 load balancer), where the switch is used to perform throughput-intensive task (packet encap/decap) and FPGA is used for memory-intensive task (physical server selection).

While this momentum is inspiring, a closer look reveals a less optimistic reality: the adoption of INC is currently limited to network operators and has not yet to be embraced by application developers, which hinders the development of new applications and their large-scale deployment. The fundamental reason, we believe, is the lack of a high-level programming framework that can abstract away the complexities associated with issues such as device heterogeneity, network topology, and function mapping. Early efforts [8, 11] attempted to improve the programming abstraction by hiding hardware details. Although this is a valuable first step

Compute in the Network

Intuition is that it benefits from the same efficiencies to reduce power consumption

- optimization/reduction of the network traffic
- high utilization of the server (hopefully)

However, content distribution has economies of scale

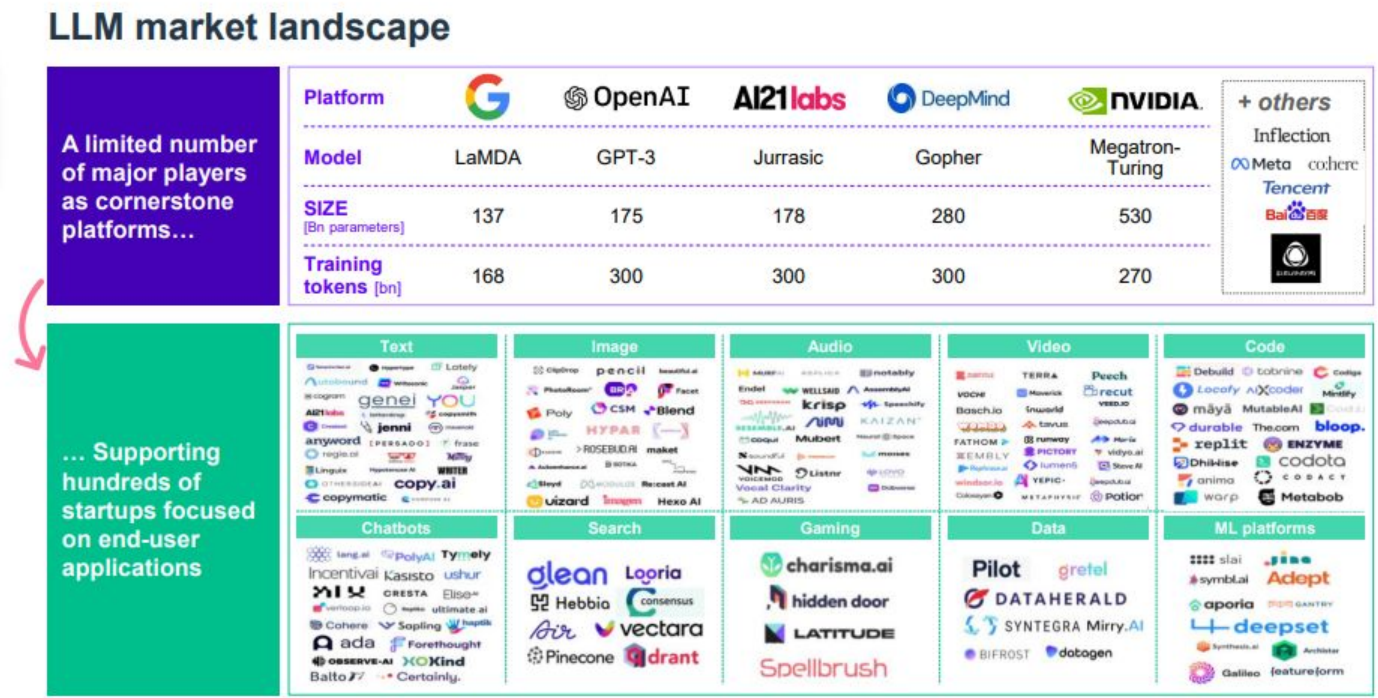
- > 80% of Internet traffic

What in-network-computation service would require the same level of integration with the network?

- no clear candidate at this point;
- ultra-low latency would mostly be the driver, but that is typically aligned with energy consumption

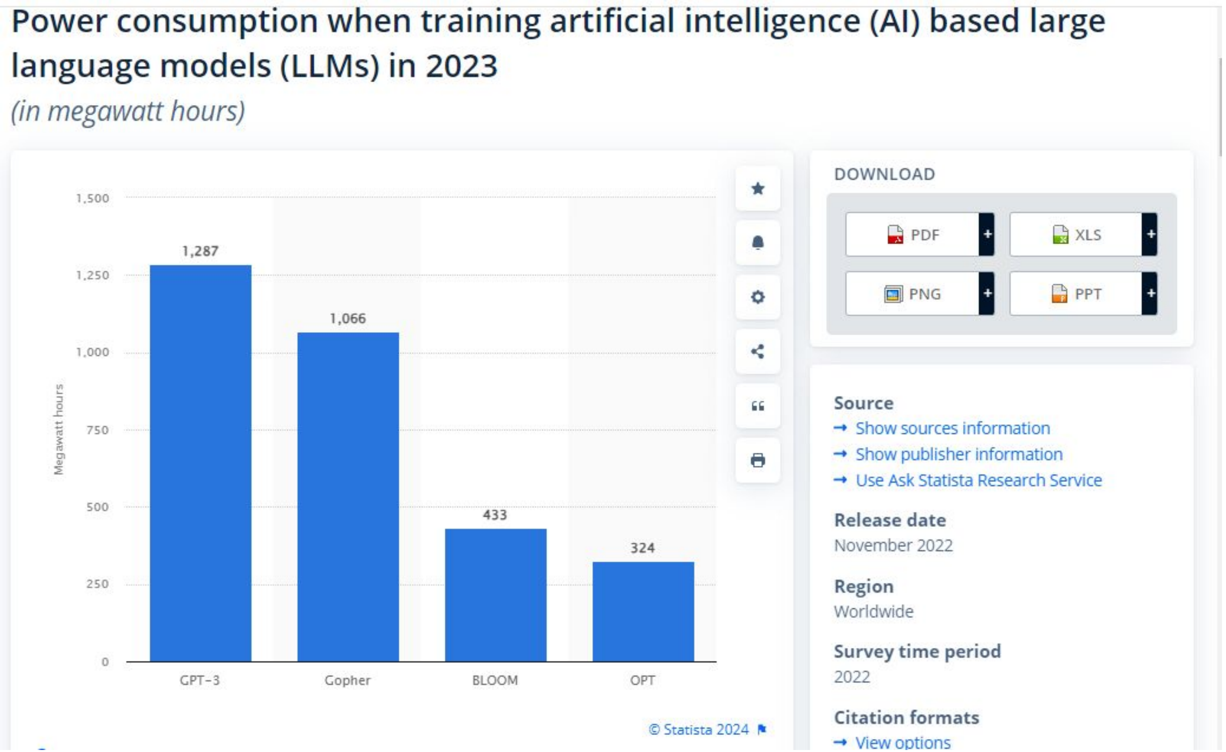
Potential candidate: (federated) AI

AI is replacing bitcoin mining as the energy hog in the Internet



The Energy Cost of AI

AI cost in energy:



Cost of Compute for AI

TECHPOWERUP

Friday, January 19th 2024

Meta Will Acquire 350,000 H100 GPUs Worth More Than 10 Billion US Dollars

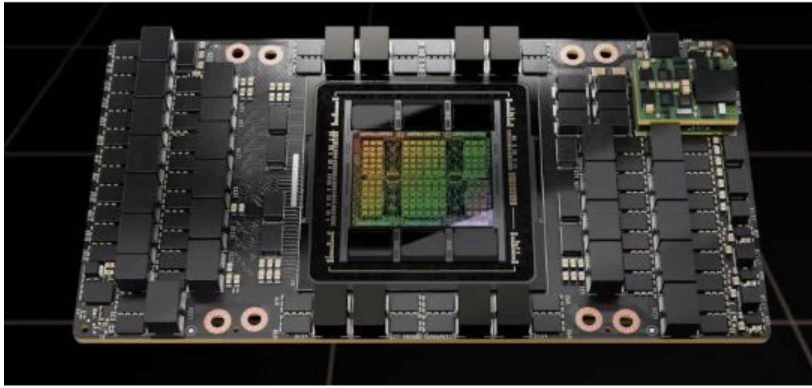
by AleksandarK | Jan 19th, 2024 06:47 | Discuss (53 Comments)

Mark Zuckerberg has shared some interesting insights about Meta's AI infrastructure buildout, which is on track to include an astonishing number of NVIDIA H100 Tensor GPUs. In the post on Instagram, Meta's CEO has noted the following: "We're currently training our next-gen model Llama 3, and we're building massive compute infrastructure to support our future roadmap, including 350k H100s by the end of this year -- and overall almost 600k H100s equivalents of compute if you include other GPUs." That means that the company will enhance its AI infrastructure with 350,000 H100 GPUs on top of the existing GPUs, which is equivalent to 250,000 H100 in terms of computing power, for a total of 600,000 H100-equivalent GPUs.

The raw number of GPUs installed comes at a steep price. With the average selling price of H100 GPU nearing 30,000 US dollars, Meta's investment will settle the company back around $10.5 billion. Other GPUs should be in the infrastructure, but most will comprise the NVIDIA Hopper family. Additionally, Meta is currently training the LLama 3 AI model, which will be much more capable than the existing LLama 2 family and will include better reasoning, coding, and math-solving capabilities. These models will be open-source. Later down the pipeline, as the artificial general intelligence (AGI) comes into play, Zuckerberg has noted that "Our long term vision is to build general intelligence, open source it responsibly, and make it widely available so everyone can benefit." So, expect to see these models in the GitHub repositories in the future.

Paul Churnock, PE • 3rd+ + Follow
Sr. Principal Electrical Engineer - DC Technical Governance...
4mo •

This is Nvidia's H100 GPU. It has a peak power consumption of ~700W. At a 61% annual utilization, it is equivalent to the power consumption of the average American household occupant (based on 2.51 people/household). Nvidia's estimated sales of H100 GPUs is 1.5-2mil H100 GPUs in 2024. Comparing to residential power consumption by city, Nvidia's H100 chips would rank as the 5th largest, just behind Houston, Texas and ahead of Phoenix, Arizona.



1,120 103 comments • 68 reposts

700W is a microwave power consumption
$10B is limited to a few companies

Break

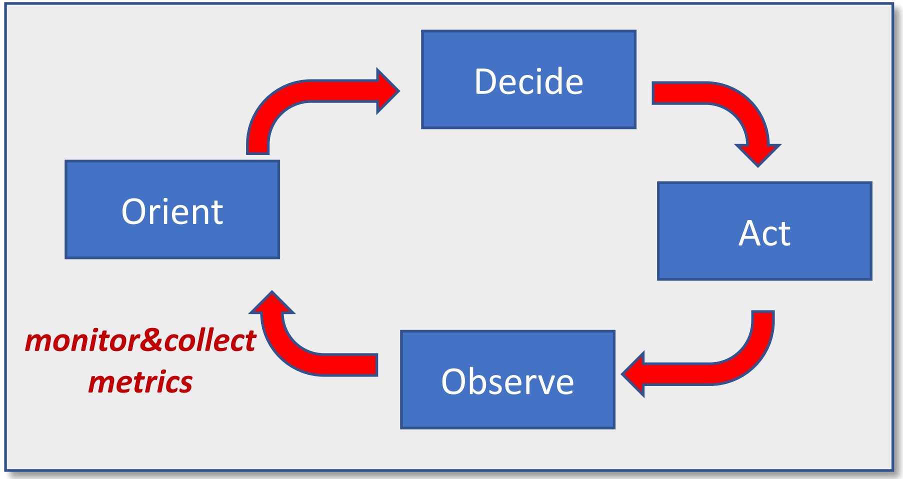
Tutorial Outline (contd)

4. Network sustainability challenges and opportunities (contd.) (Alex)
    c. Equipment level
        - Instrumentation and metrics
    d. Network level
        - Overprovisioning and peak shaving
    e. Protocol level
5. Where are we going (Cedric)
    a. Current initiatives and solution approaches
    b. Pollution-aware and context-variant routing
    c. Research opportunities going forward
    d. Standardization & Community Forums
6. Conclusions (Cedric)

Recap: Aspects where networking can contribute

- Management, deployment, network optimization
  - Provisioning, dimensioning, managing oversubscription, …
  - Energy usage is a great parameter to optimize, just like utilization, cost, etc
  - VM+VNF placement, planning of routes/segments/paths
  - Moderating tradeoffs: carbon intensity versus service levels, utilization versus service levels, caching versus access, etc
- Control
  - Selecting from greener path alternatives, carbon intensity as a cost
- Network architecture (e.g. where to cache from carbon standpoint)
- Protocol design (chattiness, traffic smoothing vs bursting, etc)
- **It starts with visibility**
  - "If you can't measure it, you can't manage it" (Peter Drucker)
  - … or assess effectiveness of solutions, or devise solutions relying on control loops…
- And **visibility starts with the right metrics** – foundation for everything else
  - Actionable and where IETF may be able to make an impact

Metrics

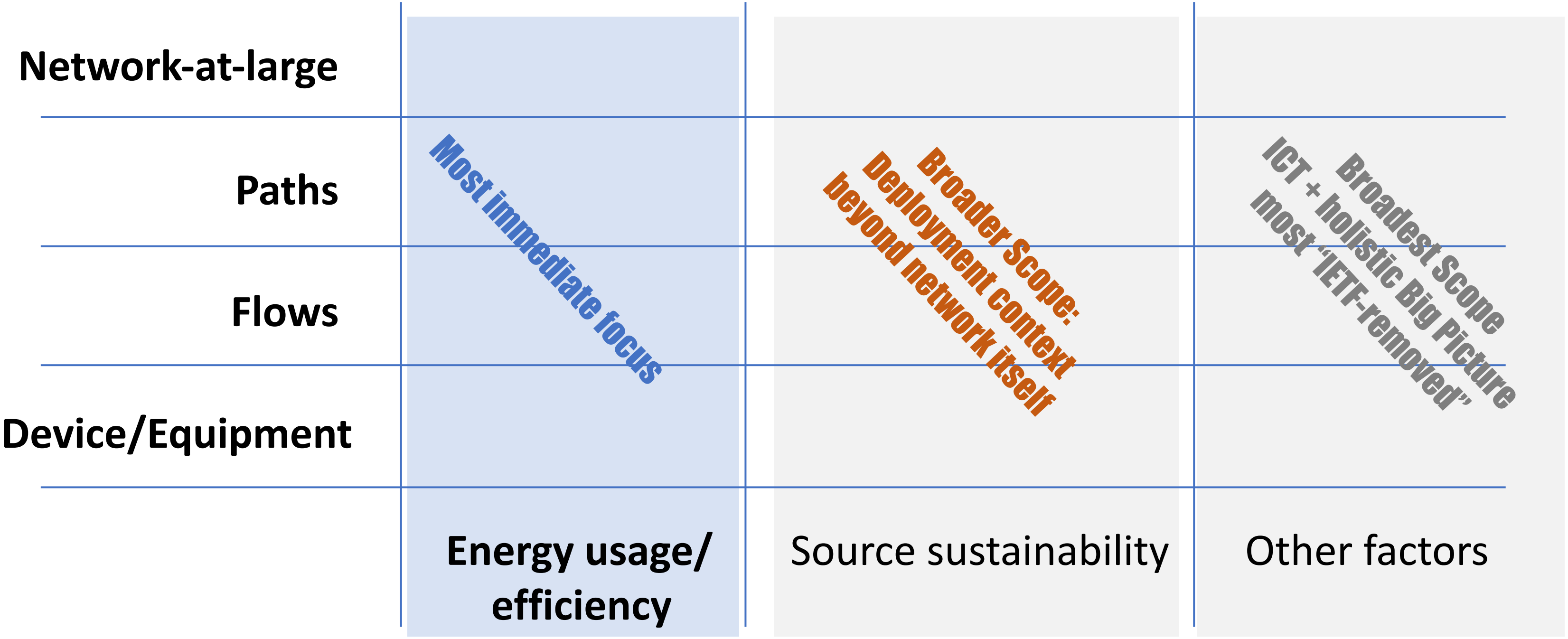


- What metrics are needed
  - To assess effectiveness of a solution?
  - To compare alternative designs?
  - To optimize network deployments?
  - To make better control decisions?
  - To make better management decisions?
- What should metrics cover
  - Energy usage efficiency (scope: network itself)
  - Energy sources (scope: network++)
  - Other aspects: HVAC, manufacturing lifecycle, …
- How can a holistic picture be provided
  - Account for the whole picture, not just a part
  - E.g. device energy usage vs device lifecycle vs deployment environment energy usage
  - Tradeoffs – e.g. edge vs core, storage/memory vs bandwidth

Green Networking Metrics

- https://datatracker.ietf.org/doc/html/draft-cx-opsawg-green-metrics-01
- Internet Draft under auspices of the e-impact program
- Starter set of metrics, augmented with with use cases / applications / motivation

Structuring the metrics space



Disclaimers for what follows:

Not a comprehensive list of metrics, some may be speculative,
some may be less straightforward to instrument than others, usefulness may hinge on particular use cases

Energy usage / efficiency metrics

- Device / Equipment level
  - Device ratings – data sheet stuff
    - Power consumption when idle, at various loads (e.g. 50% utilization, 90% utilization)
    - For subcomponents: chassis, line cards, ports
    - For defined configurations: e.g. memory
    - Does not require instrumentation but should be easily discoverable (e.g. provisioned to retrieve on request)
  - Current energy usage
    - Current power consumption

      for equipment as a whole, for subcomponents – chassis, line cards, ports
    - Power drawn since system start, for the past minute, …
  - Derived metrics in relation to efficiency
    - Current power consumption per unit of traffic – “energy per bit” (e.g. mJ/KB)

      e.g. current power consumption / in-octets rate
    - Incremental power consumption per kilobyte (limited relevance unless energy-proportionality)

Energy usage related to flows

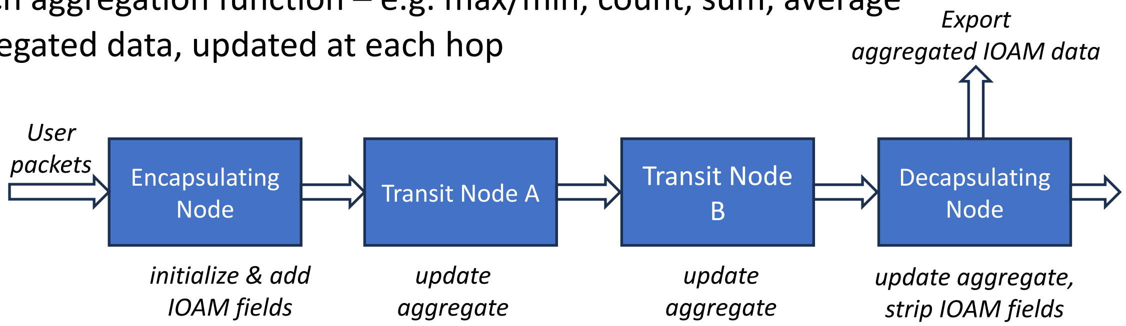
- Energy used by a flow (L2/3)
  - Upper layers: by an instance of a service, by a datastream (L7)
  - May involve correlating telemetry across flows
- Why: may provide a basis for carbon-based accounting
  - Raise awareness of carbon impact of network usage
  - Conceivable: green accounting schemes as an enabler for incentive-based schemes
- Straighforward determination: amortize energy consumed over flow duration
  - E.g. as a share of a flow's traffic
  - Could be determined on a controller
- Alternatively, observe on a device: possible extension for IPFIX
  - Flow statistics today include flow volume, duration, ...
  - Could expand energy consumption metrics
- Issues
  - Should account for packet replication & duplicate elimination
    tradeoff resilience&loss vs energy use
  - Goodput vs throughput (loss, retransmissions)

Energy metrics related to paths

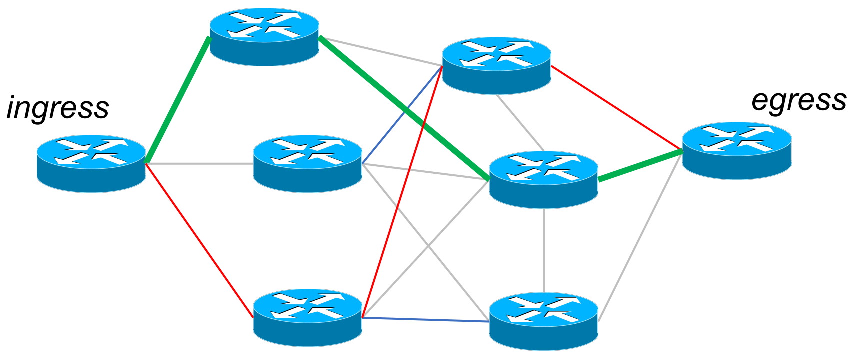
- Assess carbon intensity of paths and route alternatives
- Examples:
  - Path energy/carbon ratings (function of carbon ratings of hops)
    - Path carbon rating might be average of hop carbon ratings – or maximum – or average
    - Ratings might change dynamically (e.g. solar-powered, battery-powered hops)
  - Current PUE across a path (function of hop PUE)
  - Path Energy Traffic Ratio
    https://datatracker.ietf.org/doc/draft-petra-path-energy-api/
- Why
  - Energy-/ Carbon-/ Pollution-Aware Networking
    - Select "greener" path alternative when choices are available
    - May use to wean traffic off less energy-efficient paths
      - PUE decreases with utilization – less traffic to amortize energy use of a hop
      - As PUE deteriorates, even less traffic gets routed across the path
      - As devices get starved for traffic, they may be taken offline when feasible

Obtaining path metrics

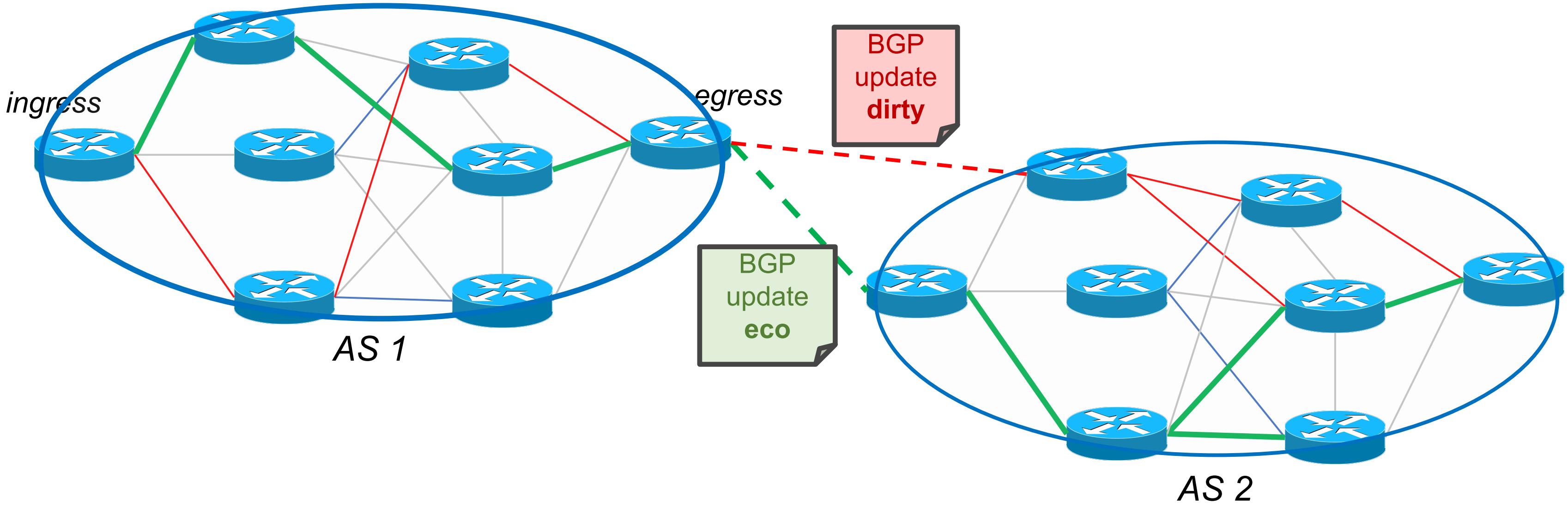
- Option 1: Compute at a controller that monitors device telemetry (current utilization, power usage) and path configurations (provisioned or discovered)
  - Straightfoward
  - Does require a controller
- Option 2: Aggregate data inside the network across path itself
  - IOAM Path Aggregation Trace Option https://datatracker.ietf.org/doc/draft-cxx-ippm-ioamaggr/
  - Basic idea: data plane packets carry IOAM metadata fields
    - Which data item(s) to aggregate
    - Which aggregation function – e.g. max/min, count, sum, average
    - Aggregated data, updated at each hop



Using eco path metrics for cleanest path selection



Using eco path metrics for cleanest path selection



IOAM Aggregation Trace Option - Discussion

- Variation of IOAM and INT
  - IOAM: In-Situ Operations, Administration, Maintenance; INT: Inband Network Telemetry (P4)
  - Basic idea: collect & append data with node data as paths are traversed
  - Processing occurs offline
  - Growing packet size issue (or complexity of correlation in case of postcard telemetry)
- Aggregation Trace Option avoids those issues
  - Data is aggregated during traversal: eg min/max, sum, avg
  - Very simple operations (comparison, addition, increment)

    Aggregate[hop_i] := function (Aggregate[hop_{i-1}], data item)

    Can possibly be performed at line speed (depending on component data)
  - Limited and fixed packet overhead, reduced data volume, greater network intelligence
  - Use cases beyond energy path metrics
    - Identify bottle necks
    - Calculate complete cost (e.g. delay) incurred across the path
    - Take a local action depending on aggregate (e.g., if average exceeds a limit)

Bister, Clemm, Dellsperger, Furrer: Towards Sustainable Networking: Unveiling Energy Efficiency Through Hop and Path Efficiency Indicators. IEEE NetSoft 2024.

Energy metrics related to network as a whole

- At the end of the day, this is what really matters
  - Total energy consumption (MWh)
  - Issues
    - How to measure – e.g. hidden devices
    - What to measure – production network versus context (incl HVAC etc)
  - Network energy efficiency (MWh/PB)
    - Widely used but surprisingly controversial
    - Efficiency = units of energy per units of utility
      - Issue: data transfer volume as a useful measure of utility
        - What about service levels, loss, retransmissions, etc
        - What would be alternative efficiency metrics

Improving network sustainability at the level of the network

- Metrics and instrumentation are an important enabler for sustainability solutions, but by themselves do very little
- Reducing energy use of network as a whole is what really matters
  Network sustainability may be more than the sum of equipment sustainability
- Networks can be optimized for many criteria, carbon is one of them
  - Minimize the number of resources & the amount of energy required to provide service
  - Applying the lessons of energy proportionality:
    Minimize resources needed to provide a service online (longer time scales)
    Apply energy saving modes, downgraded port speeds (shorter time scales)
  - Resource weaning schemes
    Facilitate putting resources in power saving modes (or on/off, where practical), downgrading port speeds
- Important focus of current research next section “Where are we going”

Special case: Mobile Networks

Exploit energy characteristics of transmission media: Dynamic Cell Sizing

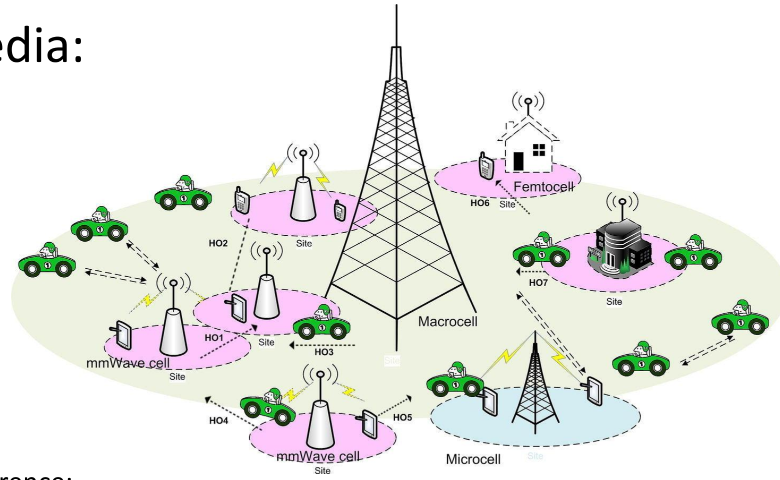
Energy savings by optimizing number of cells, cell sizes

- smaller cells → less energy per cell, but more cells

Measure demand for each cell and determine user locations

- Increase/Reduce power of towers to best match demand
- E.g.: Shrink power of tower at center of big crowd, increase power for adjacent towers so crowd is now served by multiple towers (more bandwidth)
- Scatter/direct spectrum use (when multiple frequencies in use)
- MiMo - send/receive into smaller/larger degrees within a cell

Continuous reoptimization (on the order of minutes)



Reference:
W. Saad et al: Handover and load balancing self-optimization models in 5G mobile networks, Engineering Science and Technology, Vol 42 (June 2023), https://www.sciencedirect.com/science/article/pii/S2215098623000964

Time based network availability / limited performance

Actively switch off 5G, keep just 2G running at night in country side
Not good for teleworker with night shifts...

Networking with constrained nodes

Battery powered / energy harvesting

- Constrained memory: 128KByte ...
- Not typically constrained compute!
  - Example: ESP32 : > 200 MIPS
- Non-constrained radio networks
  - WiFi / bluetooth (not only BLE!)

Delivery Traffic Indication Message (DTIM)

- Allows to only periodically enable radio to receive
- Interval 100msec .. 1sec
- Packets buffered until interval

ESP32 cpu pause: 0.8 mA
ESP32 ULP cpu pause: 0.025 mA

- Low compute, fewer RTT protocols
  highly beneficial for energy use reduction

Wi-Fi/BT Tx packet 0dBm	120mA
Wi-Fi/BT Rx and listening	80-90mA

Slide courtesy of Toerless Eckert

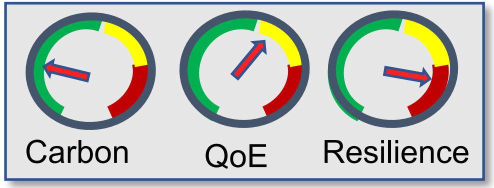
Beyond metrics and instrumentation

- What else can we do?
  - Hardware improvements
  - Power saving modes, port speed downgrades
    lighter version of taking resources offline
    needs to be accompanied by solutions that capitalize on that
  - Backward-compatible software...
    ... and communication protocols
  - Software-defined networking
- Be wary of security considerations
  - Carbon awareness introduces potential new attack surfaces (e.g. to waste energy)
  - Sustainability and net-zero mandates introduce new incentives –
    expect certification and compliance to become of growing importance

How can we reduce rip-and-replace to reduce embodied carbon?

An important caveat

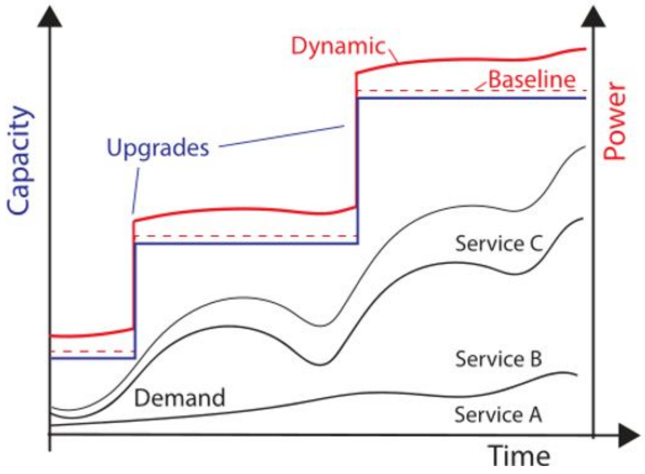
- Sustainability optimization is good, but it needs to occur under constraints
  - Purpose of the network is to provide services
  - Primary operational goals include resilience (against failures), elasticity (to meet spikes in traffic demand), QoE and SLA compliance
  - Clear tradeoffs with sustainability apply – but many of those goals are fairly rigid and non-negotiable



A use case for Intent-Based Networking?

Overprovisioning and dimensioning

- Many operators would not consider letting resources that are deployed go unused
  - Incurring capital expenses for unused equipment does not look to bean counters
  - Avoid potential complications with avoidable churn in the network –
    from control complexity to troubleshooting
- Networks are typically dimensioned for peak demand
  - At the expense of lower PUE during periods of non-peak demand



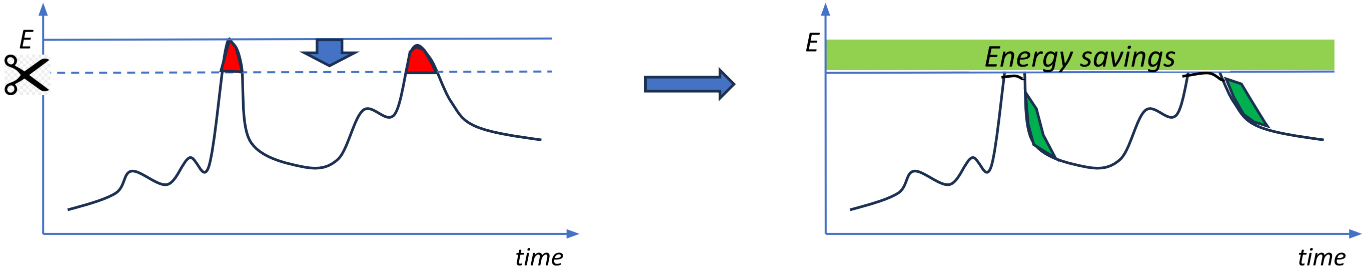
Source: Dan Schien, "Carbon Footprinting of Networks for Digital Services – Factoring in Energy Proportionality". [6]
IAB Workshop on Environmental Impact of Internet Applications and Systems, https://www.iab.org/activities/workshops/e-impact/, 2022 [5]

Bus analogy revisited

- Goal: minimize the number of busses to conserve energy (and cost) while providing acceptable service level
- But you need to run a certain number of busses to maintain customer satisfaction (keep travel times, service intervals acceptable)
  - This may mean running your bus service with excess capacity overprovisioning
  - Wasted (really, under-utilized) energy
- As long as busses have capacity, additional passengers don't matter
  - To the contrary, they increase utilization and revenue
- As buses get filled to near the levels of capacity, action is required
  - Option 1: increase frequency of schedule (and #of busses)
    (or send a new bus when needed – elastic bus service
    but then it may already be too late, and either way you need a spare)
  - Option 2: modify and/or introduce new bus routes

Peak Shaving to the rescue!

- By reducing peak demand, you reduce the capacity at which the network needs to be dimensioned – and along with it, wasted capacity that goes unused



- Compare with “Smart Grid” (energy infrastructure)
  - Many power sources like coal, nuclear are “inelastic” - cannot be rapidly fired up on demand
  - Capacity needs to be able meet peak demand to avoid power outages
  - Techniques
    - Incentives (e.g. peak pricing or rebates to incentivize reduced use during peak periods)
    - Time shifting of demand (e.g. batteries – charge during valleys, use during peaks)

(There are other aspects to Smart Grid: accommodate large numbers of dynamic suppliers, etc)

What are the possible networking techniques for peak shaving of traffic demand?

Tutorial Outline (contd)

4. Network sustainability challenges and opportunities (contd.) (Alex)
   c. Equipment level
      - Instrumentation and metrics
   d. Network level
      - Overprovisioning and peak shaving
   e. Protocol level
5. Where are we going (Cedric)
   a. Current initiatives and solution approaches
   b. Pollution-aware and context-variant routing
   c. Research opportunities going forward
   d. Standardization & Community Forums
6. Conclusions (Cedric)

Energy Efficiency at the Protocol Level

- Protocol optimization (to use less energy)
  - Generally involves awareness of characteristics of deployment context and transmission technology
  - Example: protocols for constrained network environments
  - Other properties: coding, addressing
- Protocol enablers for energy saving mechanisms
  - Management protocols & protocols to interact with instrumentation
  - Mechanisms for fast state reconvergence
- Not considered here: Protocols to manage power network itself
  - Smart Grid
  - Synchro Phaser networks

Traffic adaptation

- Traffic shaping and adaptation
  - Manage bursty vs smoothened transmission to maximize efficiency
  - Low utilization may incur same energy usage as high usage □ bundling transmission into bursts (followed by periods of silence / turning off) makes sense
  - Constrained networks can save battery by turning on/off
  - Bus may wait for more passengers to arrive before departing...

- Practicality depends on deployment context and other requirements
  - Tradeoffs with other goals: latency, loss, ...
  - Smoothening may be desirable too: e.g. avoid bursts to be able to keep links in downspeed energy-saving mode

Chattiness

- Some protocols like to send periodic updates, keep-alives, pings
- This can result in unintended consequences for certain applications and deployment contexts as it may prevent links from going to sleep

credit: https://imageresizer.com/meme-generator/edit/Hey-you-going-to-sleep

Other aspects

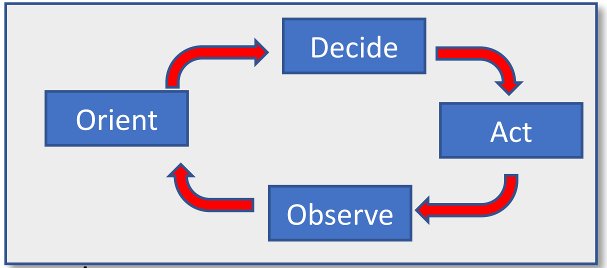
- Encoding
  - More compact encodings result in less traffic volume
  - But does it make a difference in energy efficiency?
  - Not for networks at large!
    - Most networks are underutilized, almost all of the time
    - Added encoding/decoding may be more computationally intensive
  - One notable exception (again): (very) resource-constrained devices
    - Where transmission of every bit counts
    - Driver here not sustainability per se, but e.g. prolonged battery life
- Network addressing
  - TCAM is a non-negligible consumer of energy in a device (& driver of HW cost)
  - Optimizing address allocation may result in smaller routing tables □less TCAM□less energy use
  - Savings potential of 10% or greater for some devices in some cases[1]
  - From a protocol side, reduced address sizes could have some minor impact as well (very speculative)

[1]A. Ghiasian: Impact of TCAM size on power efficiency in a network of OpenFlow Switches
doi:10.1049/iet-net.2019.0125

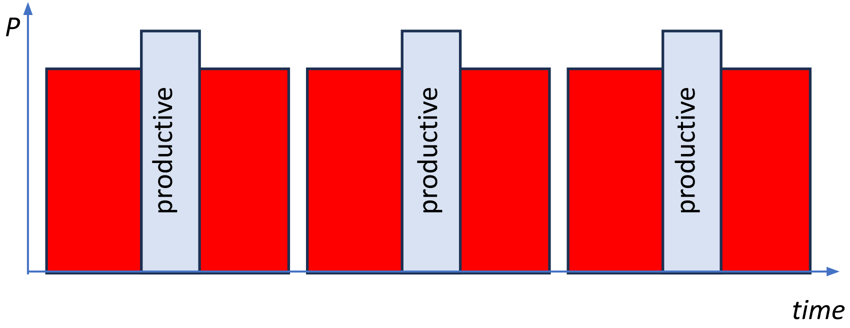
Network protocol standardization activities concerned with energy use

- Protocols define rules for interoperability, which is facilitated by standardization
- Protocol standardization for L2 technologies
  - IEEE: WiFi, bridging, Low-Power Radio short/long-reach
  - Bluetooth SIG, 3GPP
- Protocol standardization for constrained networks (resource-constrained incl. energy)
  - Concluded IETF WGs:
    6LOWPAN/LPWAN (IPv6 over low-power WAN),
    6TiSCH (IPv6oIEEE802.15.4e – low-power industrial grade),
  - Active:
    ROLL (Routing over Low-power and Lossy networks),
    CoRE (constrained ReSTful environments w/limited packet sizes & wake times),
    CBOR (Concise Binary Object Representation), MANET (mobile ad-hoc networks)

Protocol enablers for energy saving mechanisms



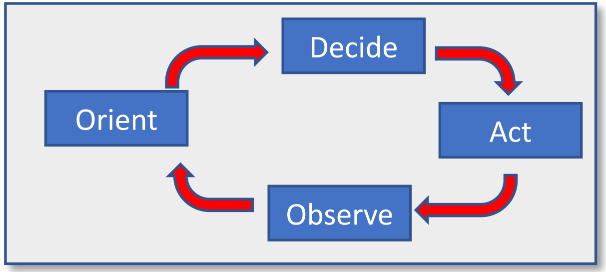
- Many solutions depend on rapid reconfiguration to exploit energy saving opportunities
  - Example 1: Power down (or downspeed) unutilized resources e.g. ports, steer traffic away from underutilized resources so they can be released
  - Example 2: Reconfigure paths and routes to exploit variances in energy mix over a day
- Many of those opportunities are short-lived
- Associated overhead cost
  - Rediscover neighboring resources
  - Propagate updated state
  - Reconvergence
  - RISK!
  - Affects network as a whole (not just device itself)



Protocol enablers

- Improve fast discovery and state reconvergence
  - Lowering time scales allows to capture short-lived opportunities
  - Reducing overhead, complexity, risk improves the cost/benefit equation
  - Move functions from management, to self-management, to control
- Example activities: IETF ANIMA – Autonomous Networking WG
  - High resilience against any type of failures for preserve remote manageability is one core use-case
  - Replace separate “out-of-band” with secure “in-band” management
  - Enable autonomic functions across the network, including discovery and state sharing function
  - Provide reliability for remotely SDN controlled port adjustments
  - “Full Stack” solution requires resilient network OS setup beyond protocols

Management protocols



- Enable rapid automated control loops
- Fulfillment: NETCONF, RESTCONF etc
- Assurance: Collection and processing of metrics
  - On-demand retrieval
  - Bulk export/collection:
    - IPFIX (for homogeneous data)
    - YANG-Push
    - Device telemetry subscription (e.g. YANG-Push and related)
    - In-situ telemetry (e.g. across and congruent with network paths): iOAM, INT
  - + notifications and events

Pretty much all of those are general-purpose, not specific to sustainability
*(unlike data models, such as YANG-defined, which often **are** specific)*

Tutorial Outline (contd)

4. Network sustainability challenges and opportunities (contd.) (Alex)
   c. Equipment level
      - Instrumentation and metrics
   d. Network level
      - Overprovisioning and peak shaving
   e. Protocol level
5. Where are we going (Cedric)
   a. Current initiatives and solution approaches
   b. Pollution-aware and context-variant routing
   c. Research opportunities going forward
   d. Standardization & Community Forums
6. Conclusions (Cedric)

Current Initiatives and Solution Approaches

- Just a few data points of relevant stuff
- Open area of research, many interesting projects

- Key issues: identifying the problem, measurement studies

- For instance: greening of streaming problem
  - how much energy is really spent
  - within that consumption, what are the most egregious culprits?

- Then, operational solutions

- Today: 1) optimization framework for greening the network; 2) pollution-aware routing;

Optimization Framework

Optimization Framework for Green Networking

Cedric Westphal, Alexander Clemm
Futurewei Technologies,
Santa Clara, USA
{cedric.westphal,alex}@futurewei.com

***Abstract*—Reducing energy consumption - and especially carbon emissions - is one of the most important challenges facing humankind. ICT (Information and Communication Technology) is a powerful tool to reduce emissions as it offers alternatives to activities that are costly in energy: video streaming saves energy vs driving to a movie theater, for instance. Still, the carbon footprint of ICT in general and networking in particular have been growing, despite better energy efficiency per transmitted bit, due to the sheer growth in Internet usage and traffic.**

The information and communication technology (ICT) sector is currently estimated to create 2.7% of all global CO2 emissions [1] and expected to continue to increase. Hence, monitoring and reducing the CO2 emissions from ICT is increasingly important. Networks are responsible for around 13% of ICT energy consumption [2], a third of which in turn is attributable to backbone (core) networks [3]. As such, it is important to offer new mechanisms to reduce the energy footprint of networks.

conference, where the main budget is air travel. For the IETF standardization organization meetings, which tabulated this in an effort to reach Net Zero emission [4], the cost of air travel amounts to 99% of the total energy cost of an event.

While ICT saves energy by such substitution effects, networks themselves still spend a lot of energy and a lot of effort goes into making them more efficient. Telefonica [5] reports that in 2021, its network's energy consumption per PB of data added up to 54 MWh. This amount has dramatically decreased by a five-fold factor over the previous five years. However, gains in efficiency are quickly offset by simultaneous growth in data volume. This report states a goal to reduce carbon emissions by 70% over the next five years.

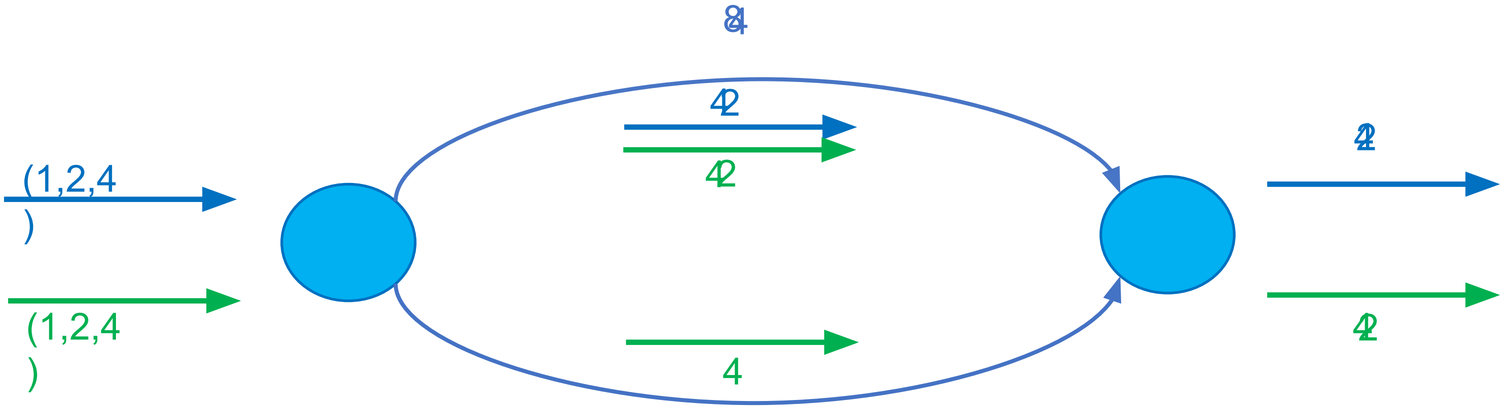
There are two ways to reduce carbon emissions: switching to clean energy sources, which is outside of

Motivation

- Optimizing the energy usage efficiency of networks is important to reduce cost and make networking more sustainable.
- Turning off individual links that are not needed is a way to reduce energy usage. Another way concerns to adapt transmission rates and having end systems adapt the rate and burstiness with which they send their traffic. Turning off links implies rerouting traffic on the on-links and trigger rate adaptation mechanisms, as fewer links imply making less bandwidth available to end system, which causes them to adapt their rates and reduce traffic demand and thereby the energy consumption.
- The problem is how to do so in ways that do not comprise network operations and service constraints, from complying with SLOs to maintaining network resilience.
- Here we define a framework that allows a network controller to dynamically configure the network in such a way that energy consumption is minimized while networking demands are being met.

Strawman Example

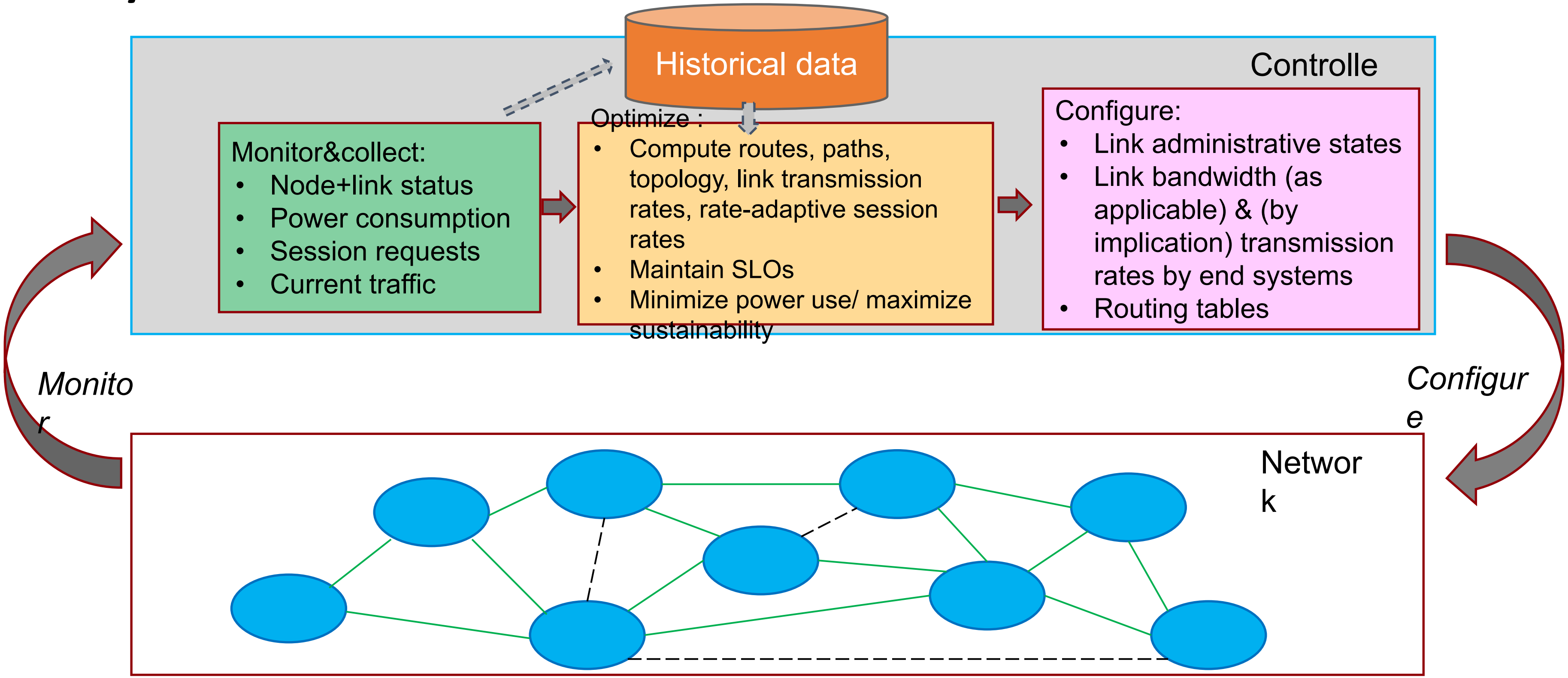
- HTTP adaptive streaming (HAS) such as DASH accounts for 80% of traffic over the Internet



Cut link energy in half, no impact on

Cut link energy in half, trigger adaptation
(e.g. adapt frame rates, color depth, resolution)
QoE impact still acceptable within bounds of SLA)

System Overview



Framework components

- Controller
  - For a given network and a given traffic demand (per a given set of session requests and traffic matrix), computes paths, routes, and topology (links administratively up / down or in various power savings states and at various transmission rates) in a way that energy consumption is minimized, while maintaining the constraint of meeting required SLOs
  - More formally: $\min \Sigma_{j,k} f_{i,j}$ under the constraint: $U(r_k) > U_w$

    where f denotes the energy cost for a link from i to j and U refers to the user utility

    For more details, please view the accompanying paper.
  - Can be applied also to maximize sustainability (can be defined as 1 / power consumption, and include other factors such as weighed sustainability ratings of links that are administratively up)
- Monitoring / Collection Function
  - Collect the required inputs, such as current topology, link and node state, energy consumption, session requests, current traffic
  - May utilize a variety of mechanisms incl SNMP polling, YANG-push, etc
- Provisioning / Configuration Functions
  - Control function that configures administrative link states based on computed topology
  - Routing function that configures routing and forwarding tables based on computed paths
  - Functions utilize a variety of mechanisms incl Netconf, P4, etc.

Refinements

- Controller can take additional inputs, such as historical data, to allow for algorithmic refinements
  - Allow for possibility of reinforcement learning, assessing effectiveness of certain actions
  - Allow for the development and incorporation of refined traffic / demand forecast models
- In addition to energy usage / consumption data, other sustainability metrics can be supported and/or factored in
  - “Sustainability scores” of devices and/or links
  - Discounted energy consumption cost based on cleanliness of power sources used to power a device in a particular deployment.

Optimization Formulation

I. Problem Formulation

A. Network Model

We consider a network graph $G(V, E)$ where the vertices V are connected by edges in E. V is the set of nodes in the network, and $e_{i,j}$ is the link from node i to node j. Each link has a capacity $c_{i,j}$. The link utilization $l_{i,j}$ is equal to the bandwidth allocated to link $e_{i,j}$ divided by the capacity $c_{i,j}$.

We consider an energy model where each link is on, and consumes an amount of energy $\epsilon_{i,j}$. This amount is zero if the link if off. We assume the link is off if it carries no traffic.

Customers (or users) want to exchange data over the network. To do so, they initiate session k with bitrate r_k, where k is the highest achievable vector drown from Q, where Q is the set of possible rate vectors for the sessions. If the session is not elastic, then Q contains only one vector. If the session is for adaptive video streaming (as most of the traffic on the Internet currently is), then Q contains a discrete list of possible rates.

k has a source and a destination within the network graph s and d in V respectively. Under a set of requests from the users, the network needs to allocate the traffic to the different possible paths between each (source,destination) pairs for each user k. $r_{i,j,k}$ is the amount of traffic on link $e_{i,j}$ for session k.

We consider the users derive a utility $U(r_k)$ for session k, where U is the user benefit function that is a positive, concave, positive function. The sum of $U(r_k)$ for all users k is the total user benefit.

We define the energy cost of an allocation. Denote by $\Bbbk$ the indicator function that is equal to 1 if its input is true, and 0 otherwise. Then the energy cost of a link $e_{i,j}$ is $f_{i,j} = \epsilon_{i,j}\Bbbk\{\Sigma_k r_{i,j,k} > 0\}$. We can also equivalently write $f_{i,j} = \epsilon_{i,j}\Bbbk\{l_{i,j} > 0\}$.

We would like to minimize the energy cost while delivering a satisfying experience to the users. This means that we would like to jointly optimize for maximizing $\Sigma_k U(r_k)$ while minimizing $\Sigma_{i,j} f_{i,j}$.

Two possible approaches are therefore:

- jointly minimize
$$-\alpha\Sigma_k U(r_k) + \beta\Sigma_{i,j} f_{i,j} \qquad (1)$$
for two positive coefficients α and β chosen to properly weight the two objectives.
- Set a value for the minimal desired QoE of the users, namely $U(r_k) > U_w$ and then minimize:
$$\min \Sigma_{j,k} f_{i,j} \text{ under the constraint: } U(r_k) > U_w \qquad (2)$$

B. Constraints

Recall that $r_{i,j,k}$ is the rate of session k going through link $e_{i,j}$. The link capacity constraint is therefore:

$$\Sigma_k r_{i,j,k} \leq c_{i,j} \forall e_{i,j} \in E \qquad (3)$$

The flow conservation constraint is:

$$\Sigma_{e_{i,j}\in\Gamma^+(v)} r_{i,j,k} - \Sigma_{e_{i,j}\in\Gamma^-(v)} r_{i,j,k} = 0, \forall k \in K, v \neq s_k, d_k \qquad (4)$$
$$\Sigma_{e_{i,j}\in\Gamma^+(v)} r_{i,j,k} = r_k, \forall k \in K, v = s_k \qquad (5)$$
$$\Sigma_{e_{i,j}\in\Gamma^-(v)} r_{i,j,k} = r_k, \forall k \in K, v = d_k \qquad (6)$$

where $\Gamma^+(v)$ is the set of incoming links to v and $\Gamma^-(v)$ is the set of outgoing links.

The problem becomes:

Problem 1: Minimize (1) under Link Capacity Constraint (3), Flow Conservation Constraints (4)-(6) and Variables $r_k \in Q, r_{i,j,k} \geq 0$.

II. Algorithm Design

Solving Problem 1 is difficult, as the rate vector Q is not linear, and as the energy cost is not concave.

We therefore propose a heuristic solution to a different problem, where we relax some of the requirements of Problem 1.

We can linearize r_k and then select the nearest lower rate within the vector Q.

This is a well defined problem for which an approximation can be found.

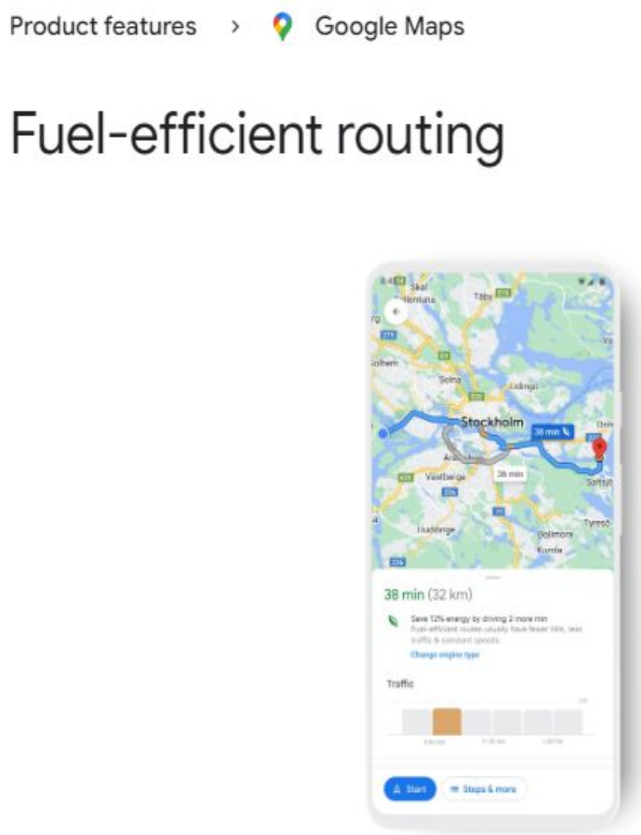
Optimization Framework Lessons

- Green networking is one of the key problems of our time
  - Switching from dirty energy to renewables is one aspect
  - Reducing the energy consumption of the network (including the embedded energy) is another
- We have presented research challenges to reduce the energy consumption
  - At different layers in the network: architecture, protocols, networks, devices
- Here, we put together an optimization framework to try to solve some of the challenges
  - Takes advantage of the elasticity of the demand caused by adaptive streaming applications
  - Need to be further evaluated and improved!

Pollution-Aware Routing

- *Carbon-Aware Global Routing in Path-Aware Networks*

- *Exploring the Benefits of Carbon-Aware Routing*



Carbon-Aware Routing in PANs.

- SCION is a Path-A
  - Can specify a sou
- Define a Carbon-a forecasts and diss a PAN architectur
- So that end point emissions of that
- If end users pick r compete on redu

Carbon-Aware Global Routing in Path-Aware Networks

Seyedali Tabaeiaghdaei
ETH Zürich

Simon Scherrer
ETH Zürich

Jonghoon Kwon
ETH Zürich

Adrian Perrig
ETH Zürich

ABSTRACT

The growing energy consumption of Information and Communication Technology (ICT) has raised concerns about its environmental impact. However, the carbon footprint of data transmission over the Internet has so far received relatively modest attention. This carbon footprint can be reduced by sending traffic over carbon-efficient inter-domain paths. However, challenges in estimating and disseminating carbon intensity of inter-domain paths have prevented carbon-aware path selection from becoming a reality.

In this paper, we take advantage of path-aware network architectures to overcome these challenges. In particular, we design CIRo, a system for forecasting the carbon intensity of inter-domain paths and disseminating them across the Internet. We implement a proof of concept for CIRo on the codebase of the SCION path-aware Internet architecture and test it on the SCIONLab global research testbed. Further, through large-scale simulations, we demonstrate the potential of CIRo for reducing the carbon footprint of endpoints and end domains: With CIRo, half of domain pairs can reduce the carbon intensity of their inter-domain traffic by at least 47%, and 87% of end domains can reduce their carbon footprint of Internet use by at least 50%.

1 INTRODUCTION

Growing concerns regarding climate change encourage companies to measure and reduce their carbon footprint, i.e., the amount of carbon emission that can be attributed to them. This also applies to their use of Information and Communication Technology (ICT), as ICT has a notable contribution of 2.7% to global CO_2 emissions [39], which is expected to grow significantly – approximately four times – until 2030 [3]. Hence, reducing the carbon footprint of ICT use is becoming increasingly relevant for enterprises, manifesting in carbon-neutrality statements of major technology corporations.

While these efforts are laudable and impactful, promising opportunities for further carbon-footprint reduction exist. Indeed, previous research has identified a range of such opportunities. However, most of these proposals apply to local aspects: intra-domain networking (i.e., within a single domain), data-center optimizations, or neighbor-domain cooperation (cf. §8). In contrast, inter-domain networking (i.e., among multiple domains), which accounts for around 13% of total ICT energy consumption, has so far received less attention. An exception is the work by Zilberman et al. [70], who identify carbon-aware networking as a high-potential research area and sketch the concept of carbon-intelligent routing, i.e., to

Evaluating Carbon-Aware Routing

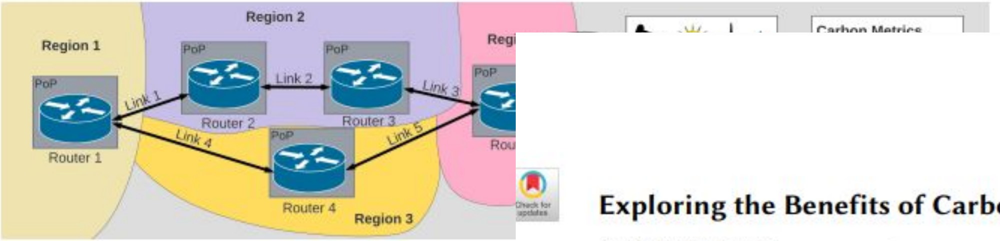
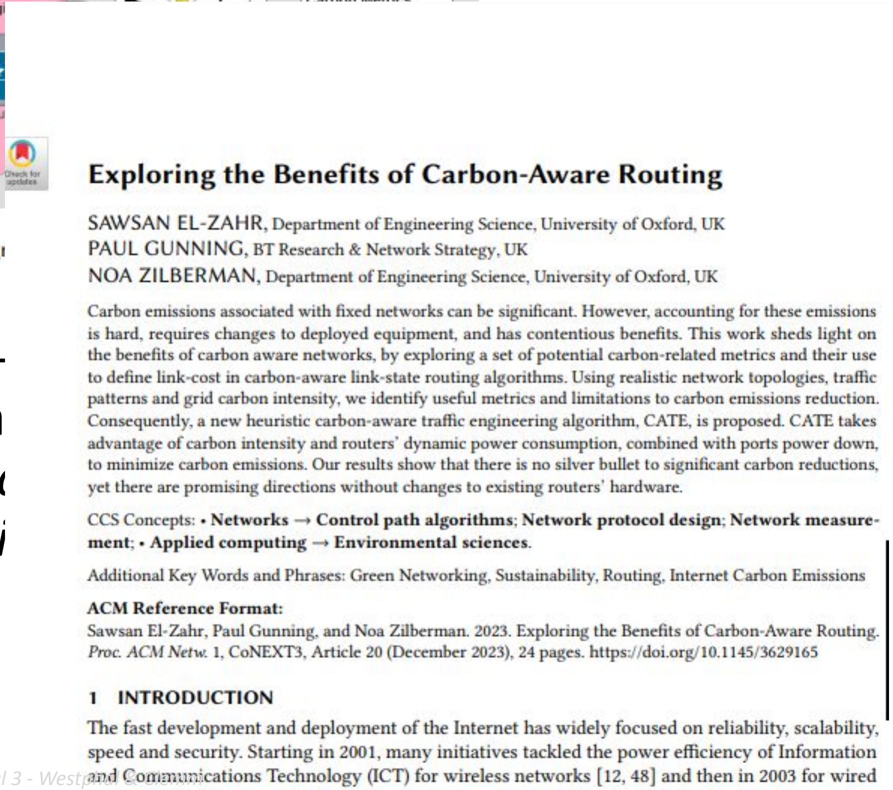


Fig. 3. Overview diag

- Evaluates the benefit of carbon-
  metrics (power, energy label, ca
- “*Routing is one way to reduce c*
  itself. Re-design of network equi
  power is necessary”



Exploring the Benefits of Carbon-Aware Routing

SAWSAN EL-ZAHR, Department of Engineering Science, University of Oxford, UK
PAUL GUNNING, BT Research & Network Strategy, UK
NOA ZILBERMAN, Department of Engineering Science, University of Oxford, UK

Carbon emissions associated with fixed networks can be significant. However, accounting for these emissions is hard, requires changes to deployed equipment, and has contentious benefits. This work sheds light on the benefits of carbon aware networks, by exploring a set of potential carbon-related metrics and their use to define link-cost in carbon-aware link-state routing algorithms. Using realistic network topologies, traffic patterns and grid carbon intensity, we identify useful metrics and limitations to carbon emissions reduction. Consequently, a new heuristic carbon-aware traffic engineering algorithm, CATE, is proposed. CATE takes advantage of carbon intensity and routers' dynamic power consumption, combined with ports power down, to minimize carbon emissions. Our results show that there is no silver bullet to significant carbon reductions, yet there are promising directions without changes to existing routers' hardware.

CCS Concepts: • **Networks → Control path algorithms**; **Network protocol design**; **Network measurement**; • **Applied computing → Environmental sciences**.

Additional Key Words and Phrases: Green Networking, Sustainability, Routing, Internet Carbon Emissions

ACM Reference Format:
Sawsan El-Zahr, Paul Gunning, and Noa Zilberman. 2023. Exploring the Benefits of Carbon-Aware Routing. *Proc. ACM Netw.* 1, CoNEXT3, Article 20 (December 2023), 24 pages. https://doi.org/10.1145/3629165

1 INTRODUCTION

The fast development and deployment of the Internet has widely focused on reliability, scalability, speed and security. Starting in 2001, many initiatives tackled the power efficiency of Information and Communications Technology (ICT) for wireless networks [12, 48] and then in 2003 for wired

Challenges and Opportunities

- We need a PUE for the network
  - Power Usage Effectiveness for the data center: ratio of total energy vs computing energy, to quantify the overhead of the data center
  - We need a metric that states how efficient the network is – how much it costs to transmit a bit over a network
  - In energy and in carbon footprint
- Trade-offs are not obvious due to opacity of the system

Challenges and Opportunities

What is needed:

- 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, bringing 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 footprint and holistic approaches to trade off carbon footprint between forwarding, storage, and computation.

Some steps in the IETF

- IAB Workshop on environmental impact of networking in December ‘22
  - See description of the workshop and material at https://www.iab.org/activities/workshops/e-impact/
  - Engagement from academia, operators, equipment manufacturers, and activists
  - Climate change as a social issue goes beyond the network
- Formation of E-Impact Program
- Formation of GREEN WG (“Getting Ready for Energy-Efficient Networking”) in mid-2024 (IETF 120 Vancouver)
  - Develop YANG Data Models for measuring energy usage
  - Develop a monitoring framework
  - Articulate use cases
- Formation of SUSTAIN RG in 2025 (first meeting at IETF 122 Bangkok next week)
- We co-authored several drafts:
  - Challenges and Opportunities in Green Networking, https://datatracker.ietf.org/doc/draft-irtf-nmrg-green-ps/
  - Green Networking Metrics, https://datatracker.ietf.org/doc/draft-cx-green-green-metrics/
  - Architectural Considerations for Environmentally Sustainable Internet Technology, https://datatracker.ietf.org/doc/draft-various-eimpact-arch-considerations/
  - Sustainability Considerations for Networking Protocols and Applications, https://datatracker.ietf.org/doc/draft-pignataro-enviro-sustainability-consid/

Other Forums

Research:

HotCarbon workshop https://hotcarbon.org/ *(next edition: July 10-11)*

e-Energy conference https://energy.acm.org/conferences/eenergy/2025/ *(next edition: June 17)*

GreenNet conference https://sites.google.com/view/greennet2025 *(next edition: June 8 or 12 at ICC)*

IEEE CNOM SIG on Sustainable Network Operations: SNO https://cnom.committees.comsoc.org/sustainable-network-operations-sno/

3GPP: https://www.3gpp.org/technologies/energy-efficiency-ee-sa5-work-and-results

ITU: https://www.itu.int/en/action/environment-and-climate-change/Pages/energy-efficiency.aspx

In conclusion

- Greener, more sustainable networks are an important problem
- Plenty of challenges and opportunities beyond hardware advances and improved energy mix for research, standardization, development
- Energy proportionality (or lack thereof) dictates many solutions
- Advances will involve multiple vectors, not one grand solution
    - HW advances drive energy per bit & scaling,
but there is also an important role for management & L2+above
    - Visibility and metrics, peak shaving as important techniques
- Activities gaining traction in IETF
- A call to action for researchers to get involved

Questions?

Please do not hesitate to contact us

Cedric Westphal cedric@ucsc.edu
Alexander Clemm ludwig@clemm.org

Thank you!