Aalto University
School of Engineering

Green Cloud Computing

T-110.5121

Lecture 31.10.2012

Tommi Mäkelä
tommi.makela@aalto.fi

Aalto University
School of Engineering

Outline

Current situation

Green IT equipment

Green infrastructure

Green cloud computing

Tools to measure greenness

http://mpcuae.com/wp-content/uploads/2011/11/Green.jpg

Introduction

- Information and communication technology (ICT) industry utilizes a large variety of equipment, such as workstations, servers, cell phones and printers.
- It also requires different types of infrastructures including datacenters and telecommunications.
- Designing, manufacturing, operating and recycling devices cause a substantial amount of emissions.
- In 2007 ICT industry produced two percent of the worldwide carbon dioxide (CO2) emissions.
  - As a comparison, aviation had very similar figure

Estimations

- Servers and datacenters are the backbone for the internet and essential for many ICT businesses.
- However, their overall electricity consumption has rather alarming pattern to increase regularly.
  - In year 2000 servers consumed 29.2 terawatt-hours of electricity and the consumption more than doubled to 61.4 terawatt-hours in 2005. The higher quantity of volume server explains most of the increased figure. The same pattern has been expected to continue during every five years.
  - During the same period of time the total electricity drain of datacenters increased from 70.8 to 152.5 terawatt-hours world widely and the annual growth rate was around 16.7%.

Estimations

- Because of the major recession in year 2008, the power draw growth rate slowed down substantially between years 2005 and 2010. Moreover, adapted virtualization technologies and other optimization measures reduced need for electricity. The electricity consumption of data centers increased around 56% world-widely whereas the same value doubled in 2000-2005.
- Nevertheless, still rather substantial growth over five years. Overall electricity consumption reached 250 terawatt-hours.
- Around 1.0-1.3% of all electricity consumption world widely.
- It has been also estimated that data centers will consume average 15% more in 2016, when compared to year 2012. Major facilities, such as cloud data centers, will drain as much as 40% more electricity in the same interval.

Motivation

- In the United States between years 2005 and 2006, business growth almost tripled the electricity bill in the IT companies. Moreover, electricity prices have created financial pressure to the IT companies in many other countries as well, including Finland, where prices have grown almost constantly during the past ten years
- Due to the increasingly high electricity prices and affordable server equipment, it takes less than three years that operational costs exceed all the initial costs.
  - Green computing strives for lower operational expenses
  - Assumingly customers appreciate greenness as well

Green data center

- Important question: what is greenness in practice?
- There is no specific definition for a green data center. However, the widely adapted definition for a green building that can be applied to green data centers
  - *Designing a building so that it uses resources, including energy, water, material, in a more efficient manner and has less impact upon people and the environment* – *Douglas Alger*
  - How much energy is consumed, natural resource required and emission released during the construction, operational and demolition phases compared to convention data centers
- These slides focuses only on energy consumption

Aalto University
School of Engineering

Outline

Current situation

Green IT equipment

Green infrastructure

Green cloud computing

Tools to measure greenness

http://mpcuae.com/wp-content/uploads/2011/11/Green.jpg

Server components

- Because of the low initial costs compared to the mainframe approach, servers are optimized to provide agility rather than efficiency.
- Even if servers operate under low workload, they tend to consume 70-90% of maximum power draw.
- Majority of electricity is consumed by processor units
  - Energy-saving can be achieved by downgrading capacity
  - Dynamic voltage frequency scaling (DVFS), for instance, allows automatically adjusting the frequency of a processor based on workload. As a result, lower voltage can be used within a processor, which declines the need for electricity.

Server components

- Main memory in a server configuration drains around 11% of total power draw. The dominant memory technology is based on double data rate memory (DDR)
  - The previous generation requires 1.8 voltages whereas currently applied DDR3 memories use only 1.5 voltages.
  - Manufactures have also released their own solutions, which utilize even lower voltages than DDR3 standard specifies. For instance, memory with 1.35V consumes 15-20% less.
  - Even though a main memory requires relatively small portion of electricity compared to processors, modern servers seem to have increasingly more main memory capacity to serve memory-intensive databases and host virtual environments.

Server components

- Storage equipment consume approximately 5% of total electricity within data centers.
  - Some companies might have even higher figure.
  - Local storage for operating systems on servers.
  - Network storage for user data and backup data.
  - Typically storage equipment is based on hard disk drives (HDD), but tape drives seem be still quite common on backup systems.
  - To reduce electricity consumption, current disk solutions could be replaced with next generation solid state drives (SSD).
  - They require around 50% less electricity than equivalent high-performance HDD with 15 000 RPM spinning disks.

http://www.victorialaptoprepair.co.uk/image/hard-drive-repair.jpg

Blade servers

- Blade servers provide computation in compact form, because they consist of high-density components
  - For instance, processor models with numerous cores, a few high capacity main memories, reduplicated 2.5” hard drives
- Blade servers are installed into a specially designed chassis with several slots and supporting modules
  - Cooling, network and power supply modules are shared among blade servers, which only have computation and storage related components. Module design makes it possible to assemble fault tolerant and versatile configu-rations. For instance, customer can choose which type of network modules are applied: Fiber Channel, Infiniband…

Blade servers

- Blade servers are commonly attached to a mid-panel
  - Transfers internal data directly between servers and provides access to network modules as well as distributes power
  - Due to the prewiring of a mid-panel and network modules with Fiber Channel, there is only few cables blocking air flow at the back of a chassis, which helps to remove hot exhaust air
- Blade servers have different cooling implementations
  - Small independent fans within every blade servers
  - Large shared fans at the back of a chassis, which probably require space from other modules and offer no energy-savings
  - Some manufacturers have optimized the shapes of fan blades and chambers to create strong airflow and high pressure

Blade servers

- Blade servers could be considered as green
  - Majority of power supplies improve their efficiency rate, when the load increases. For instance, a typical rack server has a supply, which operates at 20% load and has only 60 efficiency rate. **Blade servers with share power supplies** could easily reach optimal 50% load and 94% efficiency, because configu-rations have so few power supplies. As a result of high efficiency, power supplies generate less waste heat.
  - To guarantee high efficiency, some manufacturers also provide **power management features**, which dynamically switch off a couple of power supplies to increase load on remaining supplies. However, over fifty percent load would problematic on the reliability point of view.

Blade servers

1. A single blade server
2. A common blade chassis
3. A shared power supply unit

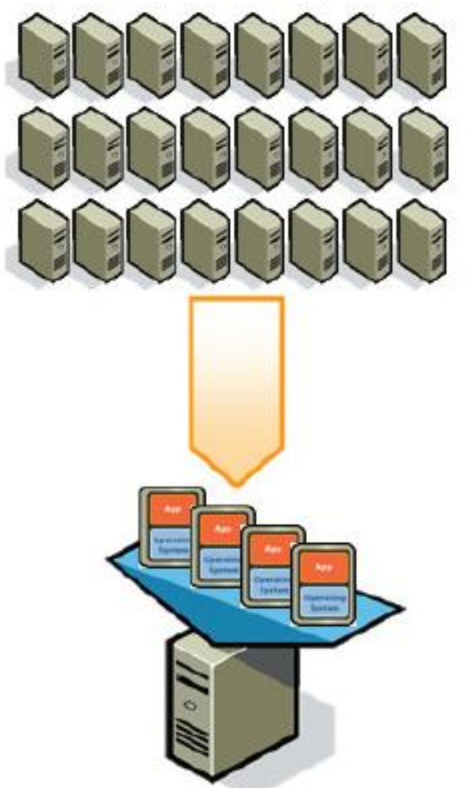
http://bladesmadesimple.com/wp-content/uploads/2010/03/c7000w16.jpg

- Blade systems
  - In this configuration, 16 individual blade servers are placed to a common chassis, which provides network connections and cooling system. However, only six power supplies are responsible for con-verting enough alternating current to direct current.

Virtualization

- Traditionally servers have only one operating system, which hosts only few services, such as email daemon
- Virtualization makes it possible to run multiple operating systems simultaneously on the same server
  – Logical abstraction on top of the physical resources
  – Logical names are assigned for each server component
  – Efficient management and dynamic resource mapping
- Consolidating several underutilized servers into a single device reduces the need for cooling, saves floor space and most importantly declines electricity consumption.

Virtualization



http://ituda.com/wp-content/uploads/2010/01/Physical_to_Virtual.bmp

- As a result of intensive consolidation, the average utilization of a single server equipment within a data center reaches 70-80%, while servers without virtualization use only 10-15% of their resources.
- According to Vmware, the combination of declined need for cooling and reduced number of physical resources drops the overall electricity consumption of servers by 80%.

Network equipment

- Network infrastructure, including routers and switches, accounts for 5-10% of total electricity consumption within data centers. However, this portion has been expected to rise, because enterprises have increasingly adopted energy intensive high-speed networks.
- The traditional topology of data center network typically consists of three-level hierarchy, with an unique set of network devices, including edge switches, aggregation switches and core switches, where the latter are the most energy-intensive devices.

Network equipment

- Measurements to reduce electricity consumption
  - Typically devices with zero utilization consume up to 70-80% of maximum power draw. Adopt advanced switches with power management features, which turn off idle components, such as ports and line cards, and lower ports speed during low traffic.
  - High-level systems with power-aware routing algorithms allow to forward traffic via most desirable routes and shut down unused and unnecessary edge switches.
  - Replace traditional network topology with alternative.
    Fat-tree: mainly edge switches and plenty of cabling
    Elastic-tree: combination of Fat-tree and power management. It has 50-70% lower electricity consumption than traditional.

Aalto University
School of Engineering

Outline

- Current situation
- Green IT equipment
- Green infrastructure
- Green cloud computing
- Tools to measure greenness

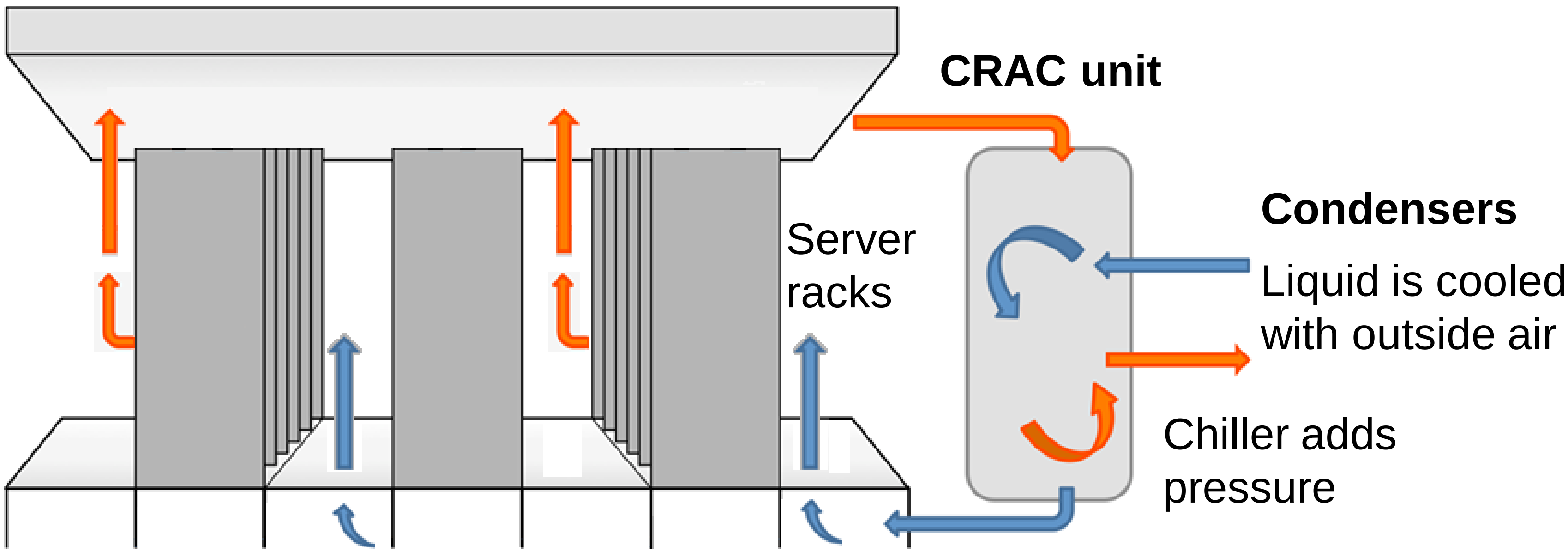
http://mpcuae.com/wp-content/uploads/2011/11/Green.jpg

Air cooling

- Computer Room Air Conditioning (CRAC)
  - The basic components of a single unit are fans and a circulation with liquid refrigerant, which low temperature is maintained by using a chiller and its energy-intensive compressor.
  - The warm indoor air is re-chilled by sucking it inside a CRAC unit, where it is directed pass a circulation or coil, which results a heat interchange. In addition to cooling hosted devices, units are used to maintain desirable humidity within facilities.
  - Commonly data centers have a raised floor, in which case CRAC units are used to create pressure underneath the floor, where the cool air is forced upwards to cold aisles though ventilation tiles in front of racks full of servers -> next slide.

Air cooling

- The basic structure of an air cooled data center
  - CRAC unit operates very similarly as a common refrigerator



http://www.thehotaisle.com/wp-content/crac.png Raised floor

Air cooling

- Measurements to save electricity
  - Measure temperatures near the servers, not inside the CRAC unit. In many cases operate at maximum capacity unnecessarily.
  - Use CRAC units with **adaptive fans**, which are able adjust their speed based on temperatures. 50% less electricity required.
  - Improve **distribution** of chilled air by placing CRAC units evenly across the room. Moreover, prevent mixing chill and warm air to each other by isolating aisles, where the chilled air is forwarded.
  - Enhance **heat removal** by installing return ducks on top of the aisles, where the server exhaust their hot air. Ducks allows to suck hot air directly back to the CRAC units for re-chilling.

Liquid cooling

- The first water cooled solution was introduced as early as 1964, when IBM as released its mainframe solution.
- In 1990 client-server model became the dominant method to provide services and at the same time air cooling was the most cost efficient method.
- The heat density of IT equipment has constantly raised and the limits of air cooling has been almost reached and as a result water cooling is not obsolete anymore.
- Despite its name, the refrigerant within circulations is commonly Freon or glycol rather than water.

Liquid cooling

- Water cooling allows exchanging heat near IT devices.
- Even though cooling is based on liquid, air is commonly used as a medium to transfer heat within closed racks, because conventional servers are cooled by using air.
- Liquids as a medium have higher thermal conductivity than air and therefore liquid cooling is able to cool down high-density IT equipment energy-efficiently.
- A research work points out that a data center with liquid cooling requires 62% less electricity than the facility with CRAC units even though both use similar chillers.

Liquid cooling

http://h10003.www1.hp.com/digmedialib/prodimg/lowres/c00582514.jpg

- If the expected lifespan of a typical data center is around fifteen years, total savings of liquid cooling are significant.
- Liquid related equipment have rather high capital investments and therefore it will take a couple of years to amortize initial costs.

1. A stack of servers
2. A heat exchanger
3. A liquid circulation

Economizer

- When outside temperature and humidity are suitable, an economizer makes it possible to cool a data center by using only low temperature ambient air and as a result highly energy-intensive chillers are no longer needed.
- Two types of economizers
  - Normally air is constantly re-chilled and recirculated inside the facility. An **airside economizer** transfers filtered outside air into premises and mixes cold outside air with hot inside air.
  - When liquid circulation is a part of cooling system, a **waterside economizer** can be used to improve energy efficiency. It utilizes cooling towers to reject heat through evaporation.

Economizer

- Airside economizer
  - Typically a data center infrastructure is designed to protect IT equipment from outside conditions, including extreme temperatures, high humidity and dirty air, which would potentially harm hosted servers. In contrast to conventional design, an air **economizer directly blows** cool outside air into premises.
  - Intel has conducted an experiment, where a data center was set up to New Mexico. It was split into two compartments and different cooling methods were measurements: conventional and airside economizer. Tests revealed that the compartment with the economizer required 74% less energy than the conventional configuration. Nevertheless, its 4.46 failure rate was almost twice as high as the other compartment's rate.

Economizer

- Waterside economizer
  - It rejects heat from a liquid circulation by using a large **cooling tower**, where some of the water evaporates and exits during the process. The minerals and particles of the evaporated water shift to remaining water and start to pile up the circulation in the long-term. Both evaporated and dirty liquid need to be replaced with fresh water, which is preferred as makeup water. Large scale data centers might monthly evaporate millions of liters of water, in which case an economizer and its cooling towers descend into a sort of environmental problem.
  - It becomes significantly greener, when recycled water from a local water treatment plant is used instead of fresh water.

Geothermal cooling

- Facility in a cave
  - The surrounding bedrock act as a large heat sink.
  - Physical protection

http://www.wired.com/news/images/full/gypsum_cave1_f.jpg

- Embed pipe system
  - Horizontally or vertically installed underground pipes are used to chill refrigerant, which loops though a **heat exchanger**, which then cools the circulation within a data center.
  - Because an energy-intensive compressor is no longer required and none of the liquid is evaporated during the process, a geothermal cooling could be considered as a green method to cool down data center facilities and hosted servers.

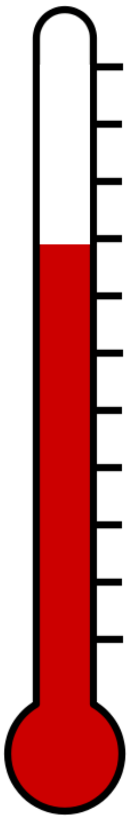
River and sea cooling

- Closed pipe system
  - In this time, pipes are placed bottom of the river or sea
  - Pipe system needs to be installed deep enough so that sea cooling can be utilized a whole year around. In the Baltic, for example, over 40 meters deep areas should be constantly below 8 degrees Celsius and suitable for sea cooling.
  - Powerful pumps needed, no energy-intensive chillers.
  - Ideal solution with liquid cooling within data centers. Naturally sea water is never mixed with circulating liquid.
  - River cooling very likely requires other supporting solution during summer. For instance, airside economizer could be applied, if outdoor air is commonly cooler than nearby river.

Brief summary

- Air and liquid cooling are methods to exchange heat **within data centers**. Both of these methods typically utilize liquid circulation to transfer heat to outdoors.
- Chillers with compressors are used to accelerate heat extraction within condensers on top of the facilities.
- Some configurations bypass chillers during winters.
- Alternatively, cooling towers with open secondary circulation can cool down the primary circulation.
- Closed secondary circulation beneath ground or bottom of the sea can chill the primary circulation.

Temperature



- Based on the real-life experiences as well as the theories of natural science, the performance of IT equipment improves under cool conditions, because the resistance within components reduces at low temperatures.
- Components also wear down slower, which potentially increase the lifespan of server investments and save capital in the long-term.
- However, a cool data center environment uses more electricity than rather warm environments.

http://openclipart.org/image/800px/svg_to_png/1309/twb_thermometer.png

Temperature

- Nowadays IT equipment should tolerate a temperature range of 18-27°C, where the top temperatures would offer the lowest electricity consumptions.
- A portion of these benefits are gained because of chillers. Raising the temperature of circulating liquid by 1°C increases the efficiency of chillers by 2-4%.
- The best way to gain energy-savings is to incrementally raise the temperature and monitor possible consequen-ces. Eventually, it is up to IT personnel themselves to find the most optimal temperature for their facilities.

Waste heat

- A green data center strives for minimal heat production. However, even greener facilities also try to utilize waste heat, because it offers two major advantages related to carbon dioxide emissions and operational costs.
- Data centers require less energy for cooling, when the waste heat is transferred to premises, such as office buildings, which then require less energy for warming.
  - In Israel, waste heat is used to warm up 11-floor office building. This arrangement results 230 000$ annual savings.
  - In Canada a data center is built directly underneath the premises of a local newspaper, which is heated with hot air.

Waste heat

  - In United State, the University of Notre Dame acquired a data center is located at shipping container, which waste heat can be utilized within a greenhouse. According to the university, waste heat makes it possible to save around 35 000 dollars annually.
- In Finland, companies Academica and Tieto have new data centers, which transfer their waste heat to the local city's district heating systems, which then is used to warm up a large variety of buildings and their service water. The data centers themselves utilize district cooling system to chill their IT equipment.

Aalto University
School of Engineering

Outline

Current situation

Green IT equipment

Green infrastructure

Green cloud computing

Tools to measure greenness

http://mpcuae.com/wp-content/uploads/2011/11/Green.jpg

Green cloud

- Energy-savings can be achieved by utilizing cloud services with large pooled resources, especially if, end users start to user thin clients to connect these services.
- However, the increasing amount of network traffic has a negative impact on energy consumption and therefore the greenness of cloud services is not unambiguous.
- Storage as a Service (Staas)
  - Based on test scenarios: When files were downloaded as rarely as 0.01 times per hour to a public cloud, the transport equipment required 25% of IT total consumption. Similarly, a private cloud used 10% of its electricity to transfer data between entities.

Green cloud

- When downloads occurred more often than once per hour, transporting data to a public cloud drained almost 90%. In those cases, where files are downloaded frequently, public clouds used 2.5 times more electricity than equivalent private.
- When cloud storage services were compared to local hard disks, the both public and private clouds used less electricity, but only under the circumstances, where the stored files were fetched less than ten times per hour and the typical size of these files was 1.25Mb. The gained energy-savings were merely one watt per user and therefore **cloud services are not significantly greener than local storage solutions**.

Green cloud

- Software as a Service (SaaS)
  - Applications are executed on a cloud platform and all the requ-ired data, such as word documents, are stored into the cloud.
  - Frames per second describes how often the full screen of a user terminal needs to be refreshed, which then indicates how much network traffic is produced and electricity drained.
  - The measurements reveal that the transport equipment of public clouds with 20-200 users required 5-10% of IT total electricity consumption, when the frame rate was only 0.01.
  - These transportation portions quadrupled to 20% and 40%, when the frame rate reached 0.1. Identical private cloud solution required around twice as less electricity for transporting data.

Green cloud

- When the servers within public and private clouds supported 200 simultaneous users and the frame rate was maximum 0.1, the economies of scale was reached, because the total electricity consumption per user was only 12-23 watts, whereas the laptop and modern computer required 17 and 70 watts.
- As a conclusion performed measurements points out that **public cloud services are energy efficient, but only if the frame rate is low and underlying servers are capable to handle great number of simultaneous users.**
- It seems that popular public cloud based office suits, including Google Docs and Microsoft office 365, could be considered greener than equivalent software running on a local workstation, because only a small portion of the screen must be refreshed, when documents are edited.

Green cloud

- Processing as a Service (PaaS)
  - It offers platform to perform intensive computation on a cloud.
  - The conducted tests to determine efficiency typically involved encoding a video file to another more desirable format. The size of video files were up to 8.5 GB and therefore several uploads and downloads would cause plenty of network traffic.
  - The results present that the servers of public and private clouds consumed less than 10% of total energy, when encoding were executed only 0.1 times per week. The portions of transport equipment in both cases were remarkably small.
  - However, one encoding per week increased the portion of public cloud network traffic to 15% and ten encodings per week similarly rise the portion to 40% of overall IT energy draw.

Green cloud

- Based on measurements, cloud services always required less energy to encode video than out of dated computers and it also seems that clouds could offer highly energy efficient computation for devices with very limited processor resources.
- Nevertheless, the number of encodings still affected on superiority between cloud services and local devices, because the public cloud started to drain more electricity than mid-range computers, when over 4 encodings were performed per week and the private cloud became less green after 8 encodings
- **The energy-efficiency of different cloud services depends on how well a cloud manages to perform a task and how it compensates the energy consumed by network devices.**

Green cloud

- Microsoft has wondered if its various cloud services are greener than its similar on-premise products.
  - Accenture and WSP were hired to research this topic
  - Three Microsoft business applications, including Exchange, SharePoint and Dynamics CRM, were selected to comparison.
- The cloud solutions offering services for
  - a large group of 10 000 users had 30% smaller carbon footprint than on-premise alternative due to smaller energy consumption.
  - a small group of 1000 users utilized the products of Microsoft, even larger gab exists, because cloud services release even 90% less carbon dioxide than on-premise applications.

Green cloud

- The key factors, which allow cloud services to have significantly low electricity consumption per user.
  - **1. Dynamic provisioning**, which strives to prevent over-provision. Providers have automated procedures to adjust available resources and experienced staff, which utilize sophisticated models and tools to plan future investments.
  - **2. Multi-tenancy**, which offers uniform physical infrastructure to multiple companies and their customers. Because of multi-tenancy, the demand peaks of individual companies can be combined to more predictable and steady demand rate, which helps to allocate resources. The gap between peak and average demand in cloud services is small in comparison and as a result cloud services require less extra resources during rush hours.

Green cloud

- **3. Highly utilized servers**, when compared to similar on-premise solutions, which typically utilize up to 10% of their resources. However, cloud services are not the only solutions with well used resources. The IT departments of companies can also virtualize their own servers to improve utilization and reduce need to purchase more hardware.
- **4. Cloud data center infrastructures** strive for extremely low PUE value of 1.1 or 1.2, which would offer up to 40% smaller overall energy consumption than conventional data centers.

- The future will show, if increasingly adopted cloud services are able to significantly decrease the carbon footprint of the whole ICT industry.

Aalto University
School of Engineering

Outline

Current situation

Green IT equipment

Green infrastructure

Green cloud computing

Tools to measure greenness

http://mpcuae.com/wp-content/uploads/2011/11/Green.jpg

Measuring efficiency

- Power usage effectiveness (PUE)
  - The most widely adapted energy efficiency metric for data centers. It has become the de facto standard in many countries.
  - Focuses to measure the energy efficiency of an infrastructure.
  - To produce a PUE value, the total electricity consumption of a data center is divided by the power draw of hosted IT equipment.
  - The score of conventional data center varies between 2.0-3.0
  - IT equipment workloads and outside air temperature vary over time and therefore PUE values are not static figures.
  - It has one notable flaw. It is possible that performed actions to reduce the power draw of IT equipment results inferior PUE score, because the portion of infrastructure increases

Measuring efficiency

- Energy Reuse Effectiveness (ERE)
  - Some people might claim that their PUE score is less than one, because they reuse wasted heat in other buildings. However, the minimum value for PUE is exactly 1.0 and the definition does not recognize any correction factors related to waste heat.
  - ERE should be used to present energy reuse instead. Its value is calculated quite similarly as PUE value, but in this case the total energy consumption is subtracted by the amount of reused energy and then divided by IT equipment energy usage.
  - Provides additional information about how well the drained energy is utilized. Providers with both low PUE values and ERE scores probably have rather green facilities.

Measuring efficiency

- Technology carbon efficiency (TCE)
  - Illustrated the relationship between energy consumption and carbon emissions. In practice, a score is formulated by multip-lying existing PUE value with electricity carbon emission rate, which depends on the type of energy source.
  - Rate is pounds of carbon dioxide per kilowatt-hour:(0.1-2.1)
  - Every power plant, however, have their own characteristic carbon emission rates, because some facilities have cleaner burning processes and more efficient filters than others.
  - TCE is highly relevant metric, because a data center with high PUE value might still be relatively green, if its electricity is produced by using hydro-electric power instead of coal power

Measuring efficiency

- Computer power efficiency (CPE)
  - It is a metric to measure the computation efficiency.
  - Defined by dividing the IT equipment power draw with the total power draw and then multiplying the outcome with the average utilization of processor units. It can be also calculated by dividing the average utilization with previously measured PUE.
  - If a data center with 2.0 PUE value and the average utilization is around 20%, the CPE value is 10%, which indicates that only a tenth of total power consumption used for computation.
- As a conclusion, PUE, ERE, TCE as well as CPE values needs to be measured so that the overall energy efficiency and the level of greenness can be determined.

Any questions?

http://meship.com/Blog/wp-content/uploads/2011/03/cloud-computing-green-environment.jpg