

Sustainability and Architecture

Thoughts from design team gatherings
in the e-impact program

Jari Arkko, Alexander Clemm, Luis M. Contreras, Hesham ElBakoury, Ari Keränen,
Suresh Krishnan, Jan Lindblad, Marisol Palmero, Carlos Pignataro, Jukka Manner,
Ali Rezaki, Sebastien Rumley, Eve Schooler, Emile Stephan, Michael Welzl

draft-various-eimpact-arch-considerations-00

Focus for this work

Environmental sustainability is a cross-cutting issue that affects everything, e.g., from economics to user behaviour, from applications to protocols to implementations and chip design, and from manufacturing to operations and recycling.

We'd like to focus on a much narrower question:

“Are there guidelines that can be helpful for protocol designers and network architects?”

The focus of this document is to try to provide actionable design advice to network and protocol designers. Matters outside the network such as applications are out of scope.

What is architecture?

Several possible answers:

- Network architecture diagrams to achieve a particular function (e.g., to collect energy consumption metrics – architectures and frameworks are being discussed in GREEN WG)
- Fundamentals that need to be considered (e.g., multi-objective optimization, tradeoffs between resilience and various costs)
- Design patterns that can help build sustainable protocols and networks (e.g., what do you need to build dynamic scaling)

Example design patterns

- Measurement & modeling
- Dynamic scaling as a way to reduce number of devices that are on
 - Also needs information, closed loops, some level of automation, and ways to reduce impacts of changes
- Transport as a way to enable increased device sleep times
- Equipment longevity as a means to amortize manufacturing costs
- Compact encoding to reduce message sizes

Measurement & modeling

Understanding where you are is a basis for any improvements
Ongoing work at the IETF (GREEN WG)

Some architectural aspects to consider:

- Need an architecture for collecting data, aggregating, etc.
- Cost vs. benefit and 'dark data' situations
- Build systems that generalize beyond the first metric (power)
- Don't expect all data is in latest format or even available, support legacy, modeling missing things, metadata
- Use of configurable pipelines for reuse and transparency

Dynamic scaling

In general, networks need to support time-varying traffic demand. Networks also need to support various resiliency and reliability requirements. And be sustainable!

Supporting these needs leads to a multi-objective optimization problem.

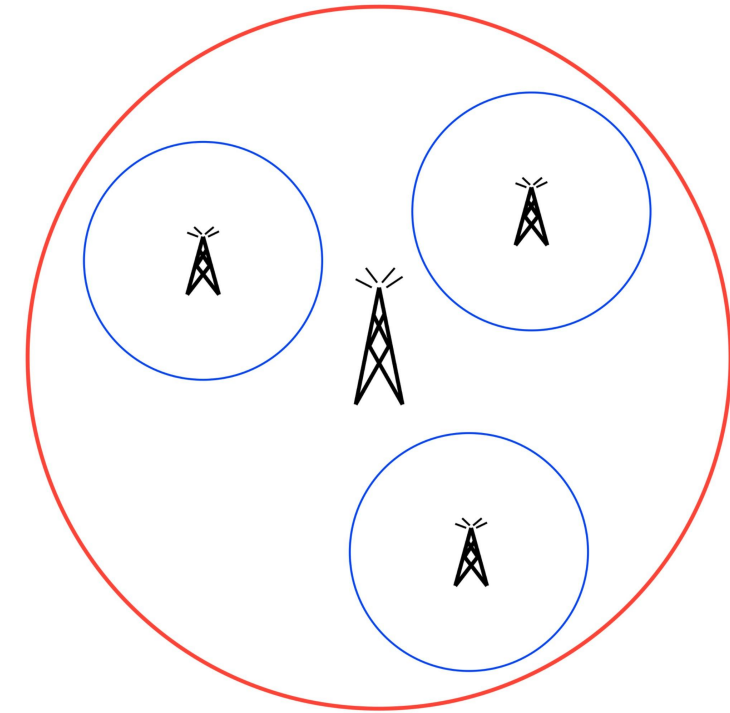
Some of the basics tools are:

- Nodes should be power proportional where possible
- There should be ways to turn nodes/links on and off (GREEN WG)

Dynamic scaling 2

However, basics may not be enough when considering entire systems rather than individual nodes:

- Where per-node power proportionality is insufficient, shut down entire nodes/parts
- Collect information about load/power/carbon to make the right decisions
- Manage the impact on rest of the network, who needs to contact which instance, moving workloads, pre-configuration (e.g., TVR)
- Understand residual impacts on jitter, latency

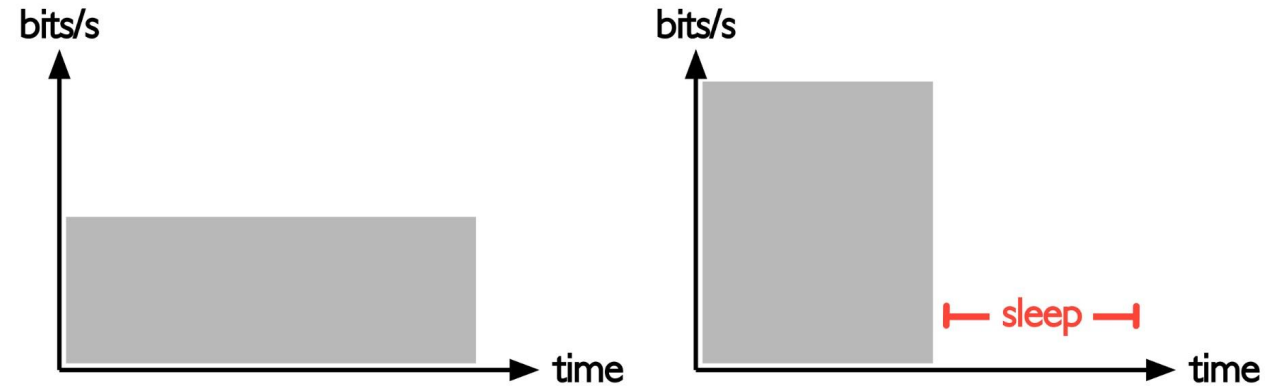


Cellular network with overlapping coverage

Transport improvements

What can transport do?

- Can we enable nodes to sleep longer?
- Sending faster may be an overall win, if nodes have energy proportionality [[NotTradeoff](#)]
- However, there are multiple factors, e.g., need to pace to avoid congestion or disturbing other traffic [[Sammy](#)]
- Send as fast as possible without increasing queue size?



Equipment longevity

Extending the life of protocols and network equipment. Consider:

- Amortization of embedded energy costs over longer period
- Tradeoffs between upgrading to more energy efficient equipment vs. running older equipment for longer
- Optimization for current vs. later use (extensibility/reuse)
- Reuse in different contexts (e.g. Core->Edge->Access)
- Tradeoffs between software programmability vs. hw upgrades (avoid obsoleting hardware)

Compact encoding

Better encoding of messages and content can reduce number of bits sent. But:

- How big is the effect? Moran et al. showed significant savings in their paper [[ref](#)] for some applications, but what would the effect be in the overall Internet traffic mix?
- At what layer is the compactness achieved?
- Are there tradeoffs related to compressing effort, e.g., when using AI-based methods

Early recommendations

Dynamic scaling clearly has significant potential for savings, but more detailed guidance would be useful

Avoiding chattiness would be beneficial. Transport-related optimizations appear to have potential for significant savings, but needs confirmation from further research

Efficient protocol message encoding helps, but does it help in significant ways across entire Internet traffic?

An ask for the SUSTAIN RG

Additional thoughts on this topic are very welcome

Research on the issues that we've labeled as needing more research would be much appreciated, too!