An Introduction into Cloud Computing

Prof. Dr. Gerhard Schneider

direktor@rz.uni-freiburg.de

Albert-Ludwigs-Universität Freiburg

UNI
FREIBURG

DAAD

Deutscher Akademischer Austausch Dienst

German Academic Exchange Service

Definition of Cloud Compting

- ## What is it?
  - I wish I knew…… ☹
  - A marketing hype?
  - A fundamental change in Computer Science?
  - A fundamental change in IT operations?

- ## To understand the concepts, we should check the historical evolution
  - History in CS is quite short ☺

Historic views

- ## High performance computing (HPC) has long been around

  - Use the fastest (and most expensive) computers in the world to advance science and industry.

  - Examples in Germany:
    Cray 2 for Stuttgart Univ. in 1990: 70 Mio DM
    Fujitsu Vector Computer for Karlsruhe Univ in 1992: 28 Mio DM
    (and it had a phantastic 2 GB of RAM)

  - Use it for crash simulation (-> car industry) or combustion simulation (-> better engines)

UNI
FREIBURG

HPC – an easy example
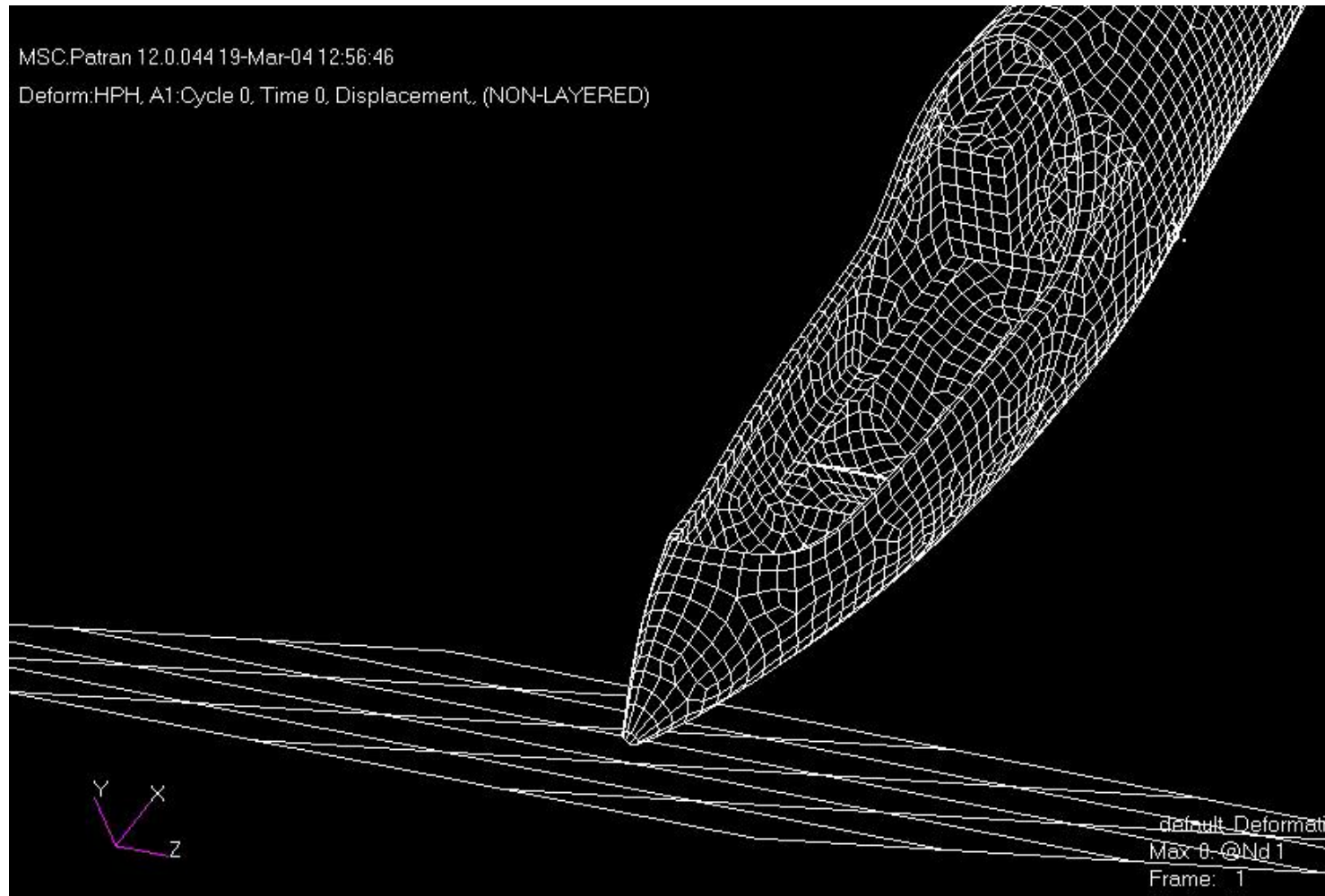
Source: somewhere on the web

HPC – consequences

- A whole new branch of science evolved: find algorithms which make the most out of your machines.
  - Consider latency, CPU weaknesses, etc
  - Every vector machine had its own approach, due to design issues
    - Once Cray, always Cray.
  - And the setup on every machine was different
- Just imagine:
  as a young scientist you train in Curitiba for machine A, come to Freiburg and get access to machine B
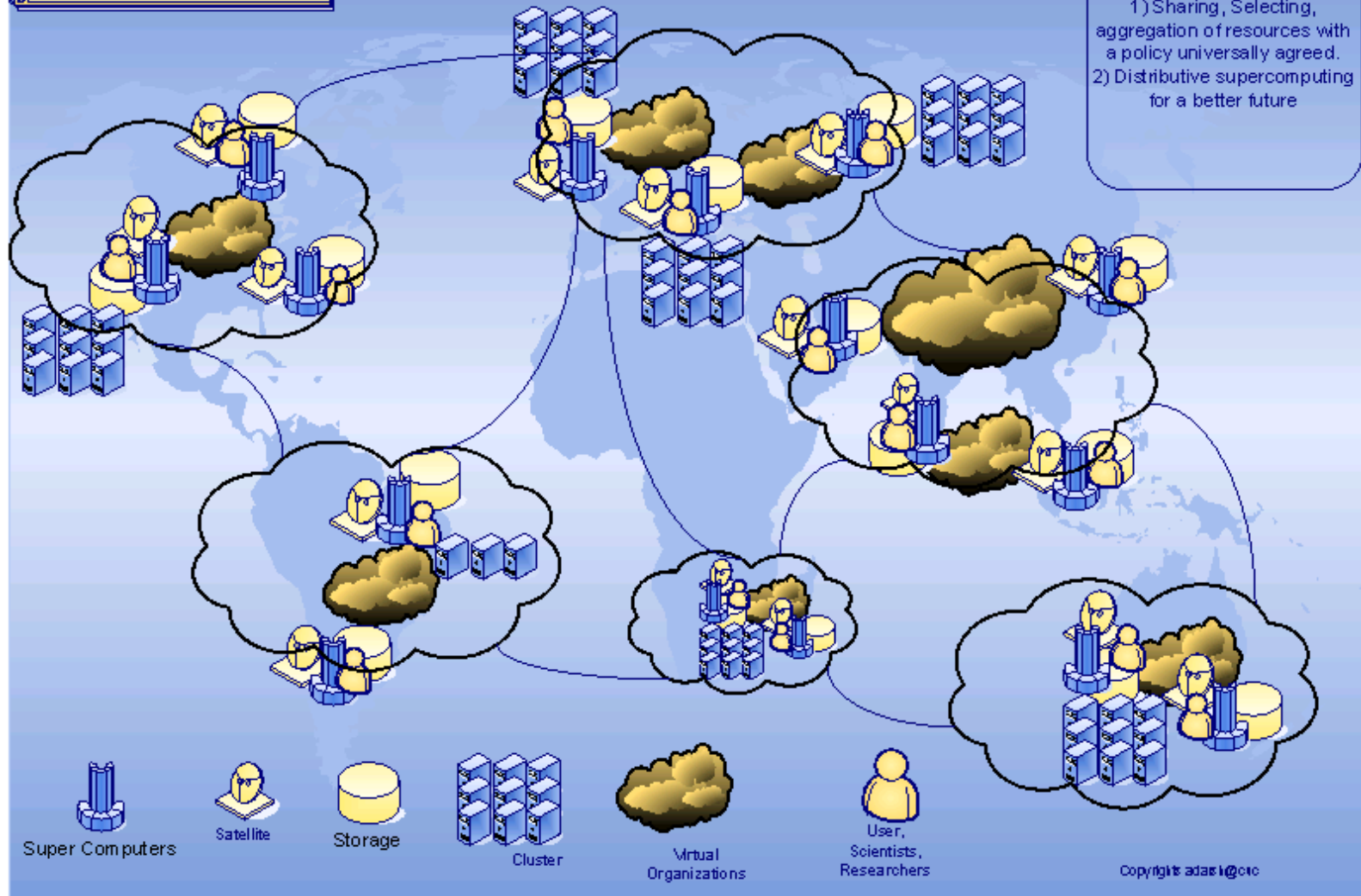
GRID

- How to get government money for new machines/projects, if people complain?
- Invent a new concept: GRID computing
- Getting CPU power should be as easy as getting electrical power – just plug into the compute grid. (Foster, Kesselman, etc 2004)
  - Fundamental new concept: the existence of the Internet ☺
  - rather than *telnet* use *remote execution*
- to succeed with new concepts you have to be close to the (US) government

UNI FREIBURG

GRID

- A new boom: try to hide the many different HPC machines behind portals
  - And implement similar mechanism on each of them
  - Quite a cultural shock – people had been trained to get the best out of each machine, which contradicts the notion of a standard
  - Job submission, file space etc.
  - Basically: submit your job, it is compiled on some target machine (with the best parameters) and executed there – the results find their way from the machine's storage onto your storage.
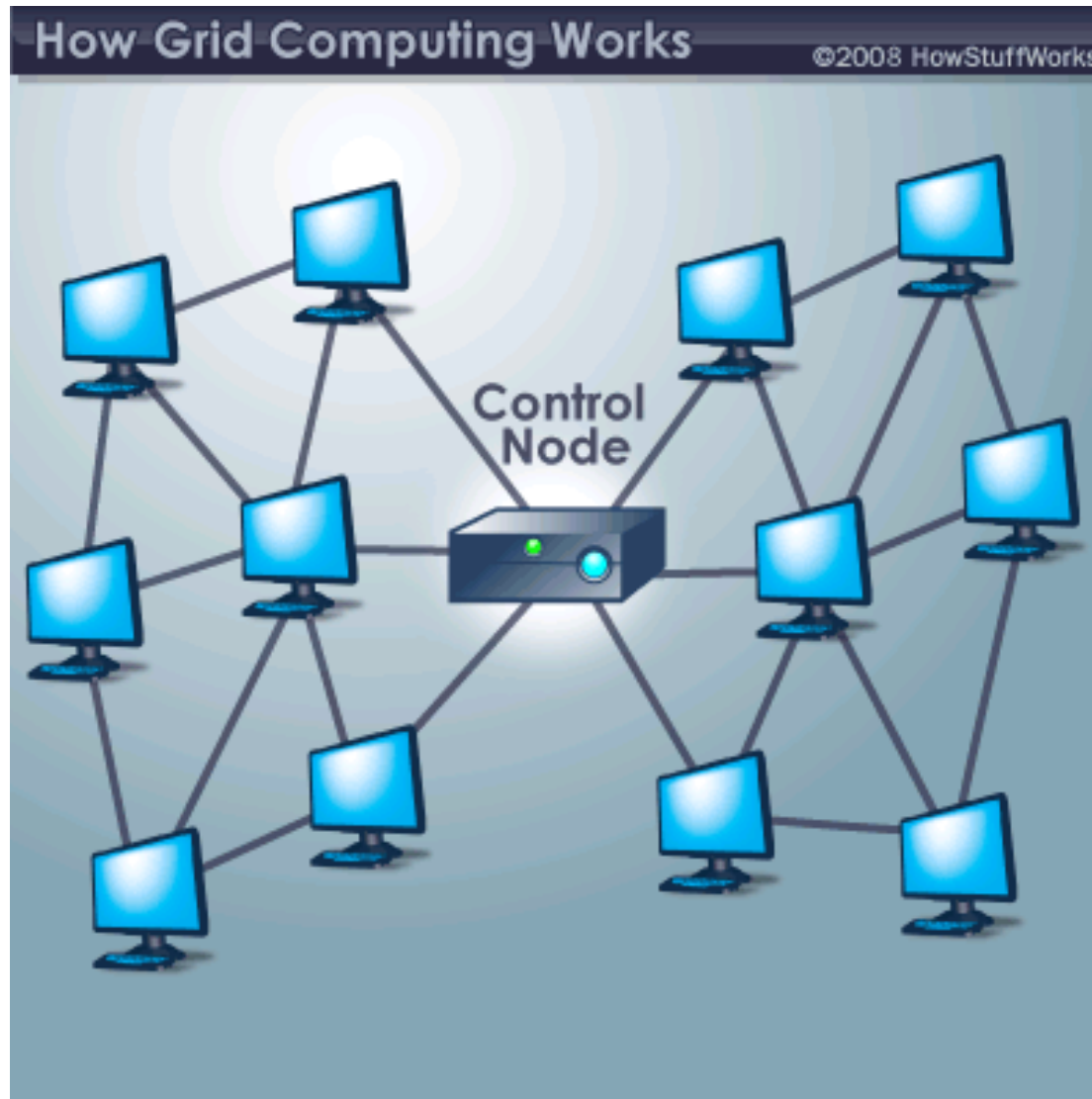
UNI
FREIBURG

GRID – from an academic perspective

Source: somewhere on the web

GRID – a wrong marketing slide

Source:
somewhere
on the web

UNI
FREIBURG

GRID – a definition

Grid is a type of parallel and distributed system that enables the sharing, selection, and aggregation of geographically distributed "autonomous" resources dynamically at runtime depending on their availability, capability, performance, cost, and users' quality-of-service requirements. (Dr. Buyya)

UNI
FREIBURG

GRID framework (1)

- Globus
  - Computing resources, file storage management, security provisioning, data movement, monitoring,
  - toolkit for additional services based on the same infrastructure,
    - agreement negotiation, notification mechanisms, trigger services, and information aggregation.

UNI
FREIBURG

GRID framework (2)

- Unicore:
  - Uniform Interface to Computing Resources
  - www.unicore.eu
  - offers a ready-to-run Grid system including client and server software and makes distributed computing and data resources available in a seamless and secure way in intranets and the internet. (good advertising…)

UNI
FREIBURG

GRID problems

- ## Why should I let you onto my computer?
  - ### What do I get in return?
  - ### How much is 1 CPU worth on your machine in comparison to my machine?
  - ### Funding system for big computers in science contradicts the „selling" of CPU time
- ## Even if you want to succeed:
  - ### How to give rights (which ones) to someone who has rights on your machine?

UNI
FREIBURG

GRID

- Did you ever compile a programme?
  - With *make*?
  - Did that work? When you moved your programme from Linux to Windows?
    - Or from one Linux to another?

- Life is easier if you submit jobs to existing programmes / software systems which are already installed in a perfect way
  - E.g. a computation in Mathematica or Nastran or…
  - According to your specification the portal chooses the „best" computer

UNI
FREIBURG

GRID examples

- Famous case: SETI@Home
  - Every interested user installs a (little) programme which gets its instructions from a portal (which was/is fed by Berkeley Uni)
- Also an easy case:
  - People agree to „donate" CPU time
    - And spend money for electrical power, etc
- Evolution: install several programmes of „type SETI" on your machine and let a broker decide on which will be executed
  - Obvious question: what does such a programme do on my computer?
  - Sandboxing? Is it a virus?

UNI
FREIBURG

Hmm ?!?

- So there a quite a few challenges associated with this approach
  - It does not scale as soon as it leaves the scientific „inner circle"
- It would be much easier if the computers which offer the services would be owned by one owner (rather than many)
  - Which excludes the scientific „big iron" computers
- In addition modern computers are no longer specialized machines, but loads of Intel CPUs (or AMD)
  - Easier to buy for companies

UNI
FREIBURG

Typical „pile" of computers

Part of bw-GRID

The Cloud

- Put this all together:
  - Someone buys or rents resources from others
    - Using legal contracts – thus complete legal control over the resources (ownership,…)
  - And builds up a service
    - Like „amazon file service"
  - And offers this to customers
- The technical background is hidden from the user and appears to be foggy – thus „cloud"
- As Scott McNealy, CEO of former SUN Microsystems, once pointed out: you can run an airline without owning planes – just sell the surplus capacity on the market

UNI
FREIBURG

The Cloud – an official definition

"Cloud computing is a model
for enabling *convenient*, *on-demand*
network access
to a shared pool of configurable computing
resources
(e.g. networks, servers, storage,
applications, and services)
that can be rapidly provisioned and released
with minimal management effort or service
provider interaction."

National Institute of Standards and Technology ([NIST](#))

The Cloud

- Nice example: **Dropbox**
  provides filespace accessible in a clever
  way from all sorts of devices
  - Linux, Windows, Mac, iPhone, iPad, Android
  - Use your data offline and resync later

- Dropbox buys its filespace from Amazon
  and remarkets it
  - A bit like the airline example.
  - Amazon delivers a „reliable" filespace to Dropbox and
    Dropbox sells it to the customer
  - From the customer's view the location of the data remains
   foggy – in the cloud

UNI
FREIBURG

The Cloud

- Caveat: this only works if someone is stupid enough to buy surplus capacity
  - Stupid in economic thinking.

- The requirements of your own business may force you to have powerful equipment
  - Amazon certainly has this need around Xmas

  and adding „a bit more" is cheap
  - 3TB disk costs 50% more than a 1TB disk

- If the „cloud provider" cannot find the resources, it may have to buy its own
 - Contrary to the idea of a cloud ☺

The Cloud

- *Nasty descriptions:*
- You buy surplus capacity cheaply to offer a service which cannot be offered by the owner of the surplus capacity
  - Because the owner is too stupid (???)
  - Because the owner is too big to play in this field
  - Because the owner thinks it is too risky (???)
- You buy extra capacity at a high cost from a provider who has waited for the emergency and you sell it to an even more desperate customer.
  - Insurance principle
  - Renting fire fighter planes from ordinary airlines, refurbish and rent to a country in need.

UNI FREIBURG

The Cloud – strange views in Wikipedia

UNI
FREIBURG
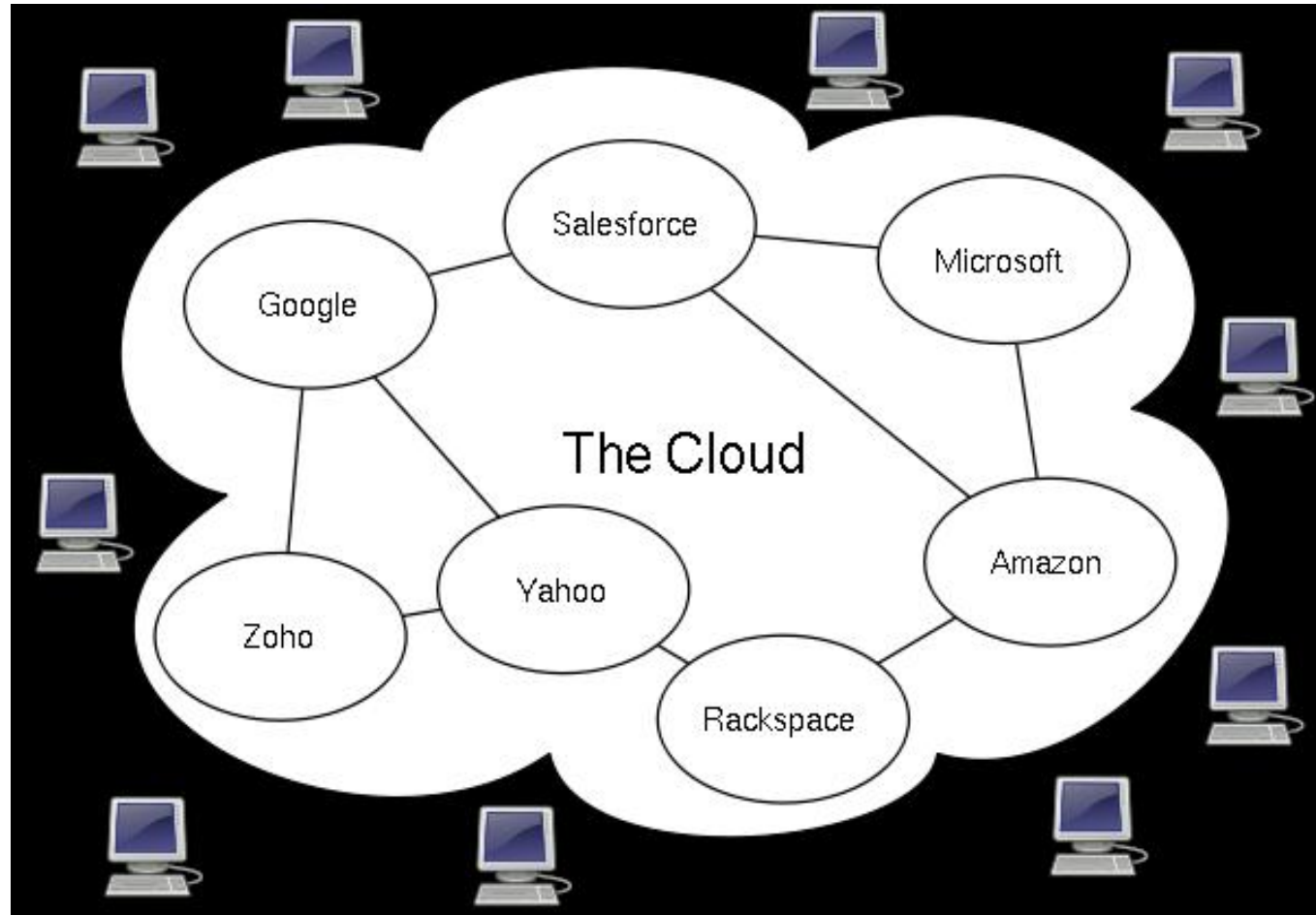
The Cloud – a slightly better functional model

UNI
FREIBURG
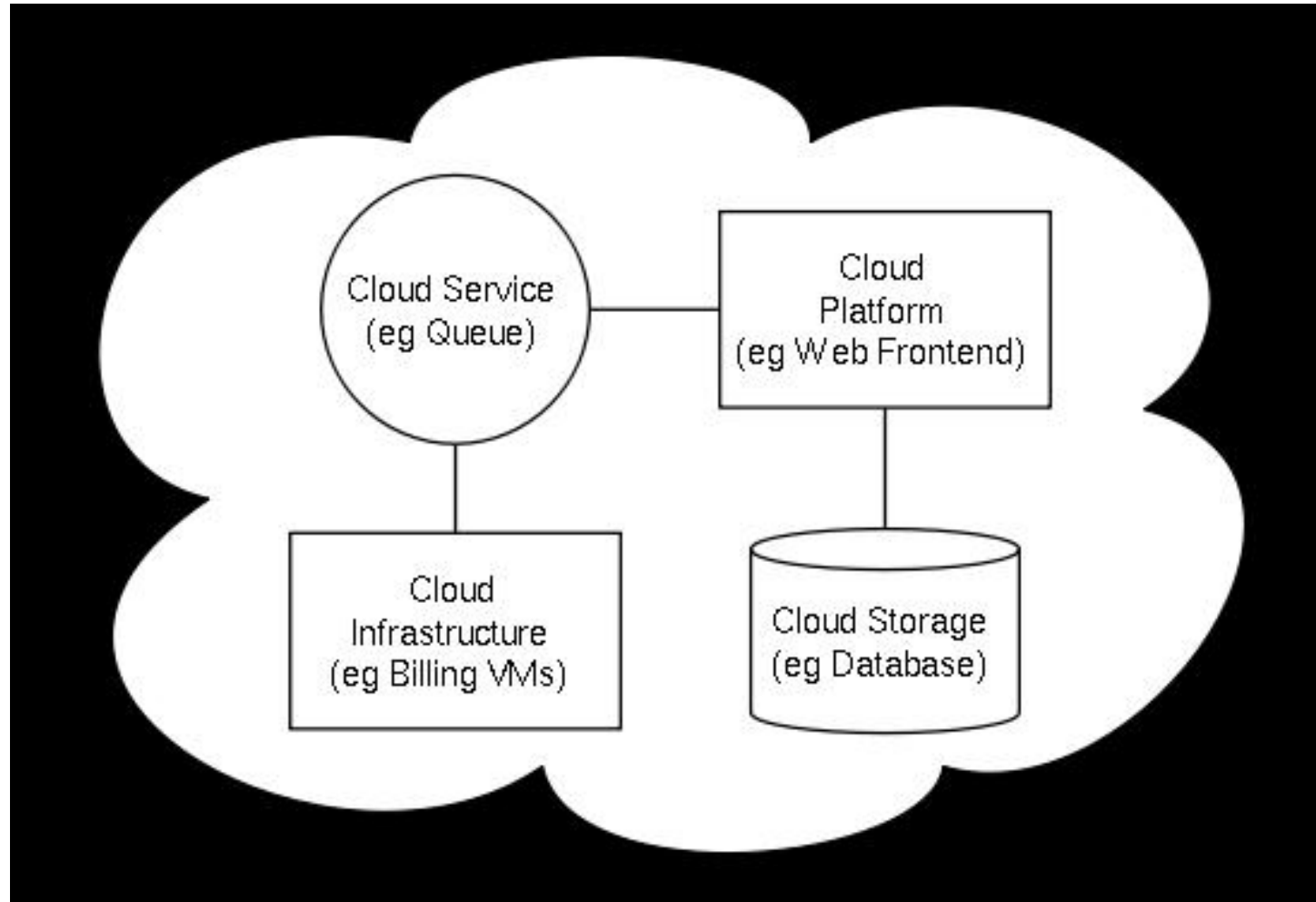
Now we know ☺

- The notion of a cloud vanishes, if the provider owns all relevant pieces
  - Google, Microsoft, Amazon,
  - Internally it may be a operated like a cloud (??)
- But we can use the concept ourselves
  - Rather than buying disk space for a short term research project – rent it from Amazon and incorporate it into your environment
  - Such that the research project never finds out
  - Organize the computer centre resources such that they can be offered in a flexible way

UNI
FREIBURG

Use of the cloud concept

- Example *file service*:
  - Fileserver offers CIFS or NFS, user must take what is offered
    - User not allowed to come up with smart ideas
  - Or export storage capacity in a flexible way:
    - Export raw disks, iSCSI, Fibrechannel over Ethernet, etc. – even CIFS, if required
- Example *Voice over IP*
  - Only do the call control in the backend
  - Ignore how smart phones are today.
    - Leave the job to the smartphone programmers
  - See in comparison what the telco monopolies in Europe did to scientific advances / liberalisation.

UNI FREIBURG

Back to square one

- Virtualisation allows to view a virtual computer as just another application
  - Just like mathematica, Nastran, etc.

- The „cloud provider" may offer a vmware service:
  - Virtual machines are accepted as input and executed somewhere
  - Amazon will offer this next year
  - The Freiburg environment allows this for years

- So we can now build our own supercomputer „on demand" by starting enough virtual machines „in the cloud"

UNI
FREIBURG

Back to square one ?

- Is there enough CPU power available?
  - Why should a cloud provider buy an excessive number of computers, just in case?
    - We are talking real money
  - No point in running 10 VMs on <u>one</u> real hardware
  - No point in running 10 VMs on 10 different real computers with differing performance
    - At least not, if you run parallel jobs
    - Service agreements are vital

UNI
FREIBURG

Different viewpoint

- Now let us take the customer's viewpoint
- And let us consider „uninteresting" side effects
  - „uninteresting" from a pure technological point of view
  - However they might be **vital** for the individual user

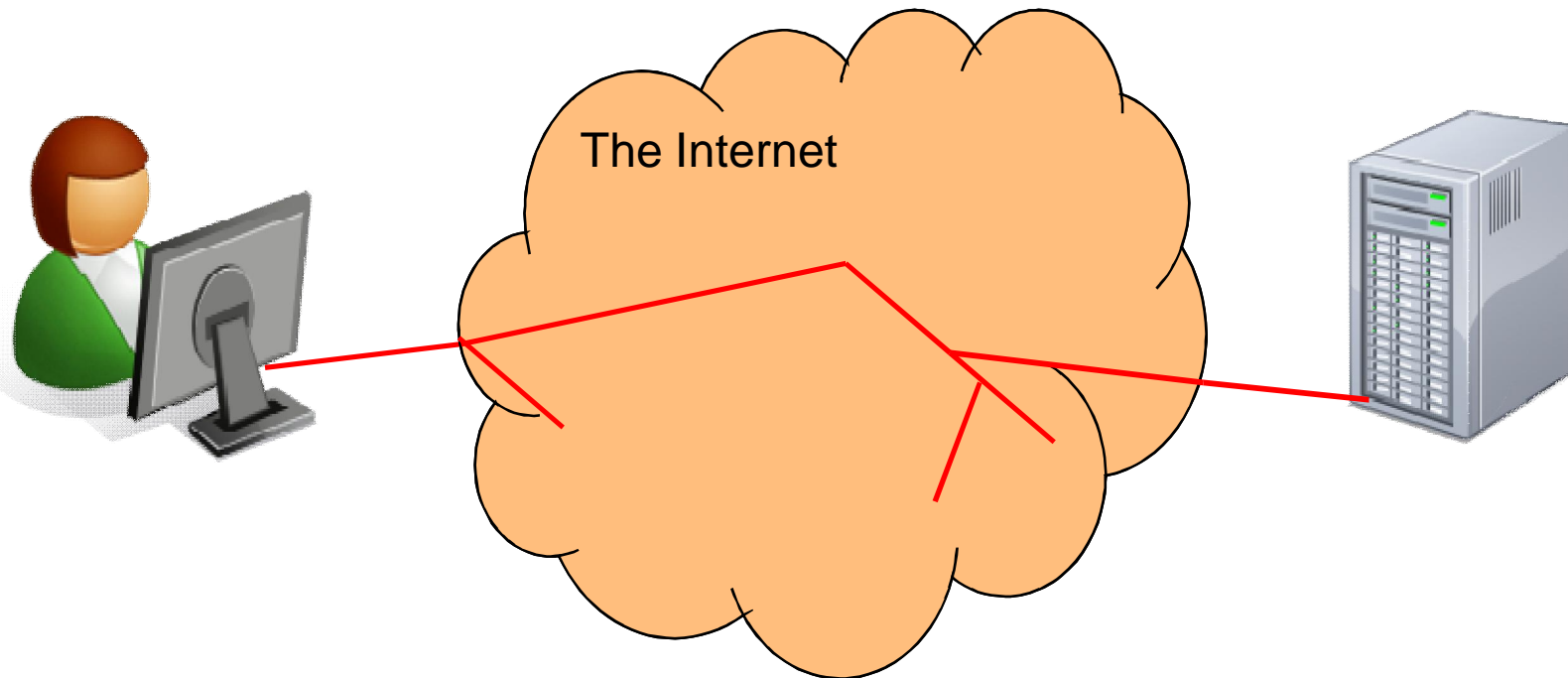- First, let us summarize what we have discussed so far

UNI
FREIBURG

Classical access

Direct line terminal – computer

Resonably well understood – if we trust the line
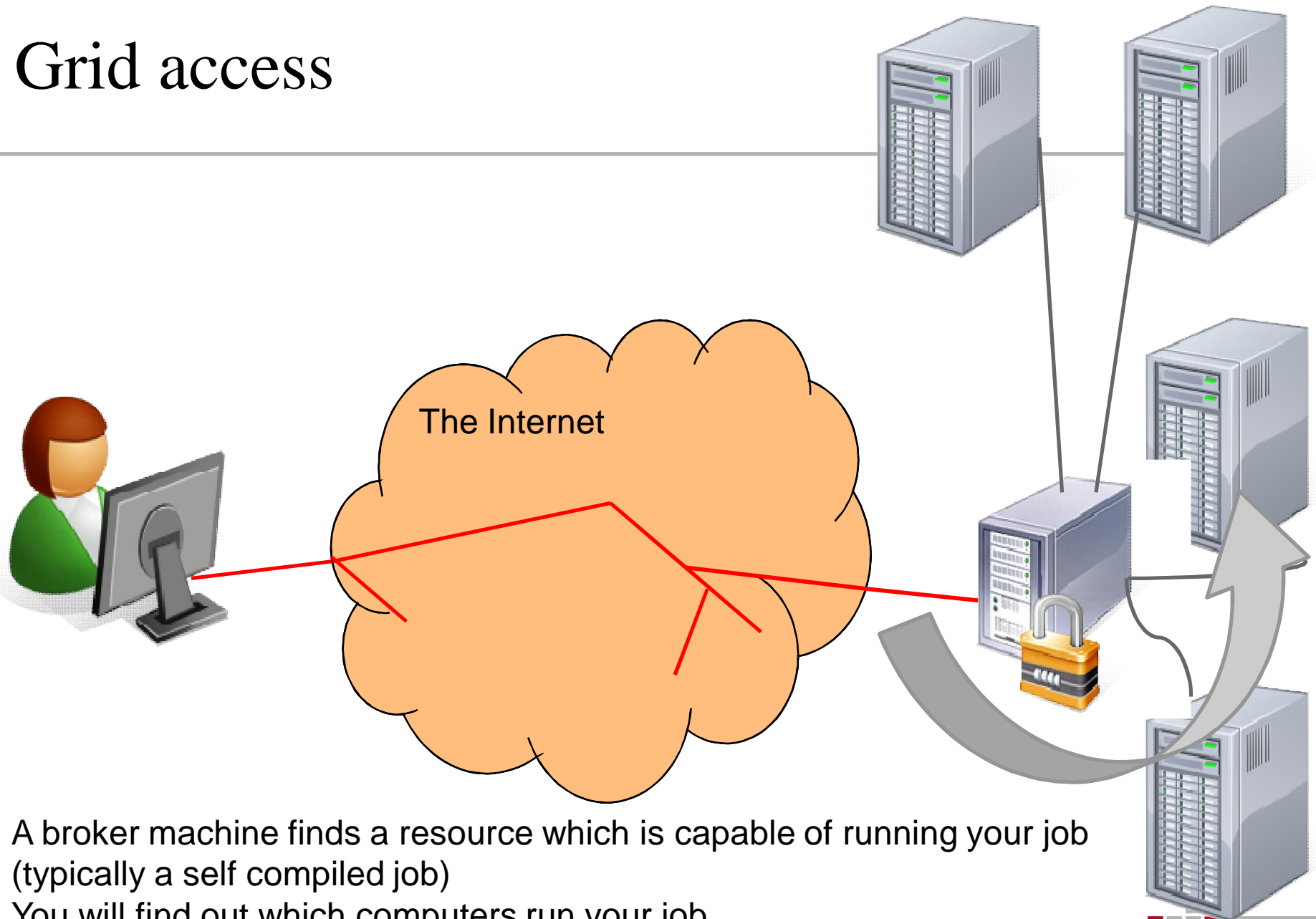
UNI
FREIBURG

Internet access

The Internet

Internet – the **network cloud**
somehow the data reaches the target and finds its way back

Plenty of legal issues / security issues / technical issues

UNI
FREIBURG

Grid access

The Internet

A broker machine finds a resource which is capable of running your job
(typically a self compiled job)
You will find out which computers run your job
Modelled on the Power-GRID: get **computer power** as easily as electrical power

Cloud access

The Internet
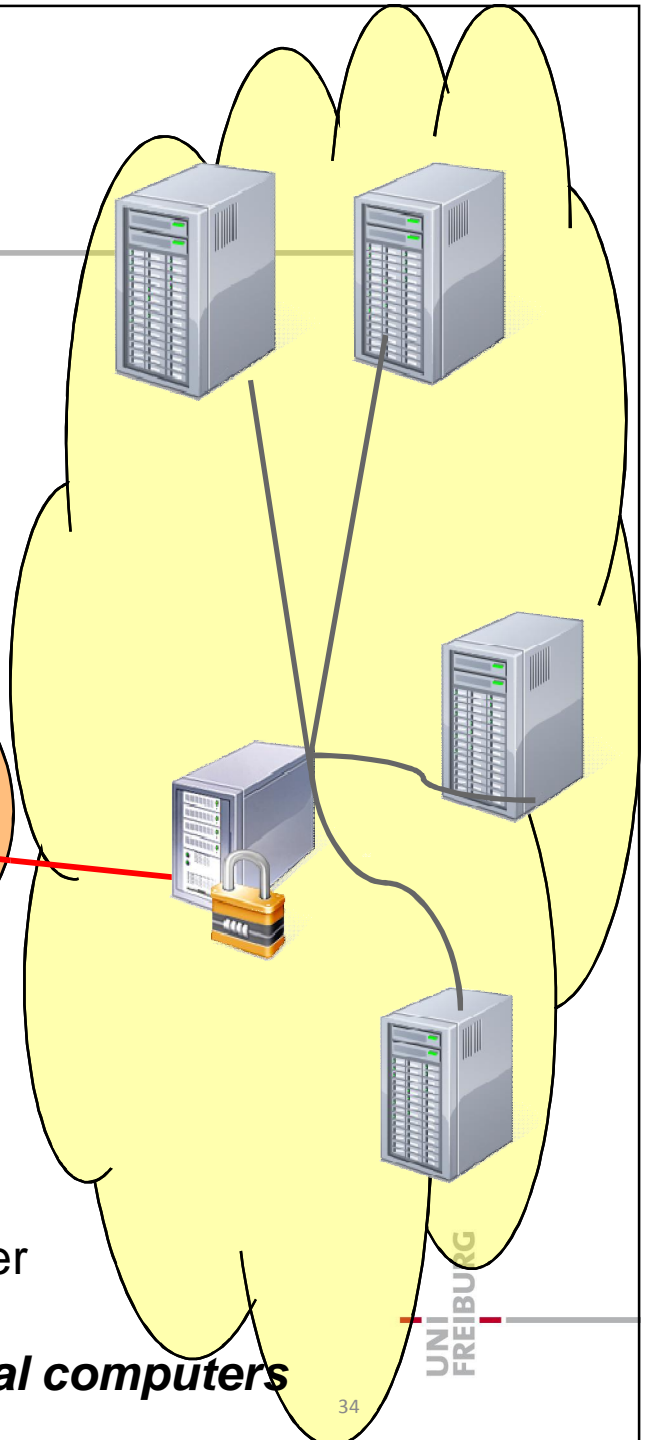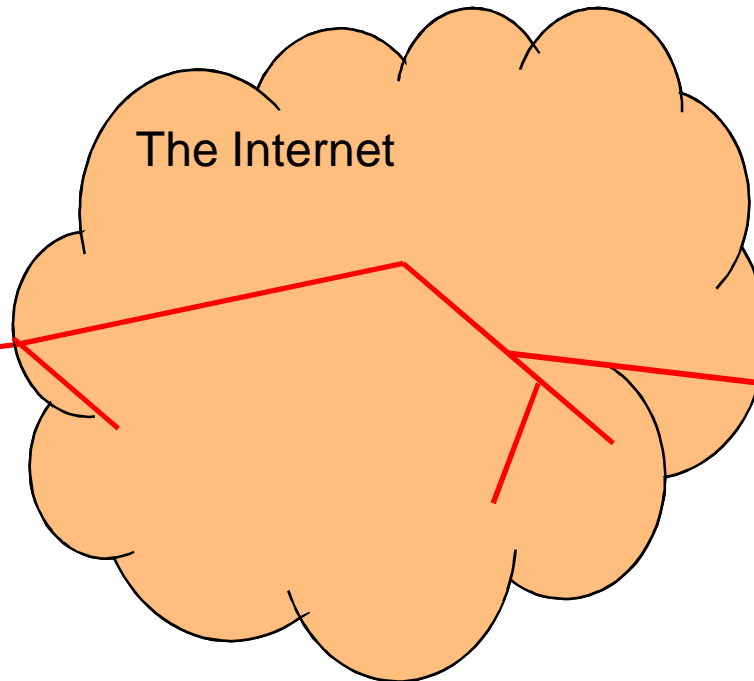
A service gateway offers you a **service**
The actual computer **task** is executed on some computer
somewhere and the result is returned via the gateway
-- *email servers, Excel server, storage servers, virtual computers*

UNI
FREIBURG

Examples

Amazon:

- 1 GB costs 0,15 € per month (for the first 50TB)
    - Your personal 10GB (photos, etc) thus cost 18€ p.a. or 90€ per 5 years – the life expectancy of your 500GB hard drive
    - And this includes a complete service (backup, redundany, invisible recovery, energy, permanent access from all over the world). But transfer (in/out) costs extra
    - Other reason: you need 50TB for a project for 5 months – NOW. You pay 37500€ vs 27000€ plus staff

Google:

- Mail account with 2GB storage – for free
    - World wide access, full service, advanced features

Reason: large systems scale better

Dangers / challenges

- What if resources are running low?
  - No sudden storage peaks in the world to be expected!
    - What if there is a Tsunami ?
  - Access bandwith can be challenging
    - Try to pump your 50TB through the Internet after a damage (as on Sept. 11, 2001)

- Large resource providers are international companies
  - Whose law applies when and where?

Challenges

Explicit questions asked to a Google VP

- How secure is my email? *Totally secure!*

- If the government wants to see it? *Of course we adhere to legal requests*

- If the <u>US</u> government wants to see my email? **Silence** (plus a few white papers via email, but no answer)

Explicit question asked to Scott McNealy (SUN)

- How do I stop the CIA from reading my email if I use a big mail provider who is operating under US law?

  He was completely thrown off course and required 3 mins to be back on track.

Challenges

- A service provider has access to my data and can physically move it to any server in the world
  - Without me noticing it
  - Move it onto a server in a country with appropriate legislation
    - Move it there 5 mins before the search warrant is executed.

- I am not paranoid, but
  - becoming a suspected terrorist is easy today
  - „public execution" of important people because the provider can look at the data and tip someone off.
  - Industrial espionage
    - Well known French successes in the past

- Is a „provider in the cloud" a reliable partner?

UNI
FREIBURG

Potential answers

- ## Legal framework?
  - International companies / different national laws

- ## Business model?
  - If Google uses your data they will be out of business
    - If we find out – move data onto a google server in Langley

- ## Self defense:
  - Encrypt your data
  - Should a government sponsor develop tools to protect its citizens and companies from friendly states?

Proper security – not TSA security

Interesting solution

- TeamDrive www.teamdrive.net

- Stores only encrypted files in the cloud

- Encryption works

  - without cloud storage provider

  - without TeamDrive Corp

  - No access to content even for the police!

- Automatic synchronisation over several computers via cloud

  - Shared workspace offline/online

- One step in the right direction

No advertising! No affiliation!

Challenges revisited

- Who is monitoring the network?

- Where is your data?

- Who „fixes" the real computers?

- Where is your job executed?

- Who monitors your jobs?

- Who infiltrates your virtual machine?

- Who keeps backups of your data? Even after you deleted your data? For how long?

- This is **no longer a local issue**, but the actions take place somewhere in the world, at a new place every day

- Is „don't use it" an answer?

Challenges for the provider

- The user
  - simulates a nuclear bomb
    - Legal on the Iranian part of the cloud, but….
  - uses the many machines for SPAM, DoS of hitherto unknown dimensions
  - uses the storage for pornographic material
- Is the provider liable? Under public pressure?
- (self-)protecting the provider:
  - keep all (connection) data forever?
    - The press will hunt you down, if a user misbehaves
  - Permanently check the data?
    - Unlike „in the Internet" you have a chance to understand!

Clouds are for individuals

Ubuntu 9.04 now has a cloud component –
you can put your computer in a cloud, so
that it offers services while it is idling
during class

- Clouds can be international – and you execute
  foreign jobs
- Storage cloud: 2GB in the basic version (10GB
  for $10 a month) to synchronize data between
  several computers

Consequences for a data center

- We in Freiburg provide a mini-Cloud!
  - Within a state legal framework
    - Seams easy for you – at first
  - But we (may) collect too much data about each user – because we are forced to do so
    - By law
    - By company rules – to prevent fraud
    - By professional guide of conduct
    - By hardware manufacturers (error logs)
- The combination of all this data is interesting
  - The more you have the more interesting it is.
- Giving the data to a trusted government site just moves the problem - and makes it bigger

UNI
FREIBURG

Combination of data

- Wide spread misconception:

  - (especially in the legal profession)

- we can hide in the enourmous sea of data

- Recent aquisition in the computer center

  - HP 8 CPUs (4 cores each) and 512GB of RAM

    - Costs a mere 40000€

  - enough to store 80 bytes for
    every human on earth
    in fast main memory!

- Cross-searches are very easy today

Consequences for IT usage in our life

Need for an „accepted new ethics for IT management"

- Which then leads to appropriate legislation
    - Accepted by the public
- Guaranteed anonymity
    - Even if abuse cannot be identified
        - The unsolved murder cases suggest that society is willing to accept this in the real world
- Do not store data to keep traces
    - Nobody tracks my car … (I believe)
- Enforce immediate/delayed deletion of all traces of deleted data
- Do not use data, which happens to be there, for other purposes
    - Toll collect
    - Unintended use may shatter confidence in the system – and change the behaviour of people

UNI FREIBURG

consequences

Advanced service mentality

- Encrypted storage / backup
  - TSM has this feature, advertised by RZ

Secure storage of logfiles

- BSI-law: admins need „all available" data to track viruses, if something gets in?
  Why is this accepted?
  Store logfiles encrypted – ask for permission to open only the relevant data

Data „protection" must overrule „marketing"

- Send to all email addresses…

The world moves into the other direction

- The latest BGH-Urteil (Az. 1 StR 76/09) w.r.t. email appears to be less helpful
  - Can email be confiscated easily?
    - While it arrives at the mail server?
    - If it lies on the mail server unread?
    - If I leave it on the mail server after reading it?
      - „in the cloud"
    - If I delete it after reading but there is a backup copy?
      - How reliable is „delete"?

Classical regulatory approach won't work

- The IT-professionals can't control the abundance of „useful" data
- The legal system neither, but it requires 3 jobs: prosecutor, lawyer, judge

Interesting research

- How can we improve anonymity?
  - Where does this help?
  - GSM – not easy
- During software design think about potentially„interesting" use cases
  - German road toll system
  - Part of computer science education?
    - Raise awareness!

UNI
FREIBURG

The more you study, the more you learn.

The more you learn the more you know.

The more you know the more you forget.

The more you forget the less you know.

So why study?

Seen in the Bodleian Library, Oxford