Aalto University
School of Science
Degree Programme in Computer Science and Engineering

Betsegaw Lemma Amersho

Simulating energy-aware networks in large-scale distributed systems

Master's Thesis
Espoo, June 26, 2017

DRAFT! — April 19, 2017 — DRAFT!

Supervisors:	Professor Antti Ylä-Jääski, Aalto University
	Professor Pekka Perustieteilijä, University of Helsinki
Advisor:	Olli Ohjaaja M.Sc. (Tech.)

Aalto University
School of Science

Aalto University
School of Science
Degree Programme in Computer Science and Engineering

ABSTRACT OF MASTER'S THESIS

Author:	Betsegaw Lemma Amersho		
Title: Simulating energy-aware networks in large-scale distributed systems			
Date:	June 26, 2017	**Pages:**	27
Major:	Data Communication Software	**Code:**	T-110
Supervisors:	Professor Martin Quinson Dr. Anne-Cécile Orgerie		
Advisor:	Olli Ohjaaja M.Sc. (Tech.)		

A dissertation or thesis is a document submitted in support of candidature for a degree or professional qualification presenting the author's research and findings. In some countries/universities, the word thesis or a cognate is used as part of a bachelor's or master's course, while dissertation is normally applied to a doctorate, whilst, in others, the reverse is true.

!FIXME **Abstract text goes here (and this is an example how to use fixme).** FIXME! Fixme is a command that helps you identify parts of your thesis that still require some work. When compiled in the custom `mydraft` mode, text parts tagged with fixmes are shown in bold and with fixme tags around them. When compiled in normal mode, the fixme-tagged text is shown normally (without special formatting). The draft mode also causes the "Draft" text to appear on the front page, alongside with the document compilation date. The custom `mydraft` mode is selected by the `mydraft` option given for the package `aalto-thesis`, near the top of the `thesis-example.tex` file.

The thesis example file (`thesis-example.tex`), all the chapter content files (`1introduction.tex` and so on), and the Aalto style file (`aalto-thesis.sty`) are commented with explanations on how the Aalto thesis works. The files also contain some examples on how to customize various details of the thesis layout, and of course the example text works as an example in itself. Please read the comments and the example text; that should get you well on your way!

Keywords:	ocean, sea, marine, ocean mammal, marine mammal, whales, cetaceans, dolphins, porpoises
Language:	English

Acknowledgements

I wish to thank all students who use LaTeX for formatting their theses, because theses formatted with LaTeX are just so nice.

Thank you, and keep up the good work!

Espoo, June 26, 2017

Betsegaw Lemma Amersho

Abbreviations and Acronyms

2k/4k/8k mode	COFDM operation modes
3GPP	3rd Generation Partnership Project
ESP	Encapsulating Security Payload; An IPsec security protocol
FLUTE	The File Delivery over Unidirectional Transport protocol
e.g.	for example (do not list here this kind of common acronymbs or abbreviations, but only those that are essential for understanding the content of your thesis.
note	Note also, that this list is not compulsory, and should be omitted if you have only few abbreviations

Contents

Chapter 1

Introduction

1.1 Context of the Study

Provide a brief history of the issues to date.

Situate your particular topic within the broad area of research.

Note that the field is changing, and more research is required on your topic.

1.2 Problem statement

Even though there is a growing concern about the energy consumption of large-scale network infrastructures, a search of the literature revealed few studies which address the issue of estimating the energy consumption of these network infrastructures. The existing proposed solutions being packet-level estimators are not scalable (in-terms of memory usage and speed) to be used in the domain of large-scale distributed networks such as cloud, grid, and peer-to-peer computing.

1.3 Aim and Scope

The purpose of this work is to propose efficient and reasonably accurate flow-level model in the context of SimGrid simulator for estimating the energy consumption of large-scale distributed networks.

In the domain of large-scale distributed network infrastructure, we can categorize energy consuming components into two broad groups: IT equipments (which includes computing servers, storage servers and networking components) and infrastructure components (which includes power provisioning, cooling and lighting components). Since SimGrid already have energy consumption model for computing servers, the scope of this study is on the network component part within the IT equipments category.

1.4 Significance of the Study

Explain how your thesis contributes to the field.

There are four main areas of contribution: theory development, tangible solution, innovative methods, and policy extension. One of these contributions must be identified as the basis of your primary contribution to the field.

In contrast to reports for industry, theory development is an expected and required contribution; for PhDs in particular, it must be "original".

1.5 Structure of the Thesis

You should use transition in your text, meaning that you should help the reader follow the thesis outline. Here, you tell what will be in each chapter of your thesis.

Chapter 2

Background

In this chapter ...

2.1 Electricity consumption of ICT equipments

ICT equipments consume a significant amount of electricity. A survey conducted by Heddeghem et al. [12] shows the electricity consumption and growth trends of three classes of ICT equipments: personal computers, communication networks, and data centers. Personal computers include equipments such as desktop, laptop and external monitors. Communication networks includes residential network access equipments (such as WiFi routers and modems), network equipments used in offices (such as routers and switches) and telecom-operator network equipments (such as base stations, routers and optical amplification systems). Data-centers house storage and computing servers, communication network equipments, and power provisioning and cooling facilities. In this classification there are overlaps, for instance, telcom operator can have office network equipments and data-centers. After carefully avoiding possible redundant measurements, the researchers estimated absolute electricity consumption and annual consumption growth rate of each category of equipments for the period 2007 and 2012. The results of the study show that the global electricity consumption of ICT equipments in all the three categories combined contributed 3.9% in 2007 and 4.6% in 2012. The estimated annual growth rate of the individual category is 5% for personal computers, 10% for communication networks, and 4% for data-centers. These growth rates are higher than that of the total global electricity consumption, which is 3%.

2.2 Data-center electricity consumption

In Section 2.1 we described data-center's global share in electricity consumption. In this section we describe the components involved within the data center itself.

Electricity consumption units with in a typical data-center can be classified into two broad groups [8]: The first group is IT equipments (which includes computing servers, storage servers and networking components) and the other group is infrastructure facilities (which includes power provisioning, cooling and lighting components).

Figure 2.1 [8] shows the electricity consumption proportion of the data-center components. This value differs significantly from one data-center to another [2], for instance, due to architectural difference[11] or energy efficiency of the components. The infrastructure facility components take the large proportion (65%) of the consumption. Though the infrastructure facil-

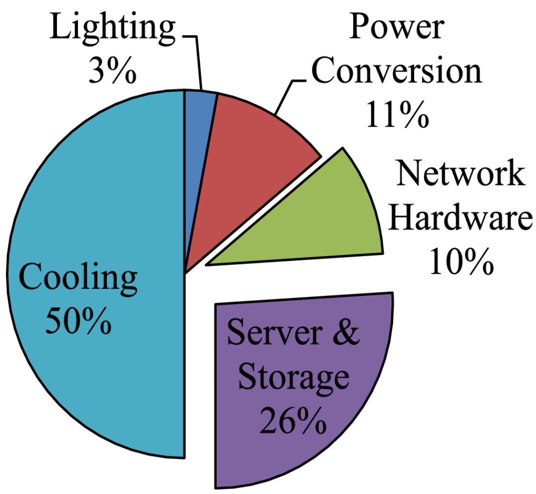

Figure 2.1: Energy consumption percentage of data-center components [8]

ity consumes relatively larger amount of electricity, the focus of this study is on the IT equipment components, particularly on the network equipments.

If we further zoom in on the IT equipments part, we can find server, storage and network equipments. A data-center servers consist of one or more CPU cores, memory and I/O devices. The energy consumption relationship among these components is shown in Figure 2.2. Combined, Memory and CPU units consume the larger amount of energy relative to other components. The fact that CPU is the dominant electricity consuming unit is exploited by Fan et al. in [9] to model the dynamic power usage of thousands of servers by using only CPU utilization as a parameter. The result of their study was very accurate, with error as low as 1%.

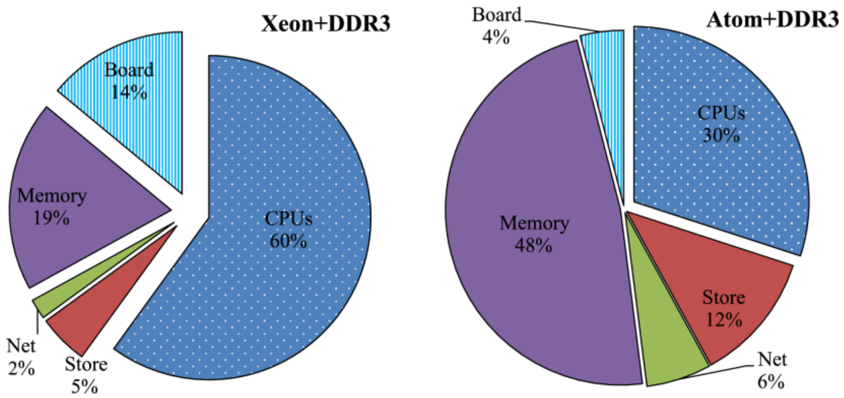

Figure 2.2: Energy consumption percentage of Xeon based (on the left) and Atom based (on the right) servers [8]

2.3 Energy proportionality

The primary reason the study of energy consumption management of network equipment becomes so important is that, in general, ICT equipments do not consume energy proportional to their workload. An ideal ICT equipment is the one which consume zero electricity when it is idle, and it consumes electricity proportional to its workload when it is active. However, the reality is, even power efficient servers consume about 50% of their peak power [3], even when they are doing nothing. This percentage can even reach 85% for network switches [10]. Figure 2.3 in [15] shows the energy proportionality of a typical network equipment. From the graph we can observe that the dynamic power consumption range is narrow. Three approaches are in common use to deal with this situation. The first one is re-engineering network devices so as to make them more energy proportional, device vendors are the prime role player in this aspect. The second approach is related to the operating rate of a network equipment port. A typical switch can operate on different transmission rate (100Mbps, 1 Gbps or 10 Gbps). An active port transmitting at 10 Gbps can consume more energy than if it transmit at 100 Mbps. Rate adaptation is the approach devised to take advantage of this situation. Instead of transmitting at the maximum rate all time, the network port can be made to adapt to the actual traffic load. This energy saving approach is

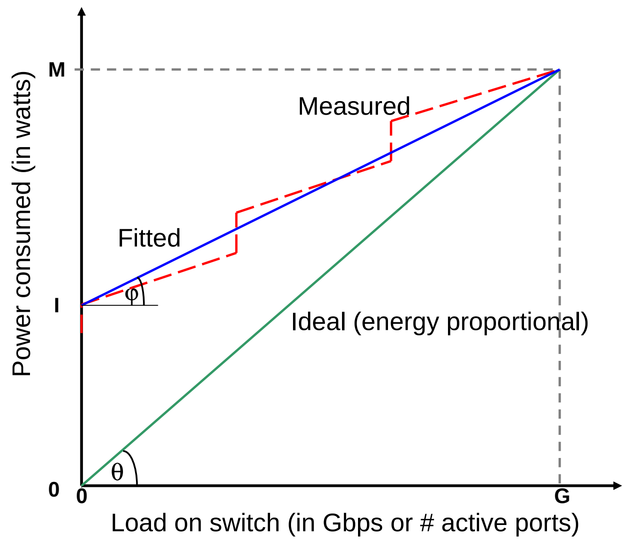

Figure 2.3: Ideal and measured energy proportionality of a network equipment [15]

known as Adaptive Link Rate (ALR). The third approach, which is known as Low Power Idle (LPI), allows a network device to send data as fast as possible and then enter low power mode between transfers. The low power mode can further be extended by a technique called packet coalescing, which allows more energy saving [4].

2.4 Packet-level and flow-level Simulators

Packet-level simulators strives to model a given network phenomenon at the granularity level of packets, thus in general they are accepted by the research community to be more accurate compared to flow-level simulators [6]. One of the most popular packet-level simulator is NS-3, which is categorized under discrete-event simulator with events corresponding to sending and receiving of packets [14]. Though packet-level simulators are accepted to be more accurate, they fail to scale well in the area of large-scale networks.

In the area of large-scale networks, flow-level simulators are the preferred alternative. Rather than modeling a given network phenomenon at a packet level, flow-level simulators treat a set of packets as a single unit [6]. The most commonly used definition for flow in the context of computer networking is coined by Claffy et al. in [7]:

"... a flow ... a unidirectional traffic stream with a unique [source-IP-address, source-port, destination-IP-address, destination-port, IP-protocol] tuple ... "

In addition to the five tuple mentioned in the definition, a flow also has a limited time duration. Claffy et al. used a time limit of 64 seconds as a flow duration in their study. Researchers such as Carneiro et al. [5], adopted this same definition to develop flow monitoring module for NS-3, a module that can generate information such as amount of packets or bytes transferred, packets dropped or transmission start and end time for each flow. Barakat et al. in [1] also used the same definition to model traffic at the flow-level for the Internet backbone link. By abstracting away fine details, flow-level models provides easy way to instantiate experiments and they also scale very well for conducting large-scale network simulations [1, 6].

The flow definition given above is not the only one. Any analytical model which capture the characteristics of a given network phenomenon can be considered as flow-level model. In SimGrid, for instance, TCP flow is modeled characterized by bandwidth and end-to-end latency[6].

2.5 Simulating and modeling energy consumption of large-scale networks

One way of conducting energy consumption or any other experiment is to use real production environment or test-bed environment, both are referred to as *in vivo* in [6]. In the former case, handling transient and varying conditions would make the data collection and prediction very difficult and often times, a production environment is not available for experimentation. In the later case, it requires setting-up a separate testing environment designed solely for the purpose of conducting the desired experiment. This approach apart from being expensive, it requires significant amount of time for experiment setup and, it is also non-repeatable as experimenting with different scenario demands a modified or new configuration.

The other alternative for experimenting is simulation, also referred to as *in silico* in [6]. Simulation, unlike real environment, allows great flexibility in terms of experiment configuration, control and repetition. In addition it can also be less time consuming and less expensive.That is why virtually in all computer network related researches simulations are widely used.

In this study we simulate energy-aware large scale distributed networks using SimGrid (Detail description about SimGrid follows in the next section).

When we say large-scale distributed network, we are referring to a set of networks residing inside in the distributed data centers and also the networks that are used to connect them.

The energy consumption E of an equipment depends on the operating power P at time t. The total energy consumption for a time period T is given by Equation 2.1 [16].

$$E(T) = \int_0^T P(t)dt \tag{2.1}$$

Due to the energy proportionality characteristic described in Section 2.3, the common approach used to compute the energy consumption is to divide the power component into two parts: static/idle power (P_{static}) and dynamic power ($P_{dynamic}$) as shown in equation 2.2. Then the total energy is obtained by multiplying the total power, P_{total} by the time duration [8, 13, 15, 16].

$$P_{total} = P_{static} + P_{dynamic} \tag{2.2}$$

For a typical network equipment such as a switch, the static part constitutes the power consumption of the chassis and the line-cards (when all the ports on the line-cards are switched off). The dynamic part, on the other hand, constitutes the power consumption of the switch ports running at a given rate multiplied by the utilization factor [15]. Equation 2.3 shows how to compute the total power for a switch, where P_{switch}, is the total power consumption of a switch, $P_{chassis}$ and $P_{linecard}$ is the power consumption of the chassis and the line card, respectively. P_{rate}, is the power consumption of a given port at a given rate and $numports_{rate}$ is the number of ports running at a given rate. The rate can take values such as 10 Mbps, 100 Mbps, 1 Gbps or 10 Gbps.

$$\begin{split} P_{switch} = P_{chassis} + (numlinecards \times P_{linecard}) + \\ \sum_{rate=min}^{max} (numports_{rate} \times P_{rate} \times utilizationFactor) \end{split} \tag{2.3}$$

2.6 SimGrid

Figure 2.4 shows the structure of SimGrid and how its core works. The top three components are the APIs that users can use to develop their simulation. Both MSG and SMPI are used to specify simulated applications as concurrent processes. The difference is that using MSG, users can simulate any arbitrary application, whereas, using SMPI users can simulate existing MPI applications, the MPI processes are created automatically from C or Fortran MPI programs. SIMDAG, on the other hand, does not use concurrent processes. It allows users to describe their application as communicating task graph. The next layer, SIMIX, implements the mechanisms that are required to simulate the concurrent process of MSG and SMPI applications. It also provides process control and synchronization functionalities. The bottom layer, SURF, is the simulation core, it simulates the execution of activities on computing or communication resources [6]. In SimGrid for each simu-

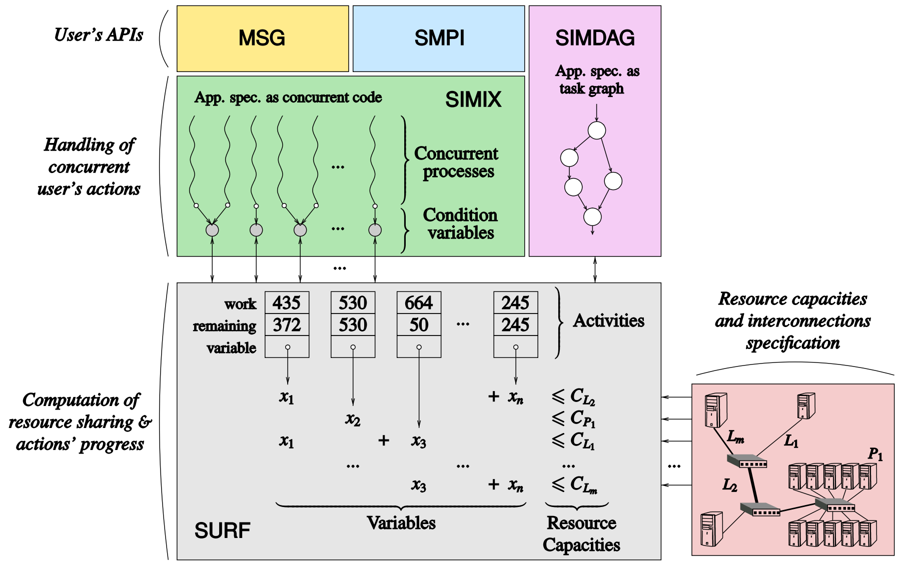

Figure 2.4: Architecture of SimGrid [6]

lated activity, such as computation or data transfer, there is a corresponding condition variable, in Figure 2.4 it is shown in SIMIX box. This condition variable synchronizes the concurrent processes of the simulated applications. The computing (P_x) and the communication (L_x) resources are shown on the bottom-right side of the figure. Computing resources are defined in terms of computing power, whereas, communication resources are defined in terms of bandwidth and latency. As shown in the SURF box, multiple activities can share the same resource (e.g., (x_1, x_n), (x_1, x_3) or (x_3, x_n)) or one activity can use multiple resources (e.g., x_1 or x_3 or x_n). Activities that share the same resource are limited by the capacity of that resource. Each activity is defined by the total and remaining work to be executed. When the work associated with the activity completes, the corresponding upper layer components receive a notification signal [6].

As we have already pointed out in Section 2.4, the primary advantage of flow-level simulation is its scalability in terms of speed and memory usage. SimGrid uses flow-level analytical model for simulating TCP network phenomenon [6]. To see the scalability of the flow-level model, the SimGrid team compared it with other widely used simulators such as GridSim and OverSim. After simulating 500,000 tasks both on GridSim and SimGrid, the results demonstrate that SimGrid is 257 times faster and 26 times more memory efficient. Similarly, the comparison result with OverSim shows that SimGrid is 15 times faster and it can also simulate scenarios 10 times larger. Concerning the accuracy, though the simulator gives very good accuracy in most case studies, there are situations where it fails to give accurate result. As an example, the comparison study of SimGrid with packet-level simulator GTNetS show that for data size less than 100 KiB there is a significant difference in prediction.

Currently SimGrid is used to simulate different network phenomenon in the area of large-scale distributed systems such as grid, cloud, volunteer and HPC [1]. Concerning energy consumption models, it houses energy models for CPU but there is no model for network equipments.Therefore, the focus of this study is to propose and implement network energy consumption model for SimGrid. The implementation of this model, together with the existing CPU energy model, allows us to estimate the energy consumption of large-scale networks that reside within or outside a data-center that are discussed in Section 2.1 and Section 2.2.

2.7 Related Simulators

2.7.0.1 ECOFEN

2.7.0.2 GreenCloud

Kliazovich et al. [13] proposed GreenCloud, a simulator that can estimate energy consumption of cloud computing data centers. GreenCloud is developed as extension to NS-2 packet-level network simulator. This simulator contains power consumption models both for the computing and communicating components residing in a typical data center. The power consumption model used for the computing component is shown in 2.4. This equation contains power consumed by the fixed parts (such as bus, memory and disk) which consume power independent of the operating frequency f of the computing component CPU and the power consumed by the CPU (P_f) operating at a given frequency f. This model allows for lowering the operating frequency of the

[1] http://simgrid.gforge.inria.fr/

CPU when workload becomes below some predefined threshold in order to decrease the power consumption.

$$P_{computing} = P_{fixed} + P_f \times f^3 \tag{2.4}$$

The power consumption model used in GreenCloud for the communicating components is the one shown in Equation 2.3. The equation shows the static power consuming parts (such as the chassis($P_{chassis}$)and the active line cards($P_{linecard}$) and the dynamic part (P_{rate}), is the energy consumed by the port running at a particular line rate for a given traffic load.

This simulator is limited in three aspects: number of allowed CPU cores, versatility, and scalability. First, only one CPU core is allowed per simulated node. This hinders the study of energy consumption of multi-core computing nodes. Second, we can not use this simulator outside the cloud computing domain such as grid, volunteer, peer-to-peer or HPC, at least that is not the authors original intention when they develop this simulator. Third, the available features of the simulators are tuned towards cloud computing applications. This limits its versatility.

The fine grain details provided by GreenCloud and the packet-level processing of the underlying NS-2 simulator is advantageous for getting accurate result when simulating relatively small networks. However, for large-scale distributed networks, it is not scalable. In related to this, the authors have mentioned that their solution gets slower and slower as the number of simulated nodes increases beyond few thousands and as the number of packets processed increases. We expect the memory footprint also to increase as the processed packets and the number of simulated nodes increases, even though the authors did not say anything about it.

Chapter 3

Environment

In this study we employ SimGrid and NS-3 simulators. SimGrid is the simulator that we want to extend

Chapter 4

Methods

Description of energy estimation models found in literature

Which model is more appropriate for SimGrid? Why?

How to compare the implementation of models on SimGrid with other packet-level simulators or measurements found in literature

Experiment design to validate the implementation with other simulators

Chapter 5

Implementation

Here I describe about SimGrid implementation of energy consumption models.
The challenges encountered
How we tackled the challenges
The limitation of the implementation

Chapter 6

Evaluation

After conducting experiments both on SimGrid and other Simulators, here we compare the result

Chapter 7

Discussion

At this point, you will have some insightful thoughts on your implementation and you may have ideas on what could be done in the future. This chapter is a good place to discuss your thesis as a whole and to show your professor that you have really understood some non-trivial aspects of the methods you used…

Chapter 8

Conclusions

Time to wrap it up! Write down the most important findings from your work. Like the introduction, this chapter is not very long. Two to four pages might be a good limit.

Bibliography

[1] Barakat, C., Thiran, P., Iannaccone, G., Diot, C., and Owezarski, P. Modeling Internet backbone traffic at the flow level. *IEEE Trans. Signal Processing 51*, 8 (2003), 2111–2124.

[2] Barroso, L. A., Clidaras, J., and Hölzle, U. *The Datacenter as a Computer: An Introduction to the Design of Warehouse-Scale Machines, Second Edition.* Synthesis Lectures on Computer Architecture. Morgan & Claypool Publishers, 2013.

[3] Barroso, L. A., and Hölzle, U. The case for energy-proportional computing. *IEEE Computer 40*, 12 (2007), 33–37.

[4] Bolla, R., Bruschi, R., Davoli, F., and Cucchietti, F. Energy efficiency in the future internet: A survey of existing approaches and trends in energy-aware fixed network infrastructures. *IEEE Communications Surveys and Tutorials 13*, 2 (2011), 223–244.

[5] Carneiro, G., Fortuna, P., and Ricardo, M. Flowmonitor: A network monitoring framework for the network simulator 3 (NS-3). In *4th International Conference on Performance Evaluation Methodologies and Tools, VALUETOOLS '09, Pisa, Italy, October 20-22, 2009* (2009), p. 1.

[6] Casanova, H., Giersch, A., Legrand, A., Quinson, M., and Suter, F. Versatile, scalable, and accurate simulation of distributed applications and platforms. *J. Parallel Distrib. Comput. 74*, 10 (2014), 2899–2917.

[7] Claffy, K., Miller, G., and Thompson, K. The nature of the beast: Recent traffic measurements from an Internet backbone. In *Proceedings of INET* (1998), vol. 98, pp. 21–24.

[8] Dayarathna, M., Wen, Y., and Fan, R. Data center energy consumption modeling: A survey. *IEEE Communications Surveys and Tutorials 18*, 1 (2016), 732–794.

[9] Fan, X., Weber, W., and Barroso, L. A. Power provisioning for a warehouse-sized computer. In *34th International Symposium on Computer Architecture (ISCA 2007), June 9-13, 2007, San Diego, California, USA* (2007), pp. 13–23.

[10] Fiandrino, C., Kliazovich, D., Bouvry, P., and Zomaya, A. Y. Performance metrics for data center communication systems. In *8th IEEE International Conference on Cloud Computing, CLOUD 2015, New York City, NY, USA, June 27 - July 2, 2015* (2015), pp. 98–105.

[11] Gyarmati, L., and Trinh, T. A. How can architecture help to reduce energy consumption in data center networking? In *Proceedings of the 1st International Conference on Energy-Efficient Computing and Networking, e-Energy 2010, Passau, Germany, April 13-15, 2010* (2010), pp. 183–186.

[12] Heddeghem, W. V., Lambert, S., Lannoo, B., Colle, D., Pickavet, M., and Demeester, P. Trends in worldwide ICT electricity consumption from 2007 to 2012. *Computer Communications 50* (2014), 64–76.

[13] Kliazovich, D., Bouvry, P., and Khan, S. U. Greencloud: A packet-level simulator of energy-aware cloud computing data centers. *The Journal of Supercomputing 62*, 3 (2012), 1263–1283.

[14] LLC, M. NS-3 simulator, 2017.

[15] Mahadevan, P., Sharma, P., Banerjee, S., and Ranganathan, P. A power benchmarking framework for network devices. In *NETWORKING 2009, 8th International IFIP-TC 6 Networking Conference, Aachen, Germany, May 11-15, 2009. Proceedings* (2009), pp. 795–808.

[16] Orgerie, A., Lefèvre, L., Lassous, I. G., and López-Pacheco, D. M. ECOFEN: An end-to-end energy cost model and simulator for evaluating power consumption in large-scale networks. In *12th IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks, WOWMOM 2011, Lucca, Italy, 20-24 June, 2011* (2011), pp. 1–6.

Appendix A

First appendix

This is the first appendix. You could put some test images or verbose data in an appendix, if there is too much data to fit in the actual text nicely.

For now, the Aalto logo variants are shown in Figure A.1.

Aalto University
School of Science

(a) In English

Aalto-yliopisto
Perustieteiden
korkeakoulu

(b) Suomeksi

Aalto-universitetet
Högskolan för
teknikvetenskaper

(c) På svenska

Figure A.1: Aalto logo variants