Energy Efficiency in Telecommunication Networks

(Invited)

W. Vereecken[1], L. Deboosere[1], D. Colle[1], B. Vermeulen[1], M. Pickavet[1],
B. Dhoedt[1], P. Demeester[1]

[1] IBCN, Dept. of Information Technology (INTEC), Ghent University, Ghent, Belgium
Tel: +32 9 331 49 00, Fax: +32 9 331 48 99, E-mail: ibcn-info@intec.ugent.be

Today's ICT provides many environmentally friendly solutions. Some typical examples are:

- o Providing alternative solutions for flights and road traffic by tele-working, phone and video conferencing, etc.
- o Sensoring systems allowing to reduce/optimize the heating in buildings

However, ICT on its own also represents a major energy consumption factor. And this energy footprint is expected to grow significantly in the coming years, mainly driven by the steeply growing file sizes and information flows.
The goal of this paper is to point out the necessity of more energy efficient solutions and present the thin client concept as such a solution.

1. Introduction

The current image of ICT is rather environmentally friendly. This is largely correct since the worldwide communication via datacom and telecom networks has transformed society and created opportunities to reduce the human impact on nature.

There is however a down side to ICT. The ubiquitousness of ICT in daily life (both private and professionally) has the drawback that the energy consumption of computers and network equipment is a significant part of the global energy consumption. It is to be expected that this share will largely increase in the coming years. Together with a growing energy price (due to shortage in fossil fuels) and the increasing awareness of the green house effect which will be translated in government policies the energy footprint of ICT will very soon be under pressure and will stimulate the demand for energy efficient solutions.

It is necessary to design new network paradigms so we can maintain the same level of functionality ICT provides but with a lower power consumption impact. In this paper we will display thin clients as a possible energy efficient alternative. In section 2 we first give an overview of the different factors that make up the energy consumption of ICT. In section 3 we elaborate a typical example of the use of thin clients. The paper is summarized with some general conclusions in section 4.

2. Worldwide energy needs

When assessing the impact of ICT on the worldwide energy production and consumption it is crucial to get a good overview of the major energy consumption factors. Firstly we give an overview of the energy consumption in the use phase of different equipment types. Secondly we give an estimate for the manufacturing phase. Finally we try to forecast the electricity consumption in the coming years.

2.1 Electricity consumption during use

When estimating the total electricity consumption the following five categories were distinguished:

1. Data centers: Servers, storage devices and network equipment, but also: cooling, backup power infrastructure (e.g. UPS systems) …
2. PCs
3. Network equipment (excluding network equipment inside data centers and PCs)
4. TV sets (including video and DVD players)
5. Other ICT equipment: All equipment not contained in the first 4 categories. (audio equipment, telephone handsets, gaming consoles, printers …)

Based on various sources we estimated the following numbers given in Table 1. The total power consumed by ICT is about 156 GW which is more than 8% of the global electricity consumption.

Equipment Type	Est. Consumption 2007 (GW)	Est. Annual growth rate
Data centers	26	12%
PCs	28	7.5%
Network Equipment	22	12%
TVs	40	9%
Others	40	5%
Total	156	

Table 1: Energy consumption for various equipment types

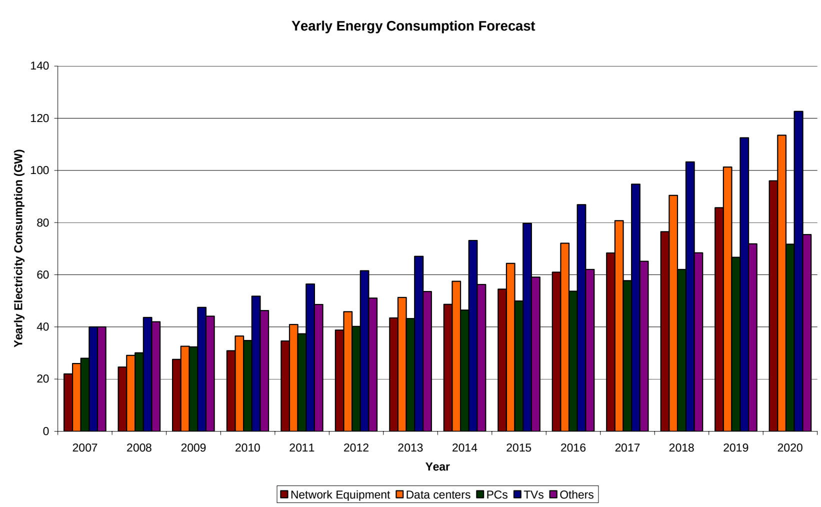

Figure 1: Electricity consumption forecast of ICT equipment during use.

When taking the annual growth rate into account we can forecast how the numbers will evolve in the next years. This is summarized in Figure 1. One sees that the power consumption is growing from 156 GW in 2007 to 430 GW in 2020. If we assume an energy consumption growth of 3% for all other equipment [1] this will result in a relative contribution of 14% of ICT to the worldwide energy consumption in 2020. Note that this is even excluding the manufacturing energy cost.

One should also take into account that the consumed electrical energy needs to be produced in a power plant. Currently the limited yield factor of power plants implies that 1J of electrical energy corresponds (on average) with 2.5J of primary energy.

2.2 Energy consumption for manufacturing of ICT equipment

Besides the electricity consumption in the use phase of the equipment one needs to take the manufacturing process into account as well. In [4] a modeling exercise has been carried out in detail leading to an average of 1550 MJ of electrical energy and 4850 MJ of non-electrical primary energy sources to manufacture a typical PC configuration.

With the electricity production yield factor of 40% the 1550 MJ of electrical energy requires 3875 MJ of primary energy. This leads to a total of about 8700 MJ per produced PC.

When assuming an economical lifetime of 4 years for a PC the average energy consumption during the use phase is about 8800 MJ. This means that the energy needs during the manufacturing phase are comparable to the needs during the use phase. This conclusion depends obviously on the type of equipment. More specifically it depends on the energy intensity of the manufacturing process and the expected economical lifetime. It will be important to take these factors into consideration when designing new energy efficient solutions.

3. Case study: Thin Clients

In this section we will demonstrate that when using thin clients power-efficiency can be increased. In this section we investigate a typical example of where thin clients can reduce the power consumption. We consider two scenarios. On the one hand we have a traditional desktop where each user is running a standalone application on a standard PC. In the second scenario the desktops are replaced with thin client terminals. The standalone applications run remotely on servers in the data center. In order to keep the complexity of the model under control we will consider the users and the data center to be in the same building. The network consists of a single switch.

It is already clear that the thin client scenario has a number of pros and cons when compared to the desktop scenario:

+ Power consumed by the thin client terminal is significantly lower

+ Server side resources can be delivered more efficiently: high-end servers are shared between all users, implying that the server infrastructure will exhibit less idle periods (the typical load on desktops is below 20%).

- Applications are run remotely, implying possibly application specific network overhead (e.g. for sending input events to the server and getting screen updates back). Additional equipment is needed (e.g. switch at server side, network interface cards consuming power …).

- Protocol overhead from the thin client protocol requires additional server side processing

- Resources at the server side must be cooled, increasing the power budget for the thin client scenario.

Given these observations, one can already conclude that the balance for the thin client paradigm will certainly depend on the following factors: application bandwidth (assuming power scales with consumed bandwidth, it is clear that high bandwidth applications, such as e.g. multimedia editing will benefit less), server resource efficiency (influenced by the achievable amount of sharing and optimal resource usage) and server side cooling efficiency.

3.1 Desktop Scenario

In the desktop scenario we consider a normal desktop PC. We assume a standalone application. That means the network card is unloaded. Therefore the Power consumption P_{tot}^{d} only depends on the CPU load λ_{CPU}^{d}. In practice this relation appears to be linear. We get:

$$P_{tot}^{d} = P_{0}^{d} + \alpha_{CPU}^{d}\lambda_{CPU}^{d} \quad (1)$$

3.2 Thin client scenario

The thin client terminal

A thin client terminal typically behaves like a desktop without a hard drive. However, the power consumption appears to be constant even with varying CPU load λ_{CPU}^{c} and NIC load λ_{NIC}^{d}. Therefore we the power consumption is reduced to:

$$P_{tot}^{c} = P_{0}^{c} \quad (2)$$

The server farm

A server also behaves like a desktop. The power consumption is now also determined by λ_{NIC}^{s}. Assuming again linear dependencies this leads to:

$$P_{tot}^{s} = P_{0}^{s} + \alpha_{CPU}^{s}\lambda_{CPU}^{s} + \alpha_{NIC}^{s}\lambda_{NIC}^{s} \quad (3)$$

The load λ_{NIC}^{s} is in reality the bandwidth received by the server b^{s}. We express this bandwidth as a function of the bandwidth b caused by the client. When assuming a share ratio of N users per server we get:

$$\lambda_{NIC}^{s} = b^{s} = Nb \quad (4)$$

Obviously, the load on the server λ_{CPU}^{s} is related to the load on the clients. The amount of work to be performed by a single server is at least the amount of work done by N desktops. On the other hand, there is the extra work needed on the server to support N sessions, and processing the protocol overhead (to receive input from the thin clients and to construct and send back screen updates). If we note the capacity of a server (according to a relevant performance oriented benchmark such as SPECint2000) as C^{s} and the analogous parameter for the desktop case C^{d} we have

$$\lambda_{CPU}^{s}C^{s} > N\lambda_{CPU}^{d}C^{d} \quad (5)$$

By denoting the extra load caused per user by ε, we have

$$\lambda_{CPU}^{s} = N\left[\lambda_{CPU}^{d}\frac{C^{d}}{C^{s}} + \varepsilon\right] \quad (6)$$

When assuming a large number of users, it is safe to assume that λ_{CPU}^{d} represents the average load on the desktops. We want to use as little servers as possible so we load them to their maximal capacity. This leads to λ_{CPU}^{s} being 1. This means:

$$N = \left[\lambda_{CPU}^{d} \frac{C^{d}}{C^{s}} + \varepsilon \right]^{-1} \tag{7}$$

$$P_{tot}^{s} = P_{0}^{s} + \alpha_{CPU}^{s} + \alpha_{NIC}^{s} Nb \tag{8}$$

The network

In this case (i.e. single business thin client environment) we consider the network to consist of one switch connecting all thin client terminals and servers. The power consumed by the switch consists of three parts. The unloaded power consumption P_{0}^{n}. The power consumption caused by the $(N_u + N_s)$ interconnections in unloaded state. And the power consumption caused by the traffic on these connections caused by N_u users consuming a bandwidth b. This leads to:

$$P_{tot}^{n} = P_{0}^{n} + \alpha_{C}^{n} (N_u + N_s) + \alpha_{T}^{n} N_u b \tag{9}$$

By denoting N_S as a function of the load factor N we get:

$$P_{tot}^{n} = P_{0}^{n} + \alpha_{C}^{n} N_u (1 + 1/N) + \alpha_{T}^{n} N_u b \tag{10}$$

Cooling

Due to the concentration of heat dissipating equipment, considerable efforts are needed to cool data centers. This cooling infrastructure of course also consumes electrical power. Therefore, if P_{ICT}^{dc} denotes the total electrical power dissipated in the data centre by ICT-equipment, the total electrical power used in the data centre is given by

$$P_{tot}^{dc} = (1 + \eta) P_{ICT}^{dc} \tag{11}$$

In this paper, we assume that the similar cooling efforts are not needed at the client side, as concentration of ICT-equipment is not an issue there. This means:

$$P_{ICT}^{dc} = N_s P_{tot}^{s} + P_{tot}^{n} \tag{12}$$

Total power consumption

The total power consumed per thin client in the thin client scenario is the power consumed by a thin client terminals and the power consumed in the data centre divided over the total number of users N_u.

$$P_{tot}^{tc} = P_{tot}^{c} + \frac{P_{tot}^{dc}}{N_u} \tag{13}$$

When incorporating (11) & (12) this results in

$$P_{tot}^{tc} = P_{tot}^{c} + \frac{1+\eta}{N_u} P_{tot}^{n} + \frac{1+\eta}{N} P_{tot}^{s} \tag{14}$$

3.3 Equilibrium

In this section we will evaluate both scenarios in terms of energy efficiency. The thin client scenario becomes favorable when

$$P_{tot}^{tc} < P_{tot}^{d} \tag{15}$$

Therefore we have an equilibrium when

$$P_{tot}^{tc} = P_{tot}^{d} \tag{16}$$

Using (1), (2), (7), (8) and (10) we get the equilibrium. We would like to know for which minimal number of users the use of thin clients becomes feasible. This number is expressed by:

$$N_{u,\min} = \frac{(1+\eta)P_0^n}{\left(P_0^d - P_0^c - (1+\eta)\alpha_C^n\right) + \alpha_{CPU}^d \lambda_{CPU}^d - (1+\eta)\left(P_0^s + \alpha_{CPU}^s + \alpha_C^n\right)\frac{1}{N} - (1+\eta)\left(\alpha_T^n + \alpha_{NIC}^s\right)b} \tag{17}$$

3.4 Experimental results

We measured the power consumption of a desktop (AMD Athlon 64 3500+), a server (AMD Opteron 2212), a thin client device (JackPC) and a switch (Force10 E1200). A typical value for η is 0.4. N is given by (7). The sensitivity of (17) to b is a small compared to the sensitivity to the other two parameters. Therefore we assumed a typical bandwidth of $5\,Mb/s$ [7]. The switch contains 12 48-port line cards. This implies that the number of users is limited to:

$$N_u \le 576 - N_s \tag{18}$$

Equations (17) and (18) give a minimal and a maximal number of users. The results are summarized in Figure 2.

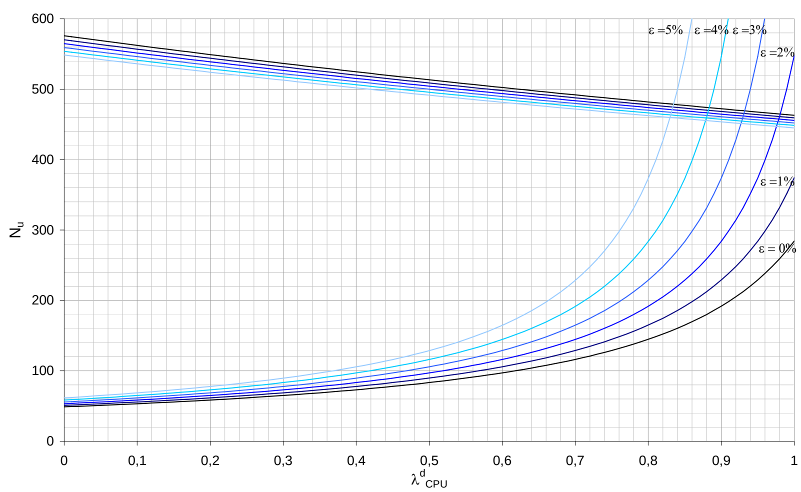

Figure 2: Impact of load and server overhead in thin client scenario

One can see that for low load the thin client scenario becomes advantageous as of 50 to 80 users. Taking into account that more than 500 users can be served and an average load of 20% is sufficient for desktop applications such as text editors and spreadsheets a thin client scenario will definitely enable power savings. It is also clear that the overhead load on the thin client server is an important impact factor. It will be crucial for implementations to keep this overhead as low as possible.

We can now project the power savings that can be achieved in the thin client scenario. We consider the gained power $\Delta P = P_{tot}^{d} - P_{tot}^{tc}$ and the ratio $P_{tot}^{d} / P_{tot}^{tc}$ as a function of the number of users. We consider the server overhead to be low. The results are depicted in Figure 3 and Figure 4. It is clear that thin clients can be up to 400% more efficient. This is however impaired by the load. This is due to the share ratio *N* becoming less advantageous with increasing load. Note that when implementing thin clients in a larger scale network the traffic cost of the network will increase and this will also reduce the energy efficiency.

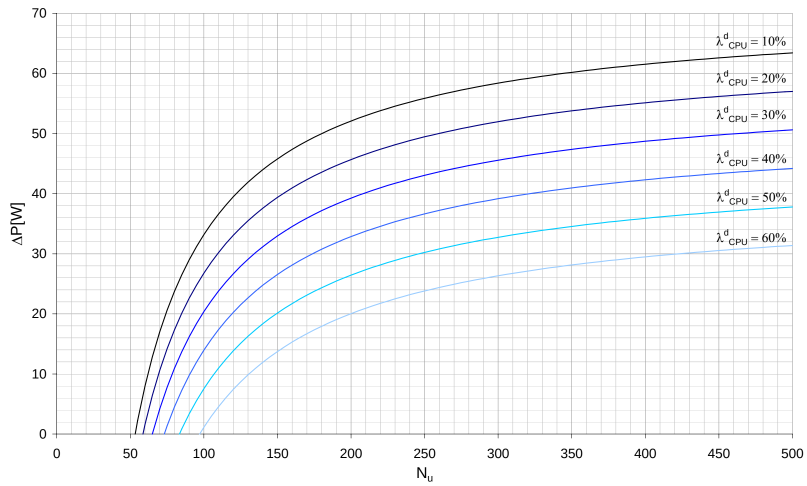

Figure 3: Power saved with thin clients

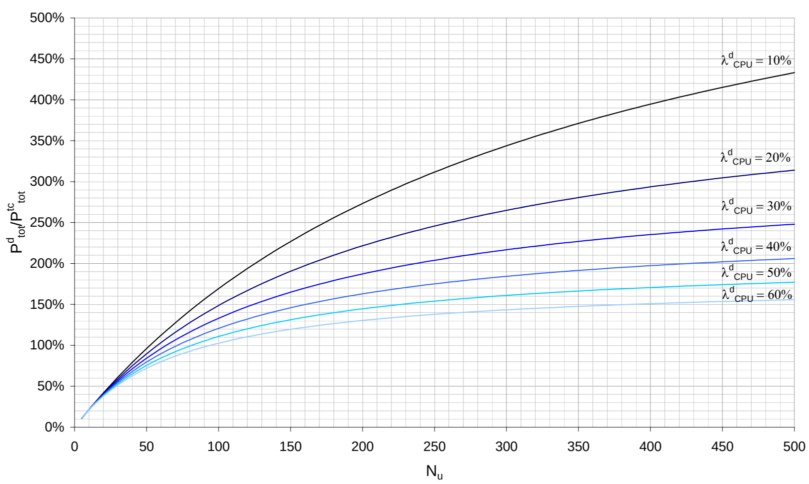

Figure 4: Power saving ratio

Currently we can only present the first results we got from our experiments. In our presentation we will be able to give more extensive figures.

4. Conclusion

In this paper we studied the future worldwide energy consumption of ICT. For 2007 the consumption during use is about 8% of the overall consumption. Without considerable measures this will grow to about 20% in 2020. These numbers do not include the high energy needs for the manufacturing of ICT equipment. Today these needs are in the same order of magnitude as the consumption during use.

In the case study we studied the possibility of thin client computing as a new network paradigm with higher energy efficiency. Thin client computing has some important advantages but also drawbacks. Nevertheless it is a promising energy efficient alternative that should be further investigated.

Acknowledgements

The work described in this paper was carried out with the support of the BONE project ("Building the Future Optical Network in Europe), a Network of Excellence funded by the European Community‘s Seventh Framework. Part of the research leading to these results was done for the MobiThin Project and has received funding from the European Community‘s Seventh Framework (FP7/2007-2013) under grant agreement nr 216946.

References

[1] Energy Information Administration: “Official Energy Statistics from the U.S. Government”, http://www.eia.doe.gov/.

[2] J. Koomey, “Estimating Total Power Consumption by Servers in the U.S. and the World”, February 2007,http://enterprise.amd.com/Downloads/svrpwrusecompletefinal.pdf.

[3] C. Cremers et al., “The Impact of Information and Communication Technologies on Electricity Consumption in Germany”, *ECEEE Summer Study*, ISBN 91-631-4001-2, pp. 455-466 St-Raphaël, France, June 2003.

[4] E. Williams, “Energy Intensity of Computer Manufacturing: Hybrid Assessment Combining Process and Economic Input-Output Methods”, *Environmental Science & Technology*, Vol. 38, No. 2, pp. 6166-6174 November 2004.

[5] P. Simoens, L. Deboosere, D. De Winter, F. De Turck, B. Dhoedt, P. Demeester, “Optimization Models for Application Migration to Support Mobile Thin Clients”, *Lecture Notes in Computer Science,* the 3rd International Workshop of the EURO-NGI Network of Excellence, ISSN 0302-9743, p. 255-269, Sitges, Spain, June 2006.

[6] Lawrence Berkeley National Laboratory, Project ‘Energy Efficient Digital Networks’, February 2007, http://efficientnetworks.lbl.gov/enet.html.

[7] L. Deboosere, J. De Wachter, P. Simoens, F. De Turck, B. Dhoedt, P. Demeester, "Thin client computing solutions in low- and high-motion scenarios", *Proc. ICNS 2007*, Athens, Greece, 19-25 June 2007.