Green computing in IEEE 802.3az enabled clusters

Dimitar Pavlov Joris Soeurt

SNE

July 5, 2012

Green Computing

"Data centres emit over 150 metric tons of CO2 per year, and the volume is increasing."[1]

"Carbon dioxide emissions account for 80% of the contribution to global warming..."[2]

- Different strategies towards environmentally sustainable IT
  - Computational efficiency (e.g. optimizing of algorithms)
  - Consolidation (e.g. virtualization)
  - Reducing / recycling of e-waste
  - Resource allocation (e.g. route data to most green datacenter)
  - **Green networking**

[1] Baroudi et al. (2009)
[2] Lashof et al. (1990)

Approaches to Green Networking

"Recent studies have shown that network devices comprise more than 15% of the total energy consumption of a datacenter."[3]

- Adjust transmission power based on cable length
  - Cables of 5m do not need same transmission power as 100m cables
- Power down circuitry when the line protocol is down
  - If the line protocol is down, why keep the hardware active?
- Use signalling to put circuitry in lower power mode when idle
  - Done by IEEE 802.3az
  - Signalling protocol to put circuitry (of both sides of the connection) in sleep mode when the transmit buffer is empty
  - State transitions operates on the microsecond level, and is therefore invisible to higher layers
  - Both sides should announce 802.3az support during autonegotiation

[3] Barroso et al. 2007

Research motivation

Research motivation

How can an application optimize its energy effiency using the 802.3az protocol in cluster environments?

- How does the protocol achieve its energy savings?
- How to model the energy characteristics of 802.3az compliant devices?
- How to apply the energy model in cluster computing?

Cluster Overview

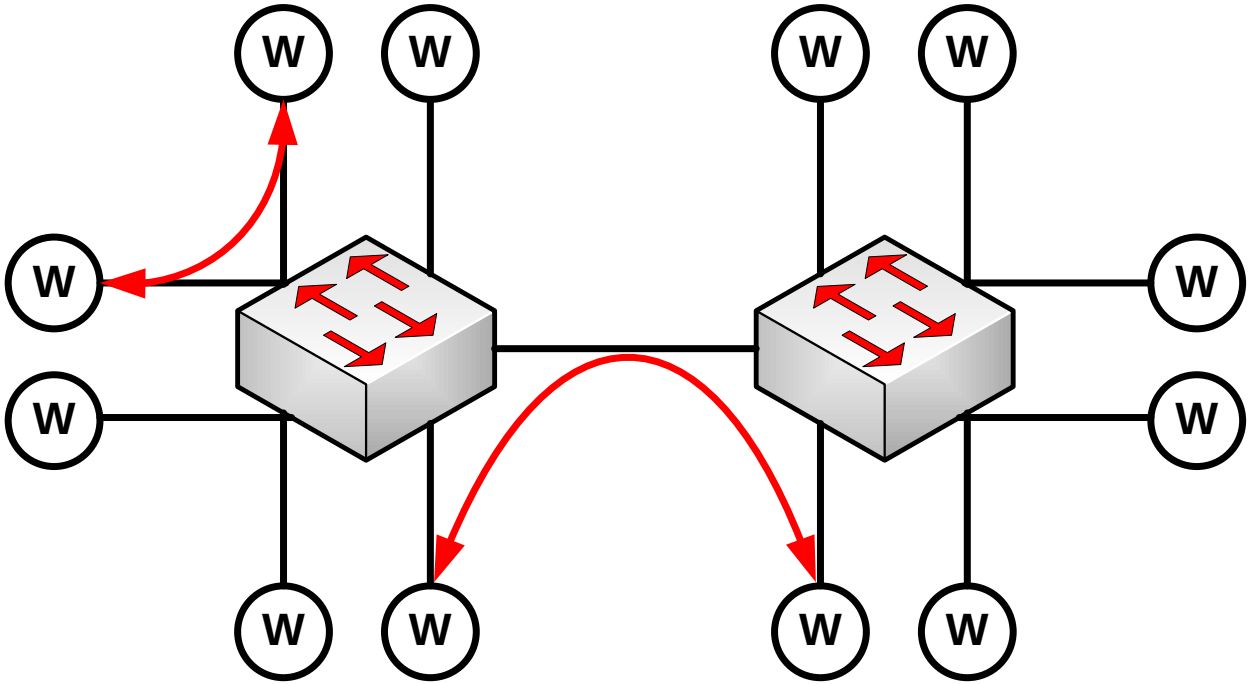
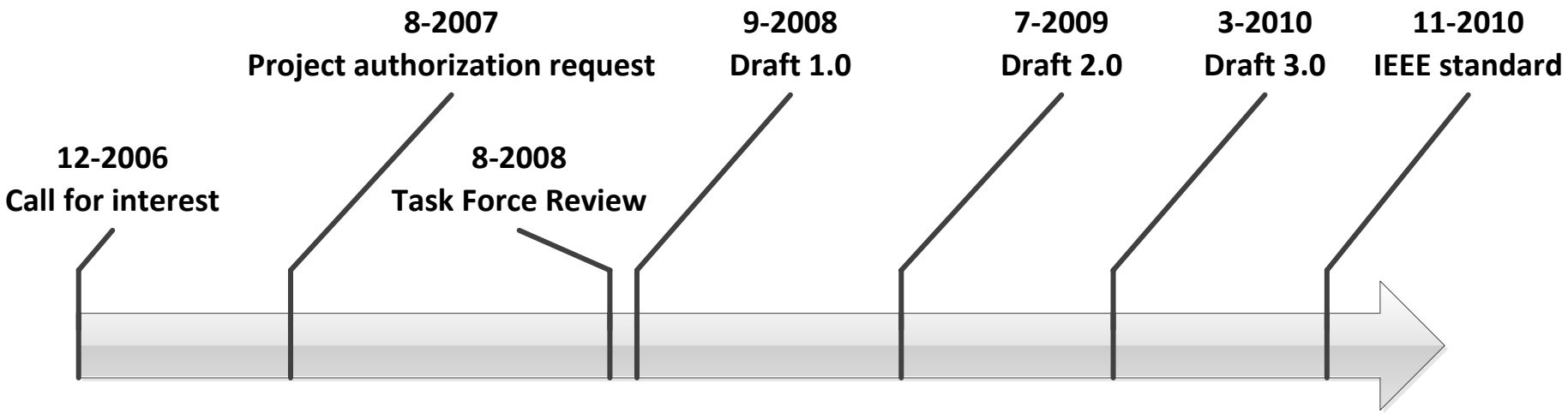


Figure: Simplified model of cluster environment

Theoretical Study on 802.3az

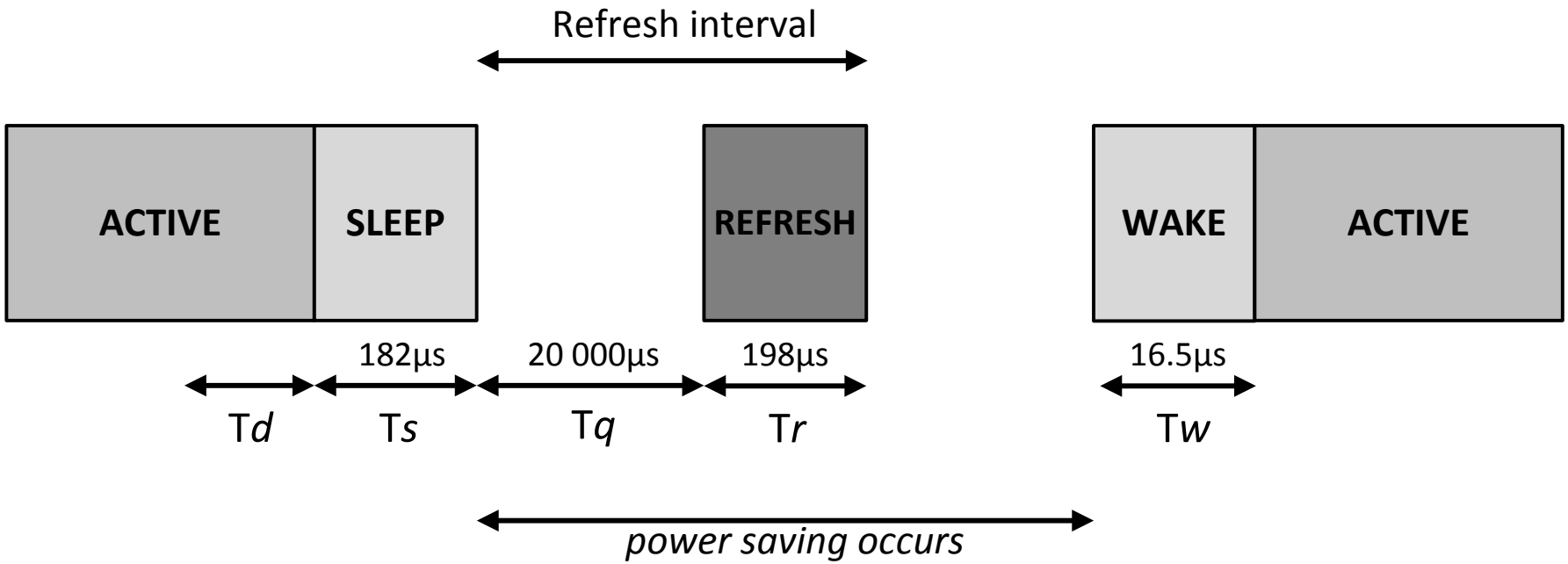
- What is the background of this protocol?
- How does the protocol achieve its energy savings?
- What earlier research has been done on this protocol?

Timeline of 802.3az



- Researched mostly theoretically before, hardware implementations of final standard only now arriving on market
- 802.3az has not been researched in the context of cluster environments

802.3az link states



- T_s = time to sleep
  - Send LPI
- T_q = time quiescent
- T_r = time to refresh
  - Detect link failure
- T_w = time to wake
  - Equals time needed for sending max size frame
- T_d = decision time

802.3az projected energy savings

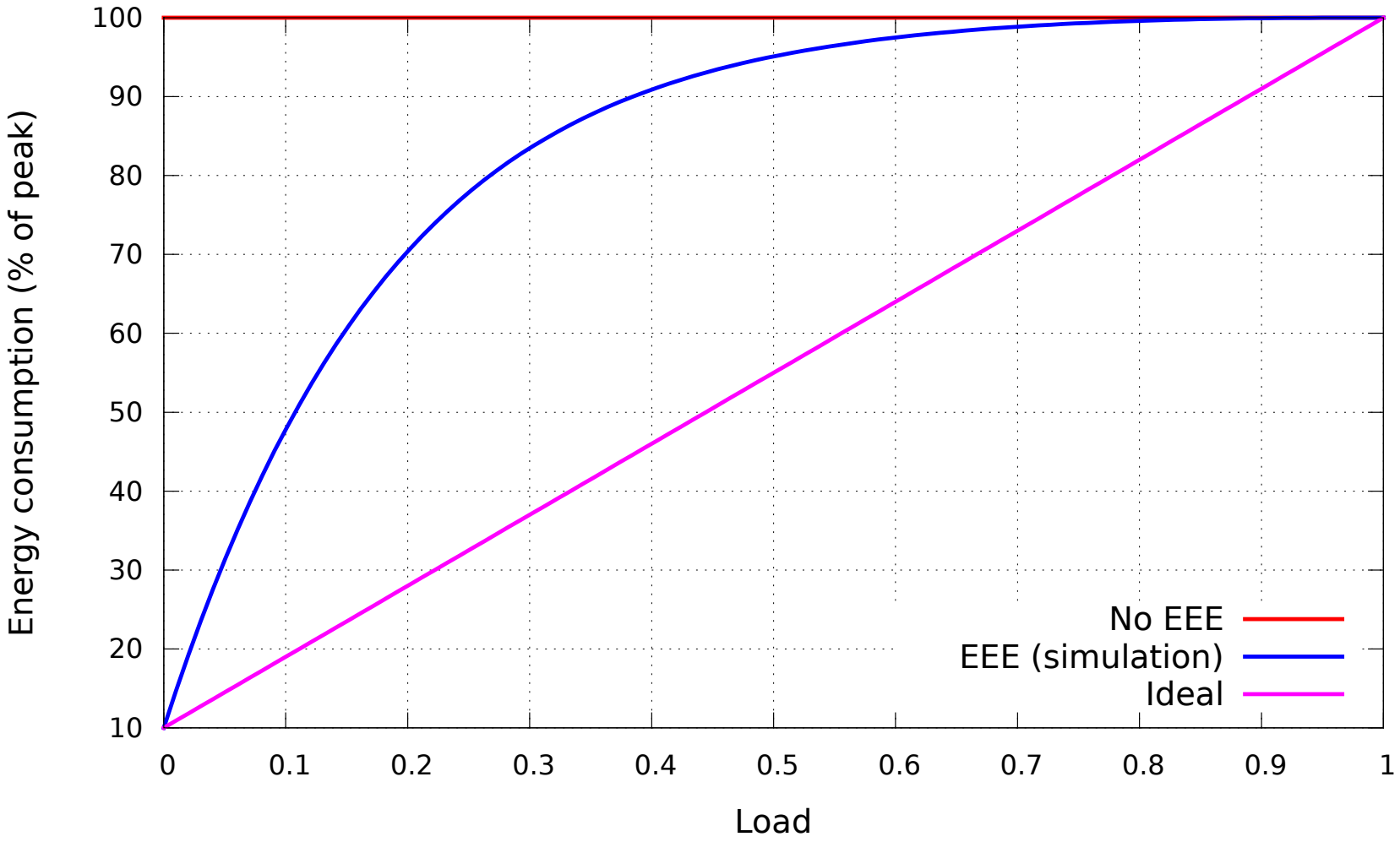


Figure: Simulated energy consumption [Reviriego 2009]

Experimental Phase

- Observe the energy behaviour of 802.3az in different situations
  - (devices, linkspeeds, throughput and protocols)
- Construct energy profiles for different devices

Equipment Overview

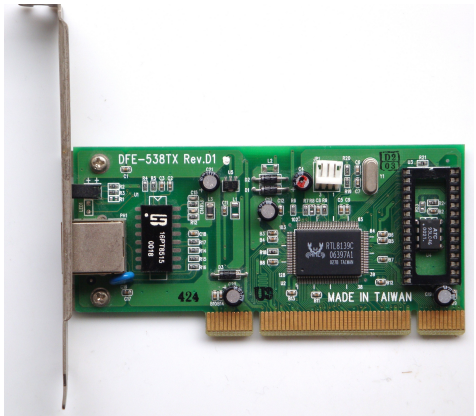
(a) Cisco SG300-28

(b) Huawei S1728GWR-4P

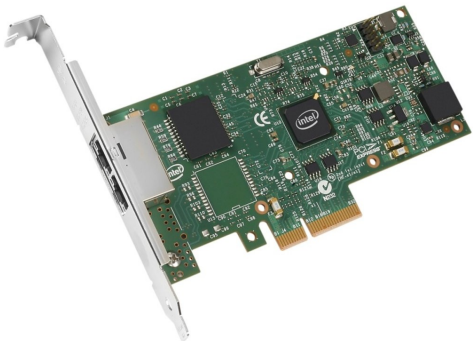
(c) Extreme x440-24p

(d) Racktivity PDU

(e) Schleifenbauer PDU



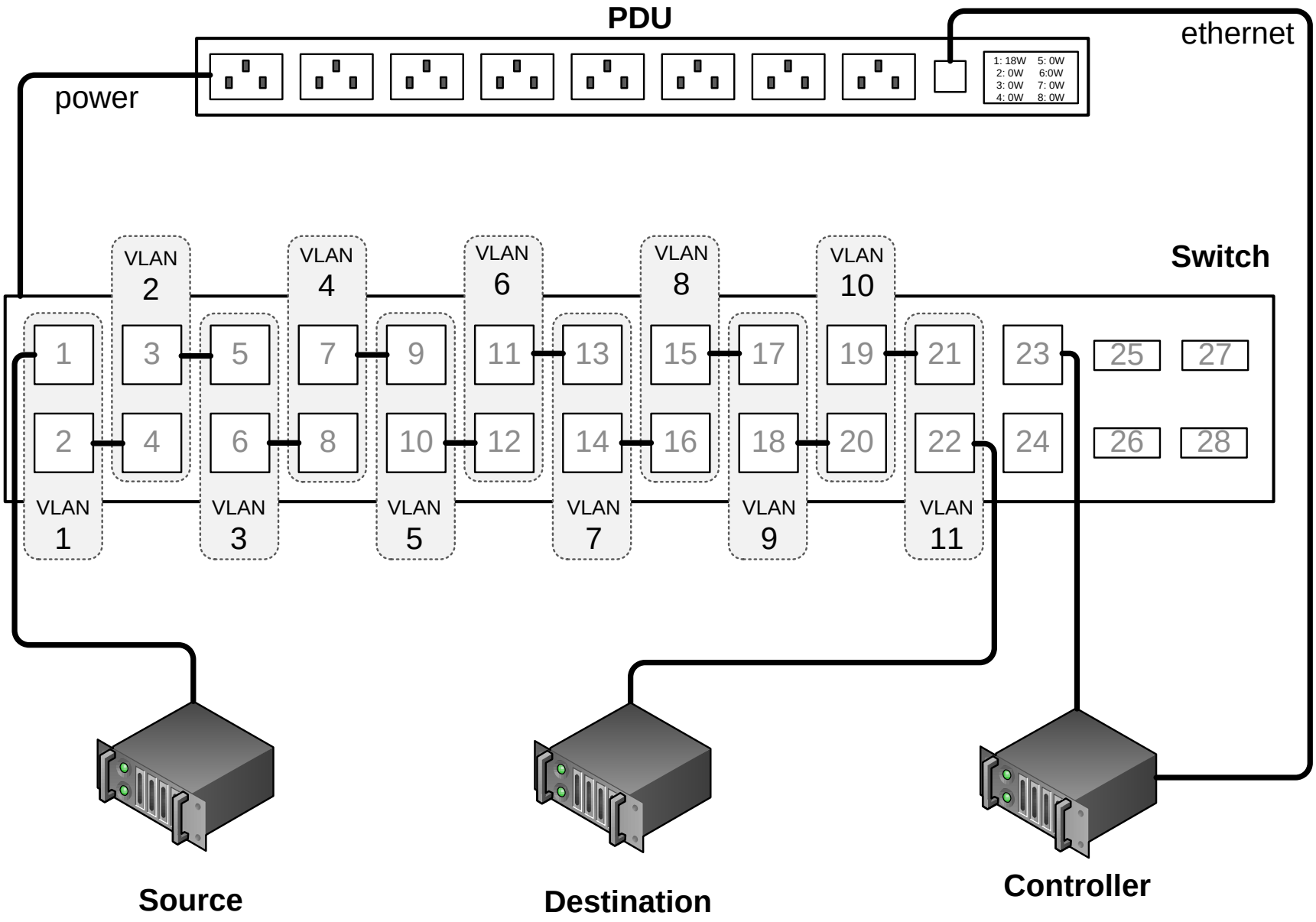
(f) Realtek 8111E



(g) Intel I350-T2 NICs

(h) Servers

Test setup


PDU
ethernet
power
1: 18W 5: 0W
2: 0W 6:0W
3: 0W 7: 0W
4: 0W 8: 0W
Switch
VLAN 2
VLAN 4
VLAN 6
VLAN 8
VLAN 10
1 3 5 7 9 11 13 15 17 19 21 23 25 27
2 4 6 8 10 12 14 16 18 20 22 24 26 28
VLAN 1
VLAN 3
VLAN 5
VLAN 7
VLAN 9
VLAN 11
Source
Destination
Controller

Test setup

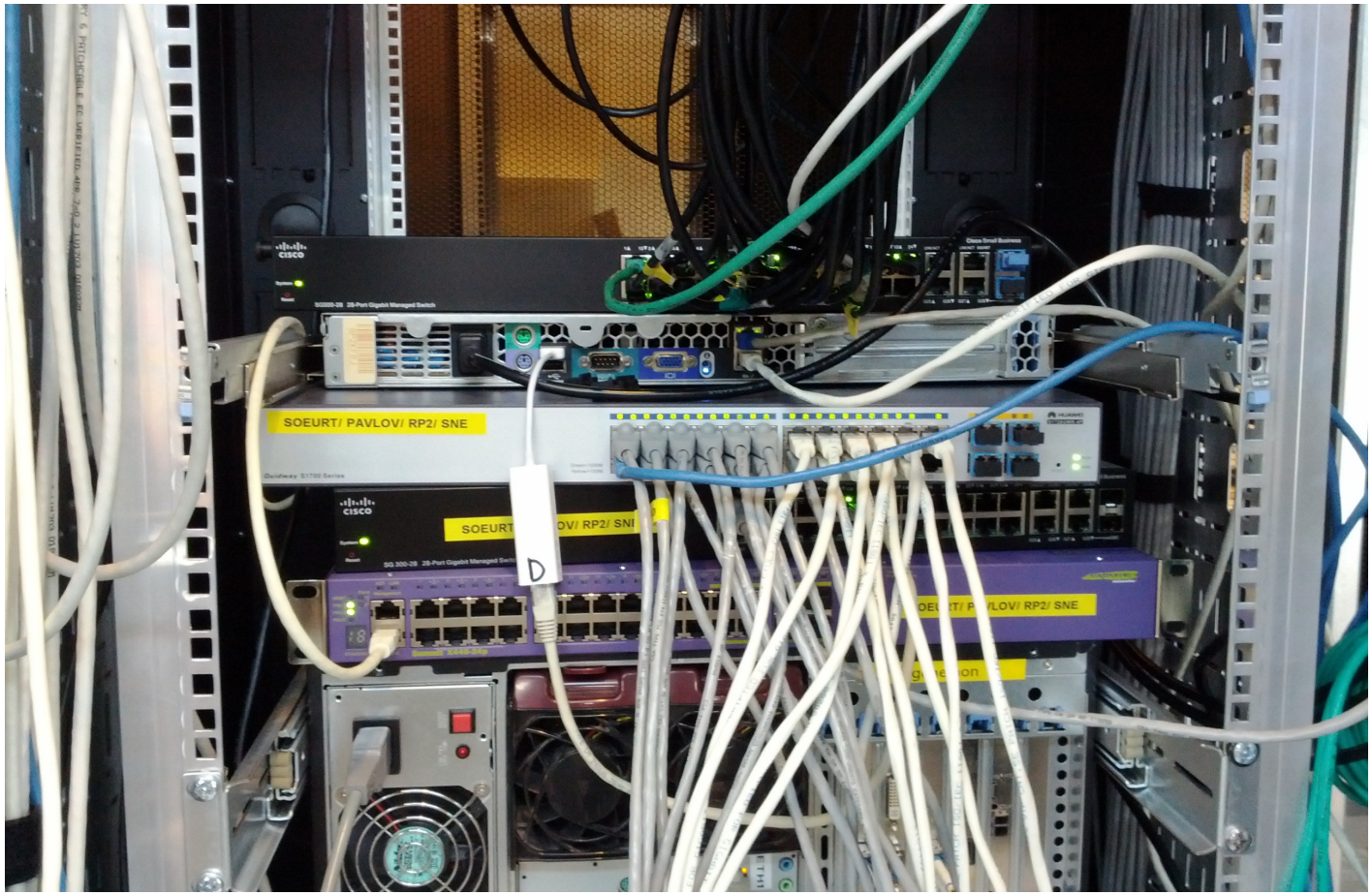


Figure: What it actually looked like...

Experiments – Maximum energy savings

- Goal: determine maximum energy savings with 802.3az
  - Fully utilize the switch to measure maximum consumption
  - Measure the minimal switch consumption when no traffic is present
  - Compare both measurements to determine maximum energy savings

Experiments – Maximum energy savings (cont'd)

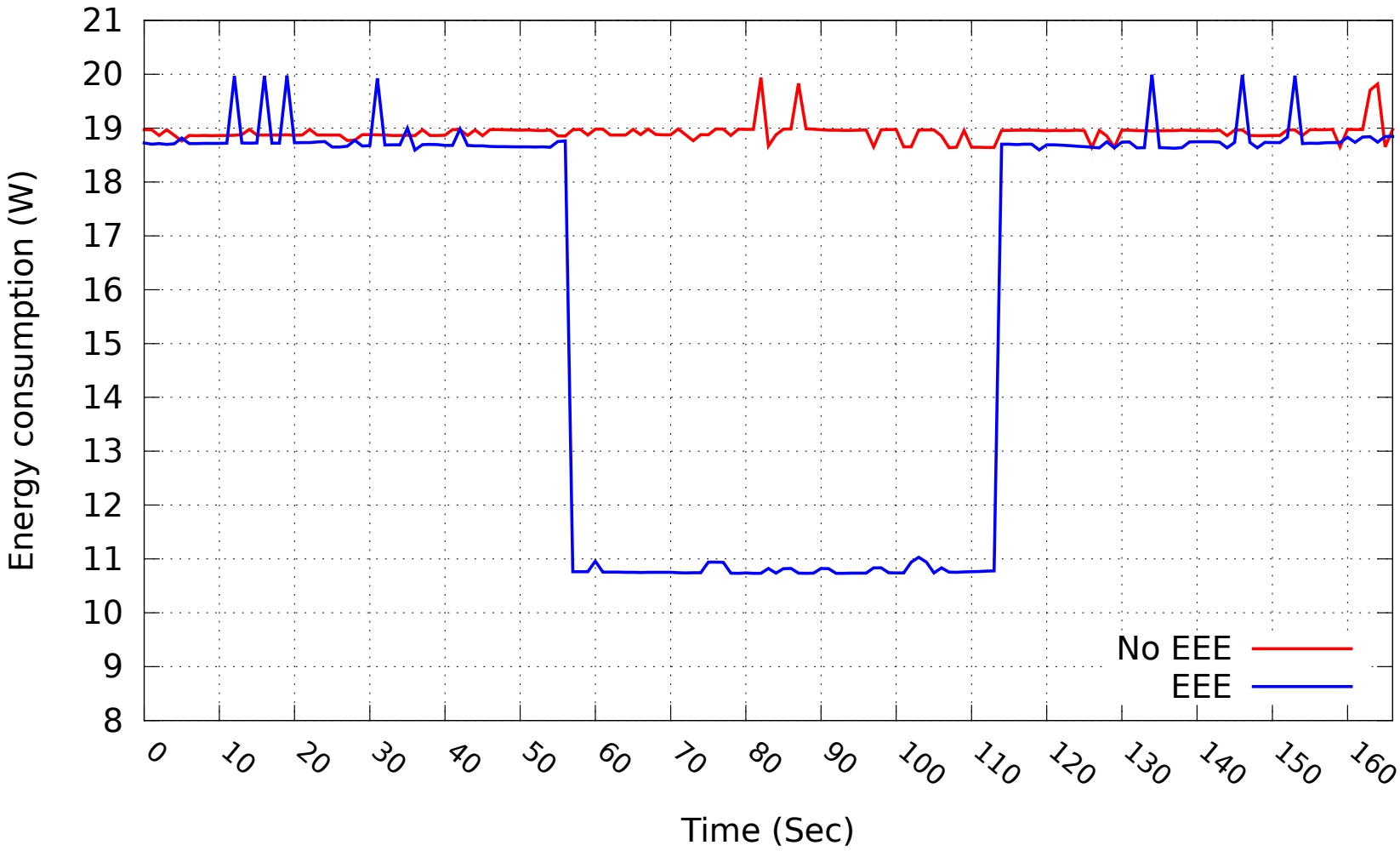


Figure: Cisco SG300-28 using Iperf – TCP/UDP traffic at 1Gbps

Experiments – Throughput vs energy consumption

- Goal: determine the relationship between throughput and energy usage
  - Generation of traffic is done with Iperf
  - The transmission rate is set with tc per test run
  - Energy usage of the switch is measured per test run
  - Traffic is generated for 5 minutes then the measurements are averaged per run

Experiments – Throughput vs energy consumption (cont'd)

- Results from TCP tests only. UDP shows unexpected results.

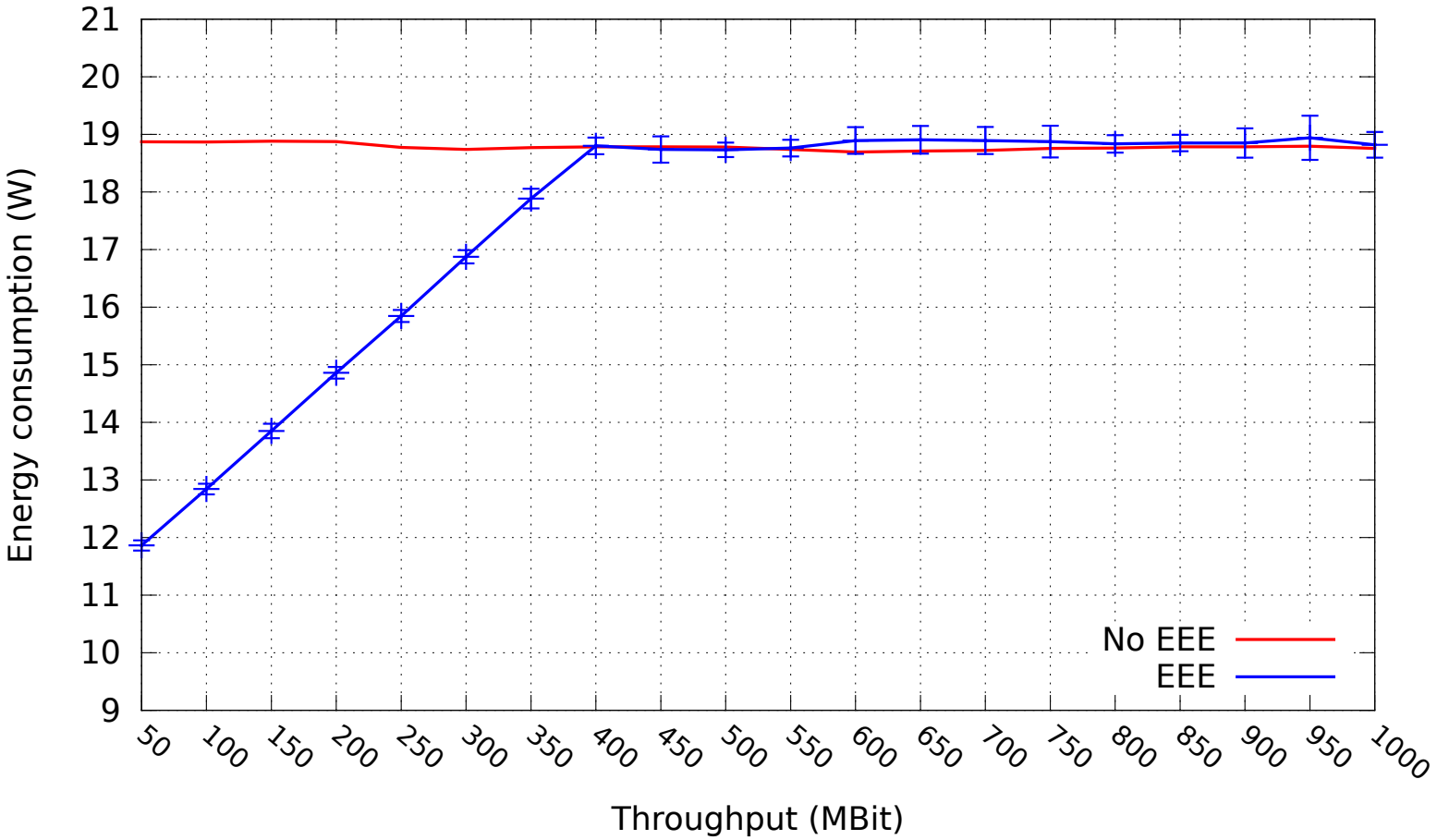


Figure: Cisco SG300-28, $E_s = \begin{cases} 10.9 + Thrghpt * \frac{18.9-10.9}{400}, & \text{if } Thrghpt < 400 \\ 18.9, & \text{if } E_s \geq 400 \end{cases}$

Experiments Summary

- Experiment shows that 802.3az has the potential to save power
- Vendor claims of 30% savings are generally true
- Odd power usage distribution – does not conform to previous research[4]
- Constructed power profiles, which were used as input to final phase

[4] Riveriego, 2009

Applying 802.3az in applications

- Distributed computing & Cluster computing
- Model for optimizing energy usage
  - Define a way to determine energy usage with 802.3az
  - Estimate switches energy consumption based on number of active ports
  - Estimate time distribution for particular tasks with a focus on parallel computing
  - Determine best transmission rate for a fixed quantity of data
- Combine output of all phases and create a prototype power calculator

Estimating time distribution in parallel computing

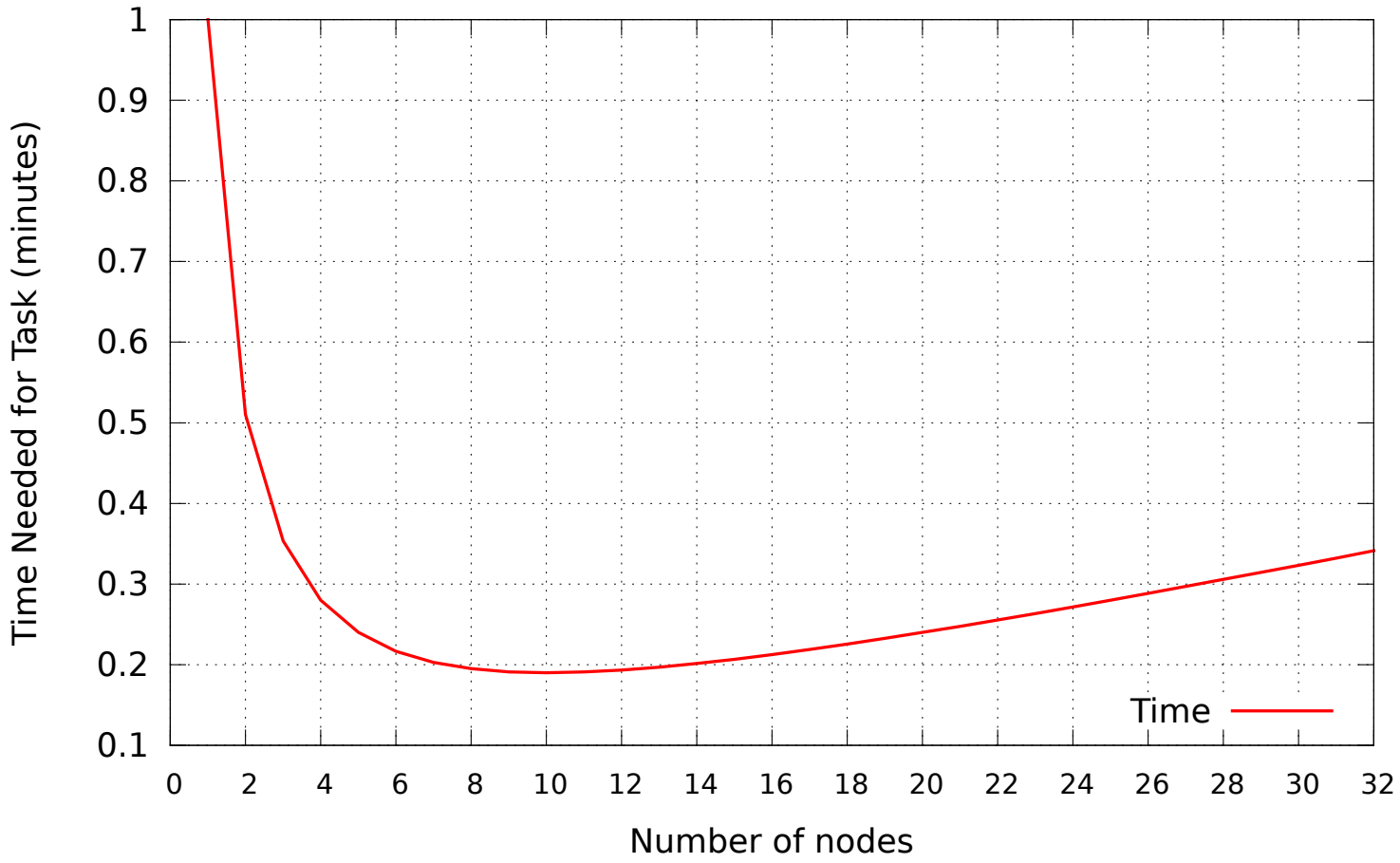


Figure: Time Needed for Task, $T_t = \frac{T_s}{n} + (n-1) * C$

Total power consumption for a parallelized task

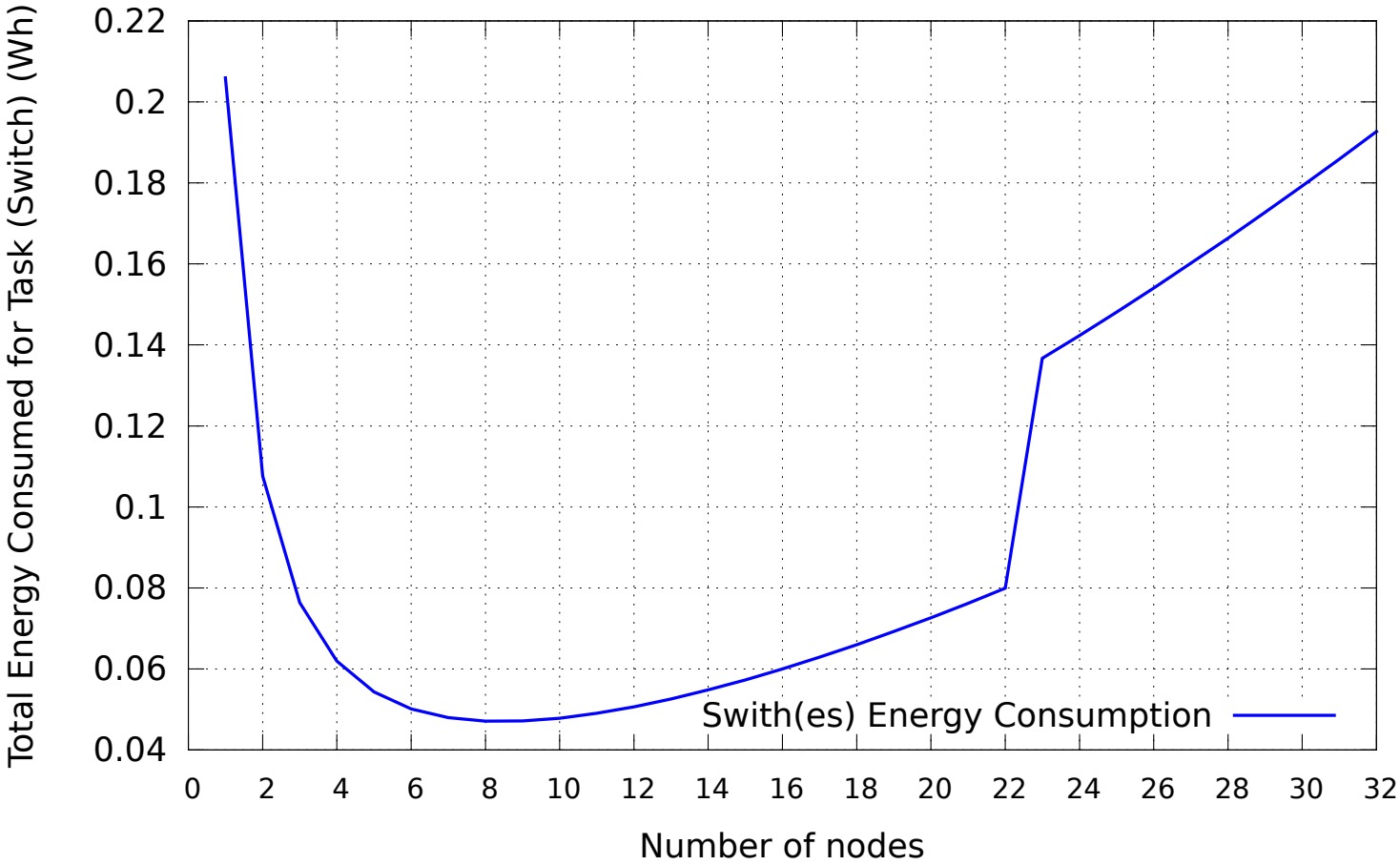


Figure: Huawei S1728GWR-4P, $E_{task} = (P_b * N_s + P_{pp} * N_n) * T$

Estimating optimal bandwidth for a fixed-datasize task

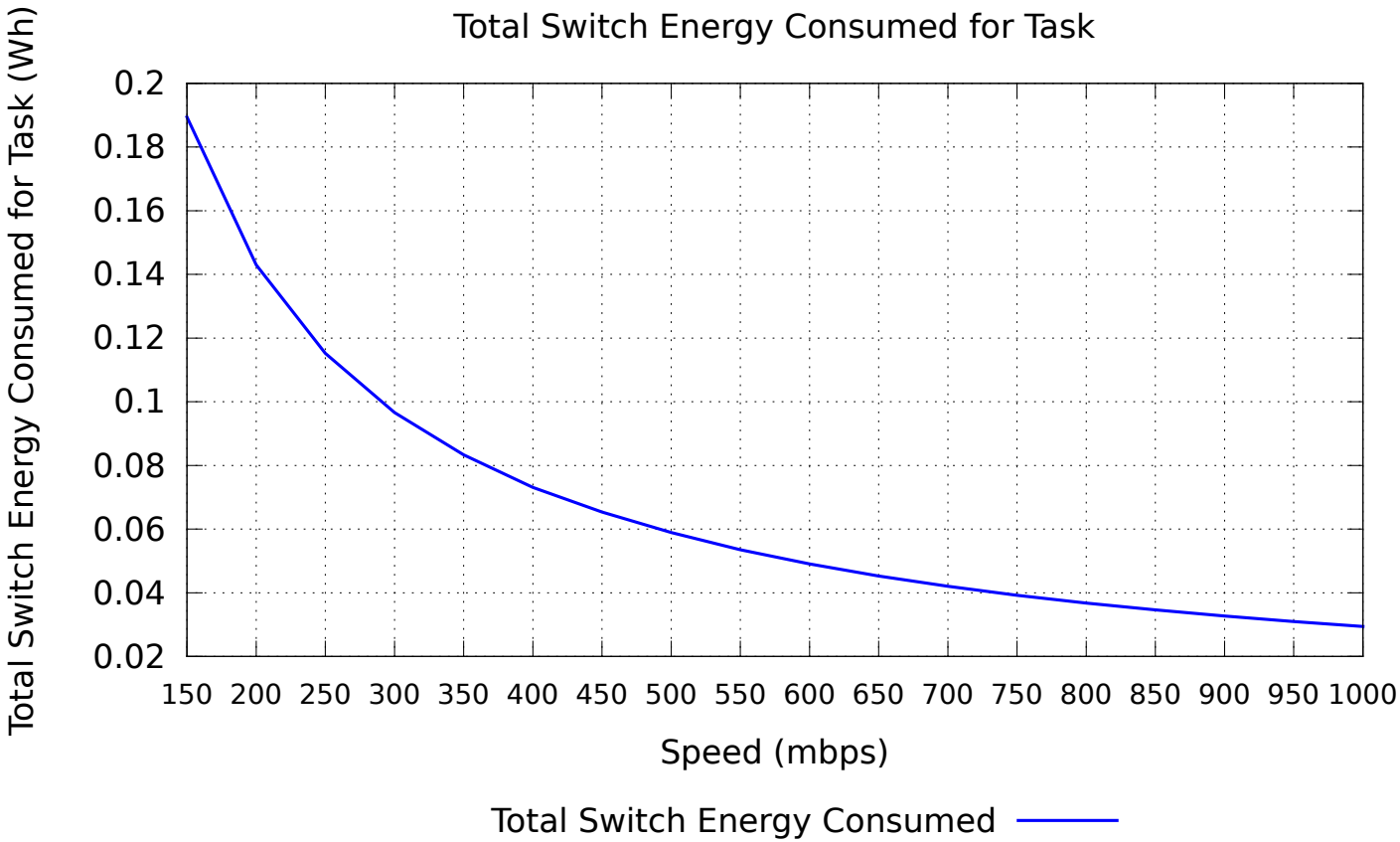
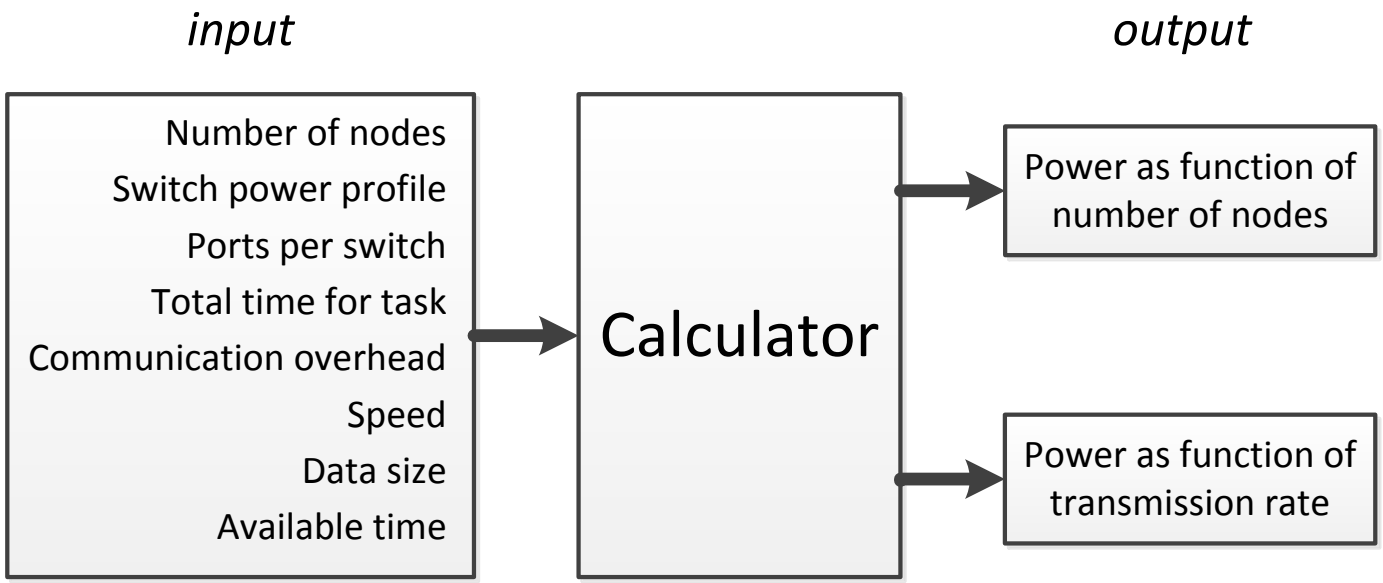


Figure: Huawei S1728GWR-4P, $E_{task} = P_{total} * \frac{(A_d * 8)}{S_t}$

Prototype power calculator operation



Conclusions

- The 802.3az protocol can potentially optimize the energy efficiency of networked environments
- The technology, when applied within a distributed computing environment, contributes to green IT efforts
- 802.3az can save energy with the vast majority of traffic patterns
- To achieve optimal energy savings, one needs to perform low-level software changes

Future work

- Extending the created energy model and applying it to other distributed computing environments
- Analyze actual cluster traffic patterns to optimize model and recommendations
- Work towards integrating computational performance into energy model
- Research applicability with multi-core architectures and incorporate into energy model
- Further investigate 802.3az in the context of UDP and 100BASE-TX
- Investigate 802.3az in the context of other transport protocols

Questions

Dimitar Pavlov - dimitar.pavlov@os3.nl
Joris Soeurt - joris.soeurt@os3.nl