INF5870 - Energy Informatics

Tanusan Rajmohan - tanusanr@ulrik.uio.no

UNIVERSITY OF OSLO

Spring 2018

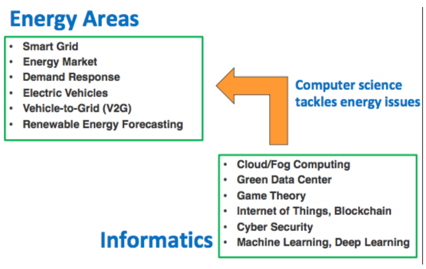
Contents

1 Goal

Exploit ICT (Information & Communications Technologies) to tackle the global warming and climate challenges.

Energy informatics is not about the traditional concepts, but it focuses on state-of-the-art computer science for sustainable future energy systems.

Course topics:

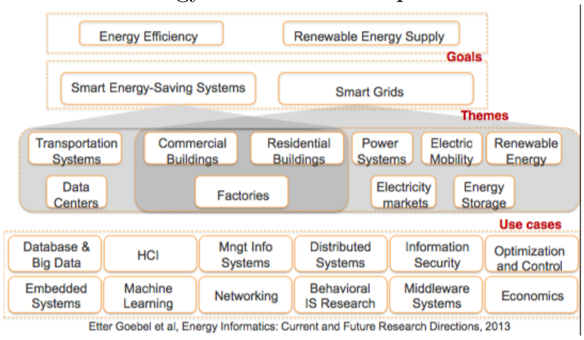


2 Introduction

Definition: **A new field covering the use of ICT** (Information & Communications Technologies) **to address energy challnges**

Energy areas:
- Smart grid; smart energy networks
- Smart building
- Smart cities
- Water systems
- Oil / Gas systems
- Transport systems
- Electric vehicles
- Vehicles-to-grid (V2G) systems
- PV systems
- Wind systems

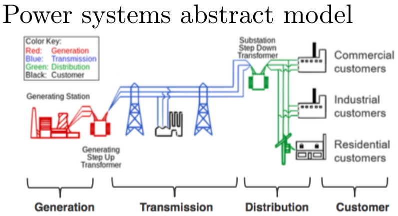
Energy informatics: scope

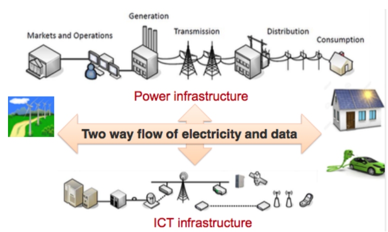


2.1 Smart Grid

Smart Grid = Power Grid + ICT

Power systems abstract model





Smart Grid concept and vision:

Cost-effective: cost-effective production and delivery of power

Green: greater use of renewable resources; support for a large number of electric vehicles

Customer-oriented: consumers can choose energy usage

Secure: resilient to various cyber attack

Dynamic: dynamic pricing and load control

Reliable: higher reliability of services

Communication and control infrastructure network to augment power grid operations.

Smart Grid needs Computation, Networking, Communications, and Control.

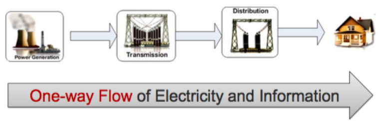
2.2 Smart buildings

Smart building has its own "Operation Systems" to monitor and management the infrastructure.

A typical computer engineering problem: how to collect data from sensors and then adjust the temperature in the building with both energy-efficiency and people's comfort considerations.

3 Overview - Smart Grid

Traditional Power Grid:
Power is generated in a plant
Power is distributed through the long-distance high-voltage transmission networks to the local community.
Power is distributed to the customers.



3.1 Problems with Current Power Grid:

Not efficient

- Transmission loss = 20 %

Not reliable

- Failure can quickly spread

Not secure

- Cyber attack

Not green

- Electricity accounts for 41 % of energy related CO_2 emissions.

3.2 Grid Definition:

Smart grid is the future grid that integrates a reliable and efficient communication and networking infrastructure that adds intelligence to the power grid to enable:

- grid automation

- user-responsiveness

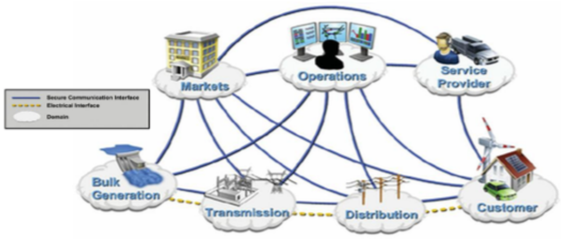
- efficient demand response management

3.3 NIST Conceptual Model for Smart Grid

NIST (National Institute of Standard and Technology, USA)

Interaction in 7 Smart Grid domains through communication flows and electrical flows.

Each domain involves its own actors and applications.



3.3.1 Customers:

- The end users of electricity
- Customers may also generate, store, and manage the use of electricity
- Normally, we have three types of customers, each with its own domain: residential, commercial, and industrial

3.3.2 Market:

- The operators and participants in electricity markets
- Participants in wholesale market: day ahead, hour ahead
- A market may involve predication, bidding, auctions

3.3.3 Service Providers:

- Organizations providing service to both utilities and electrical consumers
- Internet service providers, charging stations operators, ...

3.3.4 Operation:

- The manager of the movement of electricity
- Independent Systems Operators (ISOs) or Regional Transmission Organization (RTOs). An ISO or RTO serves as a third-party independent operator of the transmission system

3.3.5 Bulk Generation:

- Major power plants. The generators of electricity in bulk quantities. May also store energy for later distribution.

3.3.6 Transmission:

- Carriers of bulk electricity over long distance
- A system operator is responsible for an area to be electrically stable, and for the security of supply in its area.
- In the Nordic countries, the system operators have the responsibility for both the security of supply and the high-voltage grid (the transmission grid). Statnett is Norway's transmission system operator.

3.3.7 Distribution:

- Distribution of electricity, to and from customers
- May also store and generate electricity

3.4 ICT for Smart Grid

Smart Grid needs communications and networking technologies

- Smart metering and Advanced Metering Infrastructure (AMI)
- Distributed generation and renewable energy integration
- Power outage detection
- Real-time monitoring, diagnostics and protection
- Big data generated by massive numbers of sensors, meters, and telemetry
- Further analysis control, real time pricing
- Communications network support intelligent energy scheduling

Communications requirements are different from different applications

- **Type I**: Applications that basically no need for communication
  - It is good to have communication to send state information to the system operator. Small data and latency is not critical
- **Type II**: Applications that has latency in the range of ten seconds and the cycle of data scan is in the range of minutes
  - Solutions require communicating the status of controllable devices and voltage, current and power flow measurements at selected points
- **Type III**: Applications that has the highest requirements for communication infrastructure.
  - It depends on the system model and load forecast and can be improved by having more measurements (voltage, current, ...)

3.5 Smart Meter:

Smart meters perform functions such as:

- Energy consumption
- Communications with other intelligent devices in the home and utility
- Time-based pricing
- Loss of power notification
- Remote turn / turn off operations
- Power quality monitoring

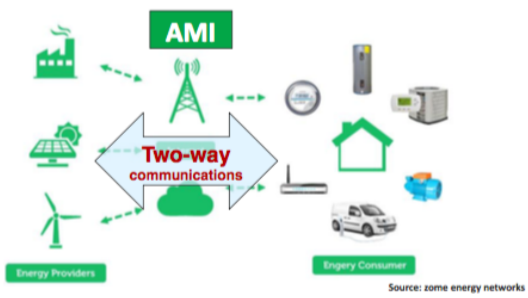
Pivacy issues:

Questions	Pattern	Purpose
Were you home after you claim sick to your boss?	Yes: Power activities during the day No: Low power usage during the day	?
Did you get a good night's sleep?	Yes: No power events overnight for at least 6 hours No: Random power events overnight	Health services
Are you on vacation?	Yes: No power events for some days No: random power events in a day	Useful information for someone malicious
Did you leave your child home alone?	Yes: Single person activity pattern No: Simultaneous power events in distinct areas of the house	Safety services

3.6 Advanced Metering Infrastructure (AMI)

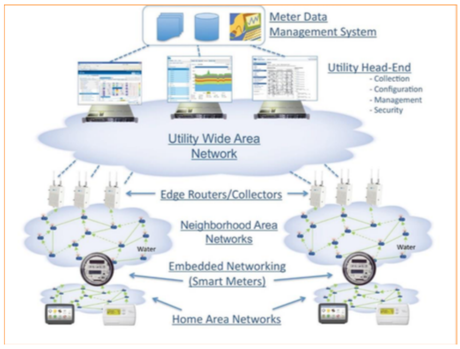
Smart meters do not work separately. They are organized as networks.

AMI is an integrated system of smart meters, communications networks, and data management systems that enables two-way communication between energy providers and energy customers.

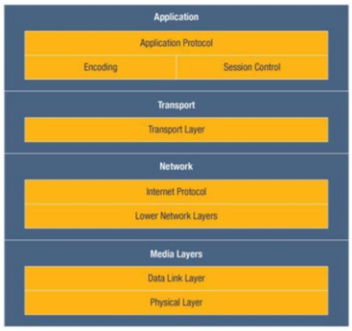


AMI integrates a number of technologies:

- Smart meters at the consumer end
- Communication networks at different levels of the infrastructure hierarchy to connect two ends
- Meter Data Management System (MDMS)
- Platform integrating the collected data into application at utility provider



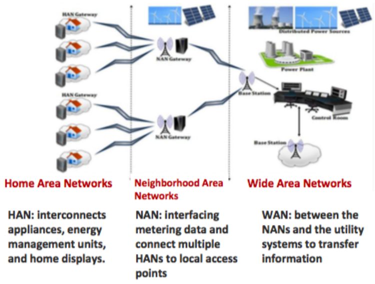
Internet Protocols for Smart Grid
NIST Priority Action Plan suggested the suitability of internet networking technologies for Smart Grid applications.
This work area investigated the capabilities of protocols and technologies in the Internet Protocols Suite to determine the characteristics of each protocol for Smart Grid application areas and types.

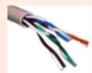


Internet Protocols for the Smart Grid per PAP01 (June 2011).

3.7 Communication Networks for AMI and Smart Grid

Power systems abstract model



Technology	Data Rates	Coverage	Medium	Band Licensed
PLC	1-200Mbps	1500m	Power cable	Free
Ethernet	100Mpbs	100m	UTP (unshielded twisted pair) cable 	Free
WiFi	5-100Mpbs	30-100m	Wireless	Free
Zigbee	0.02-0.2Mbps	10-75m	Wireless	free

3.7.1 Home Area Networks (HAN)

HAN: gathers sensor information from a variety of devices within the home, send control information to these devices to better control energy consumption, and provides access to in-home appliances. HAN can use Zigbee, WIFI, Power Line Communications.

3.7.2 Power Line Communications (PLC)

PLC is a communication technology that enables sending data over existing power cables. PLC carries data on a conducter that is also used simultaneously for electric power transmission or electric power distribution to consumers. PLC is a wired communications technology, it can compete with wireless technologies with low cost since the infrastructure already exists.

Example of PLC at home:

PLC uses the household power grid to transfer data between computers equipped with suitable adapters and other components. The data is modulated prior to transfer and sent as a signal via household power lines.

How does PLC work?

- Three phases:
  - Digital signal to be transmitted is modulated over power line. This is done by a proper coupler device.
  - The signal is propagated over the power line.
  - It will be decoupled and decoded at the receiver side.
- Major technical challenge:
  - harsh medium for data communication
  - to reduce sensitivity to the electrical noise present on power lines: signal attenuation, signal distortion and noise.
  - unpredictable and varying characteristics: time, frequency, and location.

3.8 IEEE 802.15.4 - Zigbee

Zigbee is a short-range, low-data rate, energy-efficient wireless protocol, supporting star, tree, mesh topology and suitable for WSN. It utilizes 16 channels in the 2.4GHz ISM band worldwide. It supports data rate of 250kbps, 100kbps, 40kbps and 20kb.

Zigbee for Smart Energy at Home

- Zigbee defined two smart grid application profiles.
- Home automation: lightning, window shades, monitoring, security

- Smart energy: Zigbee end devices connected can control power supply switch of home appliance; Zigbee enabled smart meters efficiently manage demand response; actively respond to different prices and effectively balance the power consumption load in the power grid.

- Challenges: limited battery energy supply, memory and processing power.

3.9 WSNs for Smart Grid:

Power generation side

WSNs provide pervasive communications and control capabilities at low cost. Both utilities and customers can transfer, monitor, predict, and manage energy usage effectively.

In the traditional power grid, energy generation facilities are generally monitored with wired sensors

- limited in numbers

- located only at a few critical places

In practice, the renewable energy generation facilities in the smart grid can be in remote areas, and operate in harsh environments

WSNs offer an ideal technology for continuous monitoring and control of the generation facilities in the smart grid.

Transmission and Distribution side

- Real time monitoring and securing of the transmission and distribution segment: substations, transmission lines and outages

- Substations transform voltage from high to low, or the reverse. Substation failure consequence can be very severe.

- Xcel Energy is the firs power company to use UAV (Unmanned Aerial Vehicles) / drones inpect substations in 2015.

- Close monitoring in case of lightning, icing, hurricanes, land slides, overheating.

- Locate failure of power lines; and detect failure and relay information to control stations.

3.10 More Considerations

3.10.1 Microgrid

"A group of interconnected loads and distributed energy resources (DER) with clearly defined electrical boundaries that acts as a single controllable entity with respect to the grid and can connect and disconnect from the grid to enable it to operate in both grid-connected or island mode." - *US. Department of Energy definition*

Q: In many respects, microgrids are smaller versions of the traditional power grid. Why shall we need microgrid?

- Resilient power supply
- Easy integration of local renewable energy into the grid
- Increased local control, reliability and security of power

3.10.2 Electric Vehicles (EV)

In Electric Vehicles and V2G (Vehicle to Grid) infrastructure, there are information exchange requirements between EV, charging stations and billing and management systems.

Q: what communication technology can be used for this requirement?

Power line communication can be a solution, utilizing installed power lines, providing strong security and enabling large number scalability.

3.10.3 Data Management

MDMS (Meter Data Management Systems) performs long term data storage and management for the vast quantities of usage data and events.

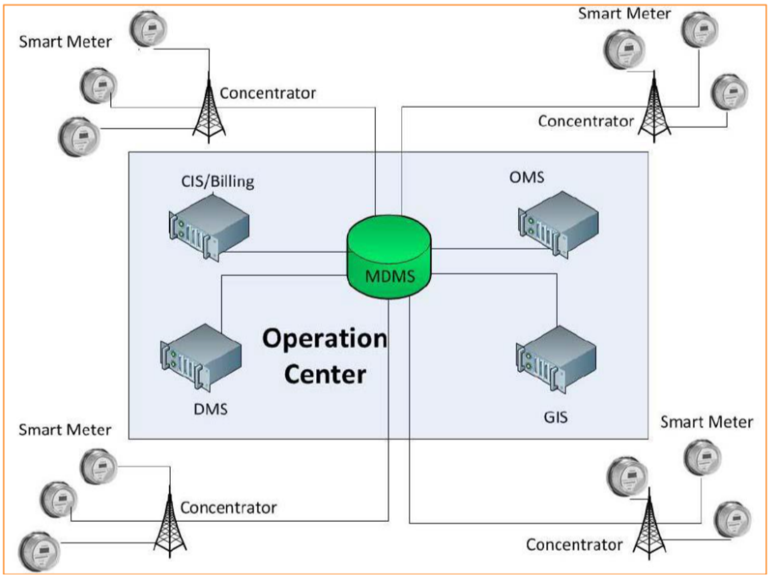
MDMS enables the interaction with operation and management systems that:

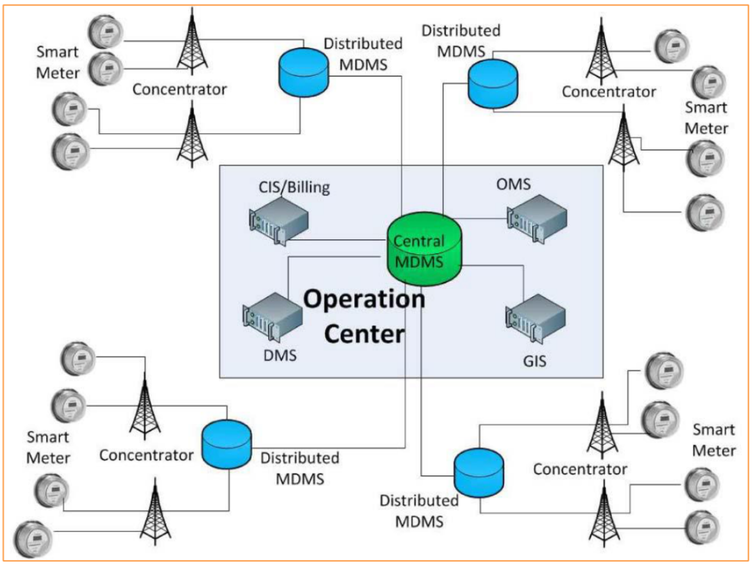
- manages the billing and customer information
- provides power quality report and load forecasting based on meter data

Q: all data from the smart meters need to go through the centralized MDMS, it makes the system non-scalable. Why?

One approach: distributed data management

Several distributed MDMSs are close to smart meters. Each MDMS is responsible for the specific area. Communicationdistance for data collection is largely shortened, and the corresponding resources needed.

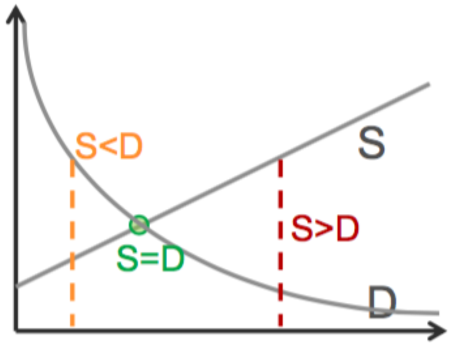




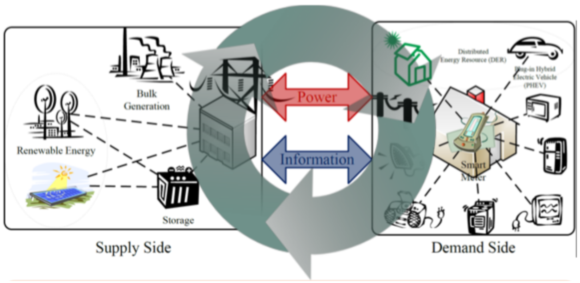
4 Demand Response Management (DRM)

Energy balance: power generation is equal to power demans

- S: power supply from generators
- D: power demand from customers
- Energy balance point: S = D
- Energy balance is very important for energy systems stability and economy **-Q: why?**
- If S > D, power generation is too much, generators have financial loss
- if S < D, power generation is not sufficient, power blackout may happen



Demand Response Management (DRM) is the main approach to achieve energy balance.



DRM studies the interaction between the supply and the demand sides by two-way flows of power and information.

4.1 Definition

According to the US. Departement of Energy:

Demand Response Management (DRM) is defined as changes in electric usage by end−use customers from their normal consumption patterns in response to changes in the price of electricity over time, or to incentive payments designed to induce lower electricity use at times of high wholesale market prices or when system reliability is jeopardized.

In plain language:

Users will change energy usage behaviors according to different electricity prices, or incentive payments, or system reliability.

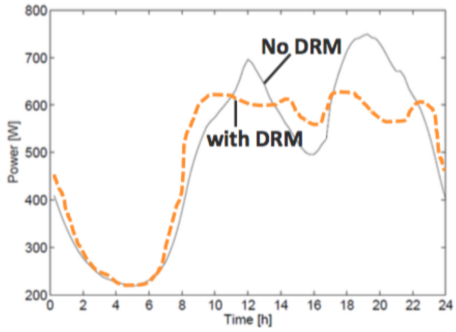
4.2 DRM two objectives

Reduce energy consumption

- encourage energy-aware consumption patterns
- Reduce power generation

Shift the energy consumption

- Mitigate power load during the peak hours
- Improve grid reliability



4.3 Two approaches to DRM

Direct Load Control (DLC)

- The utility has remote access to certain load of users
  - Air conditioner
  - Water heater
- The utility can remotely turn on or off the load when needed

Intelligent Load Control / Smart Pricing

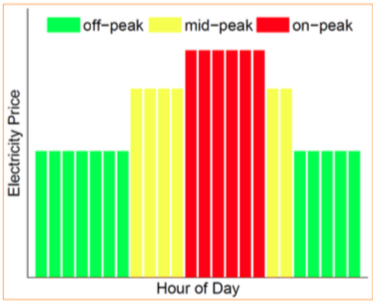
- Price-based program provides users with different prices at different times
- When users know about the price changes
  - they will naturally use less electricity when electricity prices are high
  - this will reduce the demand at peak hours
- This program indirectly induces users to dynamically change their energy usage patterns according to the variance of electricity prices, instead of directly controlling their loads.
- Three pricing models: *Time-of-Using (ToU) pricing*; *Real-time Pricing (RTP)*; and *Inclining Block Rates (IBR)*

4.3.1 Time-of-Use (ToU) Pricing

- When users consume energy at different time intervals of a day, they are charged at different electricity prices
- ToU pricing is usually released far in advance, and keeps unchanged for a long time period.

Examples:

- Ontario (Canada)
- Ausgrid (Australia)



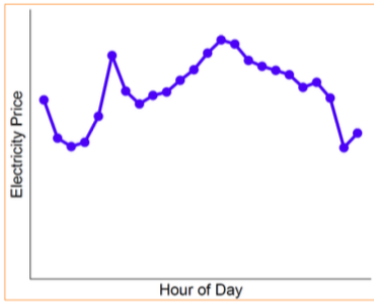
Three-level (on-peak, mid-peak, off-peak) ToU pricing

4.3.2 Real-Time Pricing (RTP)

- The electricity price usually varies at different time intervals of a day (ex.: in each hour)
- RTP is usually released on an hour-ahead pricing or day-ahead pricing basis

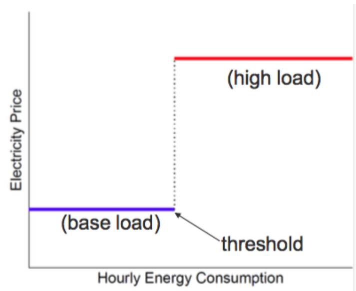
Examples:

- Chicago uses hourly-based RTP
- In Oslo, you may have RTP from Sognekraft AS in the name of "innkjøpspris" plan (source: https://www.strompris.no/)



4.3.3 Inclining Block Rates (IBR)

- Two-level rate structures
  - base load and high load
  - Price increases sharply if energy usage exceeds threshold
- Motivations
  - A user pays more when consuming more energy
  - Users evenly distribute loads among different times of a day to avoid higher rates



Adopted by Pacific Gas and Electric, USA; British Columbia Hydro, Canada

4.3.4 Electricity Pricing:

Fixed Price:

A fixed price for an agreed period of time, normlly one year. In exchange you cannot change supplier within the period

Variable Price:

Price will be as offered for the next 2-3 weeks. Maybe changed, normally with 2 weeks notice.

Purchase Price:

Price follows **the hourly prices** at the electricity exchange Nord Pool Spot (*Similar as real-time pricing*)

4.4 Home Energy Management System

In a house, a smart meter can automatically coordinate appliances to satisfy the user's need via ON/OFF control.

Energy Consumption Scheduling is actually not simple since we have many different appliances

Non-shiftable appliances (e.g., TV, lights, cooking): must be kept ON for a certain period of time.

Shiftable appliances (e.g., washing machines, Electric Vehicles, and clothes dryers): the operation task can be shifted to a different time period.

4.4.1 Energy Consumption Scheduling

Q: Given the price values, how should we schedule the power load?

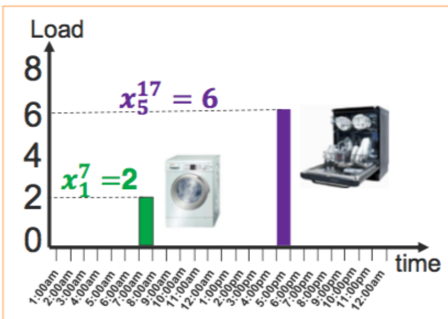
Scheduler should analyze user's energy consumption needs *and* price values.

The schedule is normally an optimal solution with different preferences **tradeoff**

- Minimize the cost of electricity
- Maximize user's comfort

Let A denote the set of appliances:
Washing machines, TV, lights, Dryer, Dishwasher, EVs (Electric Vehicle)...
For each appliance a ε A, we define a daily energy consumption scheduling vector X_a as follows:

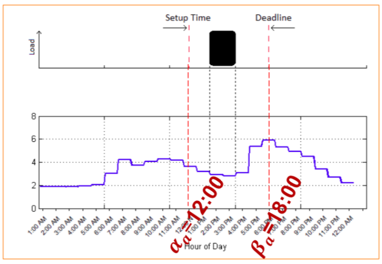
$$X_a = [X_a^1, ..., X_a^H]$$



For each appliance a ε A, the user should indicate:

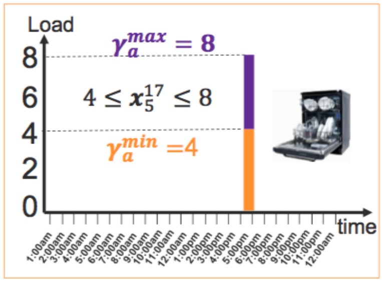
- α_a: beginning of the operation time (**setup time**)
- β_a: end of the operation time (**deadline**)

Operation should be scheduled within $[\alpha_a, \beta_a]$



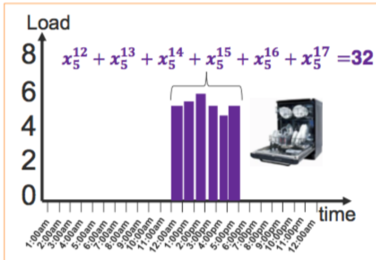
Each appliance a ε A usually has a maximum power level γ_a^{max}
Each appliance a ε A may also have a minimum power level γ_a^{min}
For each appliance a ε A, it is required that:

$$\gamma_a^{min} \leq x_a^h \leq \gamma_a^{max}$$
$$\forall_a \ \varepsilon \ \mathrm{A},\ \mathrm{h} \ \varepsilon \ [\alpha_a, \beta_a]$$



Let E_a denote the total energy needed for the operation of appliance a ε A
Given parameters E_a, a_a, and β_a, it is required that

$$\sum_{h=\alpha_a}^{\beta_a} x_a^h = E_a, \ \mathrm{a} \ \varepsilon \ \mathrm{A}$$



4.4.2 Cost Minimization

Energy Consumption Scheduling problem to minimize cost

$$\min_{x} \sum_{h=1}^{H} p^h \times \underbrace{\left(\sum_{a \in A} x_a^h\right)}_{\text{Load from all appliances in hour } h}$$

Subject to
$$\sum_{h=\alpha_a}^{\beta_a} x_a^h = E_a, \qquad \forall a \in A,$$
$$\gamma_n^{\min} \leq x_a^h \leq \gamma_n^{\max}, \qquad \forall a \in A, h \in [\alpha_a, \beta_a]$$
$$x_a^h = 0, \qquad \forall a \in A, h \notin [\alpha_a, \beta_a]$$

– where p^h denote the price of electricity at hour h. Could be ToU or RTP model

Optimization Basics

- Optimization in standard form

$$\begin{array}{ll} \text{minimize} & f_0(x) \longrightarrow \text{objective function} \\ \text{subject to} & f_i(x) \leq b_i, \quad i = 1, \ldots, m \end{array}$$

constraints functions

- $x = (x_1, x_2, \cdots x_n)$: optimization variables
- Optimal solution: x^* has smallest value of f_0 among all vectors that satisfy the constraints

Linear Programming Basics

$$\begin{array}{ll} \text{minimize} & f^T x \\ \text{subjec to} & Ax \leq b \end{array}$$

when $f = (f_1, f_2, \cdots, f_n)$; $b = (b_1, b_2, \cdots, b_n)$; A=$\begin{pmatrix} a_{1,1} & a_{1,2} & \cdots & a_{1,n} \\ a_{2,1} & a_{2,2} & \cdots & a_{1,n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m,1} & a_{m,1} & \cdots & a_{m,n} \end{pmatrix}$

- Matlab **linprog** function solves linear programming optimization problem

$$\min_x f^T x \text{ such that } \begin{cases} A \cdot x \leq b, & \longrightarrow \text{inequality constraint} \\ Aeq \cdot x = beq, & \longrightarrow \text{equality constraint} \\ lb \leq x \leq ub. & \longrightarrow \text{bound constraint} \end{cases}$$

x = linprog(f,A,b,Aeq,beq,lb,ub)

For more information: http://se.mathworks.com/discovery/linear-programming.html

4.5 More Considerations

Demand response applications: Data centers

Daily services supported by data centers

- Gmail
- Facebook
- Dropbox
- DNB (supported by Green Mountain data center in Stavanger)

Q: Why place data centers in Finland or undersea?

Data centers are huge energy consumers, in particular cooling systems, and

- pay a lot for electricity bill
- make power grid instable during peak hours

Cooling system uses sea water from the Bay of Finland and reduces energy use

Q: can DRM help data centers to reduce energy cost?

Allocate computation tasks to locations with cheaper prices

5 Energy Market and Game Theory

5.1 Game Theory

5.1.1 Game theory as a discipline

- Game theory is a discipline that is used to study problems of conflict among interaction decision makers

Three key elements in a game

- **Players**: there are multiple players. Each player can be an individual, a group, an organization, and so on
- **Strategies**: a plan of actions carried out
- **Playoff**: amount received for a given strategy

5.2 Strategy

- **Definition**: a strategy is a plan of actions by which a player has a decision rule to determine their moves for every possible situation in a game
- Two types of strategies: a pure strategy and a mixed strategy
- *A pure strategy*: at every stage in the game, it specifies a particular move with complete certainty
- *A mixed strategy*: applied some randomization to at leat one of the moves. The randomization is a set of fixed probabilities, where the sum of the probabilities is 1

5.2.1 Strategy: travel example

To travel from home to office such that you reach as soon as possible, one can use public transport, private vehicle, one can cycle to work.

- A **pure** strategy: One always takes public transport or always uses his/her car/bike
- A **mixed** strategy: One sometimes uses train, sometimes car and some other times bike

5.3 Two types of games

- Non-cooperative game
  - a game with competition between individual players. Only self-enforcing (e.g., through credible threats) alliances (or competition between groups of players) are possible due to the absence of external means to enforce cooperative behavior (e.g., contract, law)

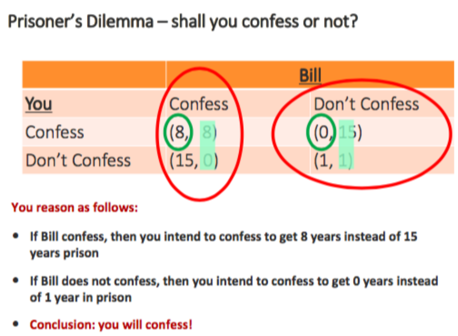
- Cooperative game
  - a game with competition between groups of players due to the possibility of external enforcement of cooperative behavior (e.g., through contract)
  - Cooperation generally leads to higher payoffs. For example: countries cooperative on trading (reduced tariffs) leading to boost in exports

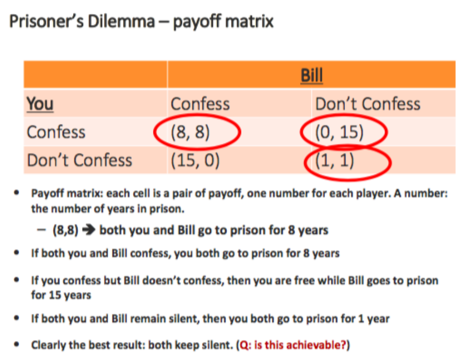
5.3.1 "Prisoner's Dilemma"

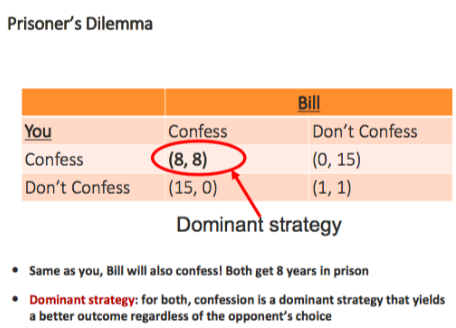
You and your friend Bill are arrested and thrown into prision in separate cells. For three days, neither you nor Bill has told nothing, then.. You are told by FBI: "We have your friend Bill and he is starting to talk". Should you confess?

How to choose strategy to minimize the years in prison?

- Each prisoner is rational and selfish. Namely, he wants to maximize his own benefit and does not care about the other person's benefit
- Each prisoner is put in separate room and does not know the other person's choice.
- When you and Bill can talk, you may not trust even if Bill claims to cooperate







Nash Equilibrium(NE): Acombinationof strategies is called a Nash Equilibrium if neither player has an incentive to change strategy, given the other player's choice mutual best response.

- Best response: the strategy which produces the most favorable outcome, taking other players' strategies as given

Both confess is a Nash Equilibrium

Paradox: both you and Bill confess and go to prison for 8 years, whereas if you both keep silent and you would have spent 1 year each in prison!

Diagnosis: equilibrium need to be efficient. Non-cooperative equilibrium in the Prisoner's dilemma results in a solution that is not the best possible outcome for the parties.

Conclusion: individual's best choice is not the group's best choice. An individual's rational choice may lead to group's non-rational choice

Cooperative solution: they could both become better off if they reached the cooperative solution (which is why police interrogate suspects in separate rooms)

Both don't confess is not a Nash Equilibrium

	Bill	
You	Confess	Don't Confess
Confess	**(8, 8)**	(0, 15)
Don't Confess	(15, 0)	(1, 1)

Bill will confess, then be in prison from 1 year to zero

- **In both don't confess , rival will always want to deviate**
- **(1,1) will change to (15,0) or (0, 15) since confess leads to less years in prison, which eventually leads to (confess, confess)**

5.4 Enegy Market

5.4.1 Deregulated Energy Market

In **regulated energy market**, prices are all determined by the regulatory/government bodies:

- energy prices
- transmission and distribution prices

And, you cannot choose supplier

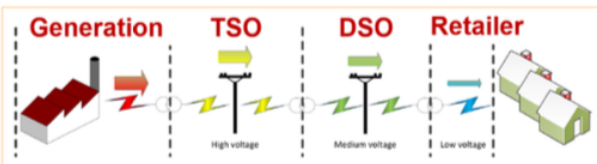
In **deregulated energy market**, prices are determined by "invisible hand" of the market. There is competition among a set of suppliers. Norwegian electricity market was deregulation in 1991. Deregulation allows different power suppliers to offer services to consumers. It allows you to choose your supplier.

5.5 Energy market players

5.5.1 power grid operators

Transmission system operators (TSOs): The TSO operates the transmission assets and is responsible for the power balance on the transmission system (e.g., Statnett)

Distribution system operators (DSOs): power distribution system to operate the distribution grid and transmit electricity to residental customers (e.g., Hafslund)



5.5.2 sell and buy

Generating company (genco): The generators own production assets, whose generation is offered through the electricity market.

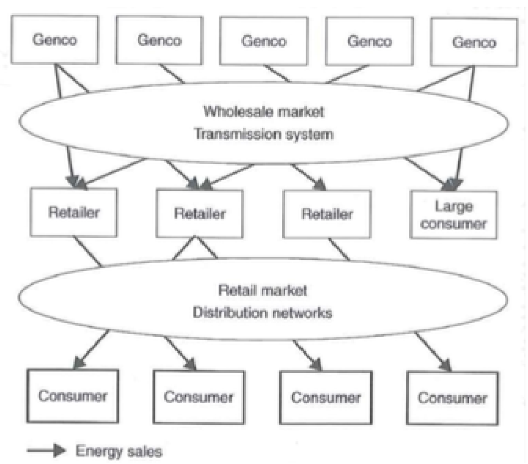
Retailer: The retailer buys electricity from the electricity market, then sell to the end-consumers.

Customers: Those eventually use the electricity for any purpose (from watching TV to heating to industrial production processes). There is a difference between small and large consumers, since the latter ones may be allowed to directly participate in the electricity market.

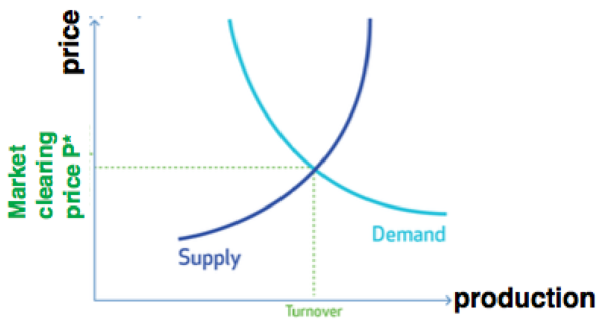
5.5.3 rule and operate the game

Regulators: regulators effectively 'police' the energy market. The regulator is responsible for the market design and its specific rules. It also monitors the market in order to spot misbehavior in electricity markets (collusion, abuse of market power, etc.) (e.g., NVE - Norwegian Water Resources and Energy Directorate)

Market operator: power exchange platform used by market players to negotiate purchases and sales of electricity (e.g, Nord Pool).



Supply: power supply from generators
Demand: power demand from users
Market clearing price P* when power supply is equal to power demand (i.e., Supply = Demand)



Nord Pool: power exchange platform used by market players to negotiate next-day purchases and sales of electricity. Norway uses day-ahead trading system where buyers and sellers send orders. The market price is important to establish equilibrium between supply and demand

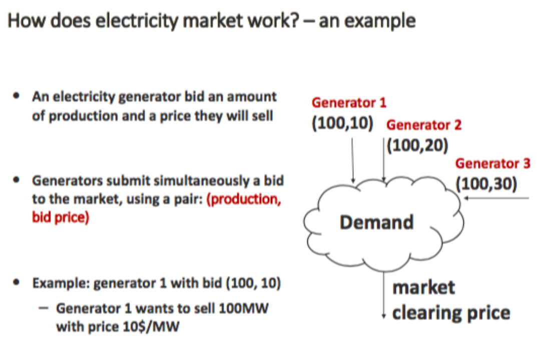
5.6 Day-ahead market

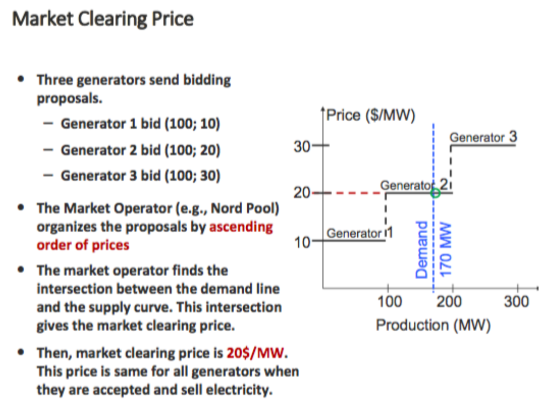
A buyer, typically a utility, needs to assess how much energy it will need to meet demand the following day, and how much it is willing to pay for this energy. The seller, e.g., a power plant, needs to decide how much it can deliver and at what price.

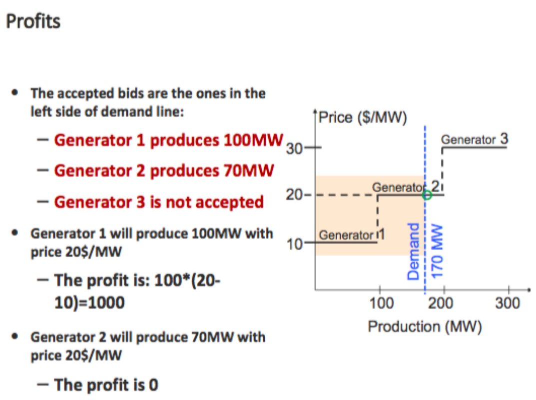
Day-ahead market: contracts are made between seller and buyer for the power delivery the following day, the price is set and the trade is agreed. The day-ahead market at Nord Pool becomes a *concrete visible hand* to creates equilibrium. It is an auction based exchange for the trading of prompt physically delivered electricity.

Auction in electricity pool

- All generation bids and consumption offers are placed at the same time
- No-one knows about others' bids and offers
- An algorithm decides about bids and offers that are retained
- Eventually, the system operator is informed about the trades that occurred







5.7 Electricity Market Strategies

- Electricity produces game using their price and production. By changing these two parameters, the market sharing changes.
- **Cournot strategy**: reduce production. Two firms compete simultaneously on the quantity of output they produce of a homogeneous good.
- **Bertrand strategy**: changing price. Two firms compete simultaneously on the price of a homogeneous good.

5.7.1 Curnout Model

- Non-cooperative: generators are non-cooperative and independently decide their own production. Generators set the production simultaneously.
- Rational and selfish: each generator will choose production to maximize its own profit
- A generator should decide how much electricity to produce in order to maximize its profit without knowing the decision of the others.
- Electricity price will be determined by the demand curve and supply curve where the total supply is equal to the total demand.

6 Green Data Center

6.1 Data Center Basics

Daily services supported by data centers: Search, Gmail, Google map, Youtube, Driving, Translator

Norway is very active in the data center business

Norway can provide unique cooling technology (cold seawater, close to fjord.

Many data center players: Green Mountain, Lefdal Mine, Digiplex, Nordavind

6.2 Data Center Definition

A data center is a building/container used to house computing equipment such as servers, along with associated components such as networking devices, storage systems, power distribution units, and cooling systems.

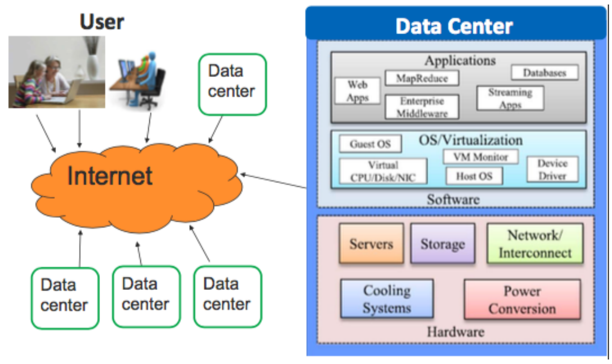
<u>Common features</u>:

- Power: equipped with a guaranteed power supply
- Communications: equipped with high bandwidth connectivity
- Reliability: duplication of networks, power and IT infrastructure
- Environment: maintain a specified temperature and humidity
- Security ensure that the facility and its data remain secure

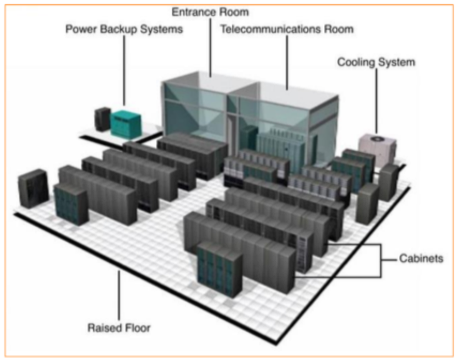
6.2.1 Modular Data Center

- Rapid deployment and can connect to water and power supply and run. Built-in cooling, high PUE (power usage effectiveness)
- Very suitable for disaster recovery, e.g., flood, earthquake (offered by Cisco, IBM, SGI, Sun/ORACLE).
- *Small*: 8 foot, power < 1 MW, 4 racks per unit;
  Medium: 20 foot, power: 1-4 MW, 10 racks per unit;
  Large: 20 foot, power > 4 MW, 20 racks per unit
- Negative aspect: the containers can be damaged while being moved or transported, always needing recalibration.

6.2.2 Hardware/Software Components



6.2.3 Physical Layout

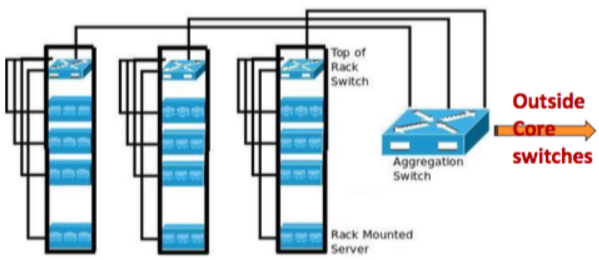


6.3 Top of Rack (ToR) architecture

Servers are connected to switchers within the server racks. Two high-speed and individual switches connect to the aggregation switch.

ToR architecture refers to the physical placement of the switch in the top of a server rack. Servers are directly linked to the switch. Each rack usually has one or two switches (Q: why two switches?)

All the ToR switches are connected with the aggregation switch. Only a small amount of cables are needed to run from server rack to aggregation rack. Aggregation switch will be further connected to the outside core switches.



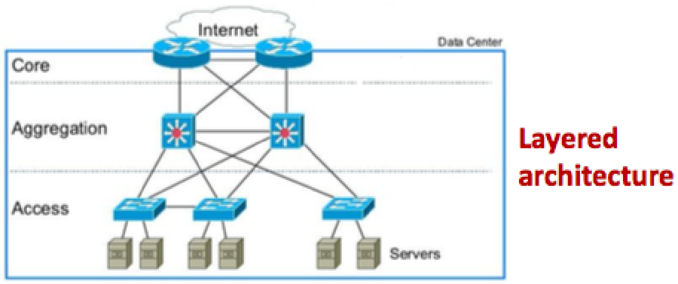
Advantages:

- No large cabling infrastructure is required. Cable management is also easy with less cables involved.
- Low cabling costs since all servers connections are terminated to its own rack
- Flexible "per rack" architecture. Easy "per rack" changes.

Disadvantages:

- High management burden: Each rack switch is individually managed, management burden will increase significantly by adding many new switches to the data center. For example in a data center with 50 racks, where each rack contained 2 ToR switches, the result would be 100 switches totally. There are 100 copies of switch software that need to be updated, 100 configuration files that need to be created and archived.
- Potential scalability concerns: ToR design typically requires higher port densities in the aggregation switches.

6.4 A 3 Layer Data Center Network Architecture

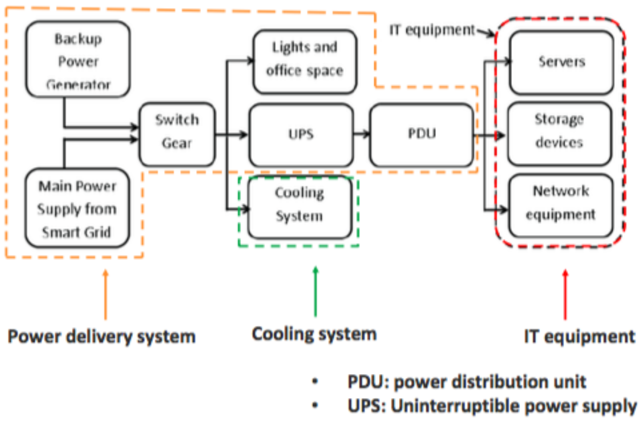


- Servers: Racks of equipment that require network access.
- Access Layer: Equipment directly connected to servers
- Aggregation Layer: Equipment that aggregates access layer devices to provide connectivity among Access Layer domains
- Core: Equipment that interconnects aggregation layer devices either within a data center or across geographic locations with outside world

Features of Data Center Architectures

- Regular, well-defined arrangement
- Hierarchical structure with rack / aggregation / core layers
- Mostly homogeneous within a layer
- Supports communication between servers and between servers and the external world
- Contrast: ad-hoc structure, heterogeneity of WANs (Wide Area Networks)

Major Components in a Data Center



6.5 Data Center Major Components

6.5.1 IT equipments

- There are three kinds of IT equipments hosted in a typical data center

- servers for data processing
- storage equipment for data storage
- network equipment for data communications

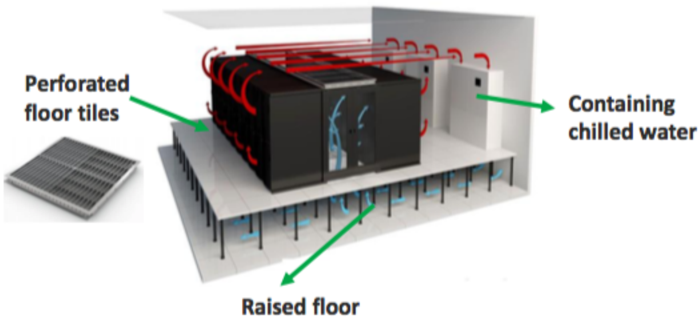
- These three types of equipments are collectively known as "information technology" (IT) equipments

6.5.2 Power delivery system

- The power delivery system mostly contains power conversion units, voltage regulators and backup equipment.
- Power backup is often provided by Uninterruptible Power Supply (UPS) unit which prevents the IT equipment from experiencing power disruptions and possible serious business disruption or data loss.

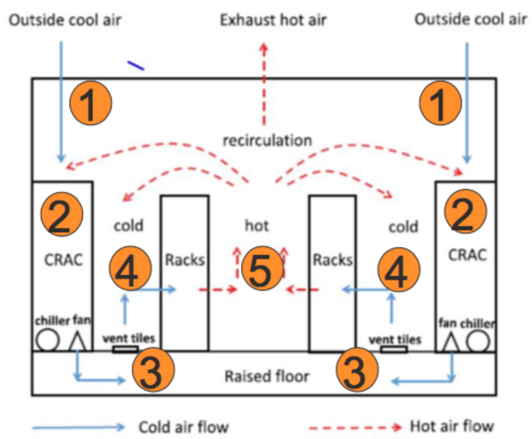
6.5.3 Cooling systems

Cooling in data centers is often provided by computer room air conditioning (CRAC) units.



6.6 Air flow in Data Center

1. Outside cool air enters the top of a CRAC unit.
2. Air is conditioned by passing chilled water pumped from a chiller located outside the building.
3. The chilled air goes through a raised floor.
4. The chilled air is then supplied to the IT equipment (primarily servers).
5. The cold air, while passing through perforated floor tiles, is pulled by the fans located inside the servers.

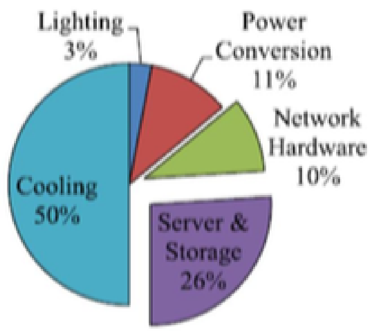


Data centers are huge energy consumers, in particular cooling systems, and pay a lot for electricity bill, make power grid instable during peak hours. Cooling system uses sea water from the Bay of Finland and reduce energy use. Google says it spends about 0.0003 kWh of energy on an average search query. Energy per search seems low, but Google had 2.3 million search per second in 2016.

6.7 Energy Consumption in Data Center

Energy Consumption in Data Center

- The largest energy consumer in a typical data center is the cooling infrastructure (50%), while servers and storage devices (26%) rank second in the energy consumption hierarchy. – according to the statistics published by the Infotech group. (**Note:** that these values might differ from data center to data center)
- A breakdown of energy consumption by different components shows: the cooling infrastructure consumes a major portion of the data center energy followed by servers and storage, and other infrastructure elements.



Energy Efficiency Metrics for Data Centers

In order to quantify the energy efficiency of data centers, several energy efficiency metrics have been proposed to help data center operators to improve the energy efficiency and reduce operation costs of data centers.

Two important energy efficiency metrics are:

- Power Usage Effectiveness (PUE)
- Data Center energy Productivity (DCEP)

6.7.1 Power Usage Effectiveness (PUE)

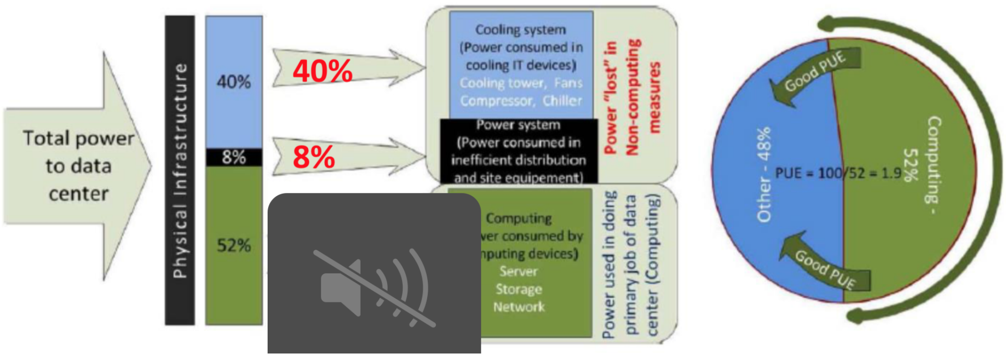
- PUE: the most commonly used metric to indicate the energy efficiency of a data center-
- PUE definition: ratio of total data center energy usage to IT equipment energy usage

$$PUE = \frac{\textit{Total Power Consumption of a Data Center}}{\textit{Total Power Consumption of IT Equipment}}$$

- **Total power consumption of a data center**: the sum of power used by all components in a data center, including cooling, lightning, and IT equipment.
- **Total power consumption of IT equipment**: the sum of power used by IT equipment (servers, storage, and network).

- PUE is a value greater than 1 since data centers draw considerable amount of power as non-IT power.

An example



Distribution of data center power among cooling, power distribution, and computing units. In this example, 25% power goes to computing while 48% power goes to non-computing. Then, PUE = 100/52 = 1.9

PUE features:

- PUE measures the total power consumption overhead caused by the data center support equipment, including the cooling systems, power delivery, and other facility infrastructure like lightning.s

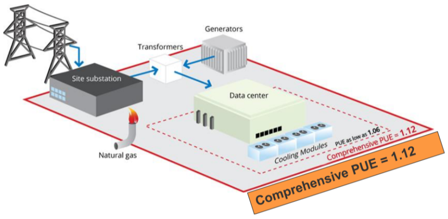
- PUE = 1.0 implies: there is no power overhead and all power consumption of the data center goes to the IT equipment.

- A higher PUE: a greater portion of the electricity coming to the data center spent on cooling and the rest of the infrastructure.

  Q: PUE is a good metric?

- A good PUE value may not be enough to guarantee the global efficiency of the data center. PUE metric does not consider the actual utilization (applications and workloads) of computational resources, some computation may not be necessary.

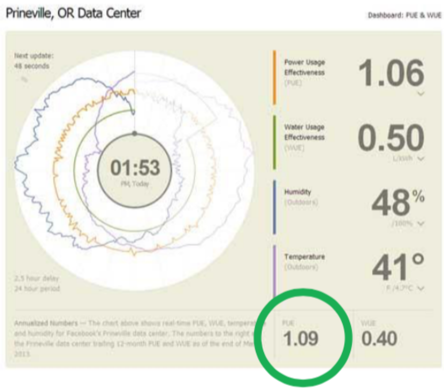
Google's PUE



- A PUE of 2.0 means that for every watt of IT power, an additional watt is consumed to cool an distribute power to the IT equipment. A PUE closer to 1.0 means nearly all of the energy is used for computing.
- Google reports: the average PUE for all Google data centers is 1.12. Bet sites can be less than 1.06.

Facebook data center PUE

- Facebook provides dashboards showing real-time PUE measurements for two of its largest data centers Prineville, Oregon, and Forest City, North Carolina.
- The measurement report PUE averages of 1.08 and 1.09.



6.7.2 Data Center energy Productivity (DCeP

- Energy efficiency and energy productivity are closely related to each other.
  - Energy efficiency focuses on reducing unnecessary power consumption to produce a work output
  - Energy productivity measures the quantity of useful work done relative to the amount of power consumption of a data center in producing this work.
- DCeP measures the useful work performed by a data center relative to the energy consumed by the data center in performing the work

$$DCeP = \frac{Useful\ Work\ Produced}{Total\ Data\ Center\ Power\ Consumed\ Producing\ this\ Work}$$

- DCeP metric allows the user to define the computational tasks, transactions, or jobs that are of interest, and then assign a measure of importance of economic value to each specific unit of work completed
- DCeP allows the continuous monitoring of the productivity of a data center as a function of power consumed by a data center.
- DCeP metric tracks the overall work product of a data center per unit of power consumption expended to produce this work.

  Q: Easy to use this metric?

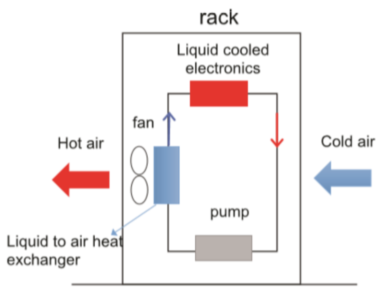
  - Subjective and may not be easy to define "Useful Work"

6.8 Green Data Center

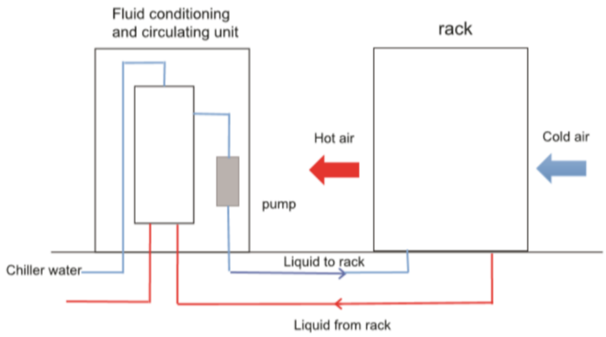
- "Green Data Centers": energy-aware, energy-efficient and CO2 emission minimization designs, protocols, devices, infrastructures and algorithms for data centers.
- Reduce energy consumption in an individual data center
- Reduce energy consumption across globally located data center

6.9 Techniques to Improve Energy Efficiency of Data Centers

6.9.1 Smart Cooling and Thermal Management



Internal liquid cooling loop: electronics components, e.g., CPUs, inside the rack are cooled by a a liquid cold plate, with heat released to air through a heat exchanger. Hot air will load the room, which can increase the room temperature. This impact can be reduced by direction the heat to a water chilled system.



External liquid cooling loop: an external refrigeration chiller system is installed outside the server. The chiller system cools the CPUs using a liquid coolant.

6.9.2 Power Management - Chip Level

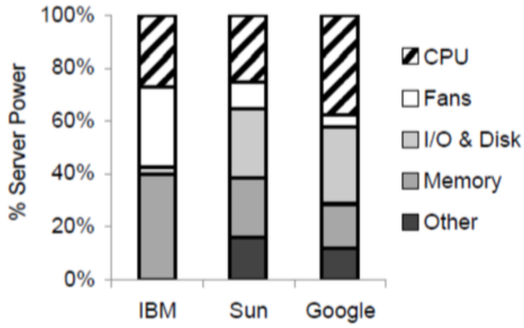
- Dynamic voltage and frequency scaling (DVFS) is a commonly-used techniques to save power at chip level power management
- **Main principle**: DVFS is able to reduce the power consumption of CPU by reducing the operating frequency, as shown by:

$$P = CfV^2 + P_{static}$$

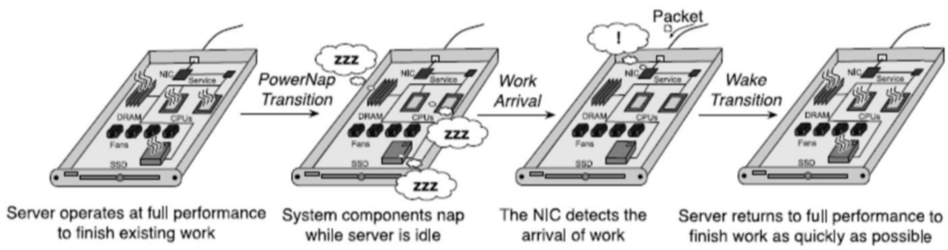
- where **C** is the physical capacitance, **V** is the supply voltage, and **f** is the operating frequency.
- The voltage is determined by the frequency at which the circuit is clocked and can be reduced if the frequency is also reduced. This can yield a significant reduction in power consumption because of the V^2 relationship.

6.9.3 Power Management - Server Level

- A server has many components, including CPU, fans, disk, memory, etc.
- DVFS can be highly effective in reducing CPU power. However, CPUs account for a portion of total system power. See typical server power breakdowns for the IBM p670, Sun UltraSparc T2000, and a generic server by Google.
- We need to approach that manage the power in the entire server, e.g., transition into a low-power state or sleeping state when there is no job.



PowerNap for a Server



- Each time the server is done with all pending work, it transits to the nap state.
- In the nap state, nearly all system components enter sleep mode, which are already available in many components. Power consumption is low, but no processing can occur. System components that signal the arrival of new work, expiration of a software timer, or environmental changes, remain partially powered.
- When new work arrives, the system wakes and transitions back to the active state. When the work is complete, the system returns to the nap state.

Network Equipment Level: switches

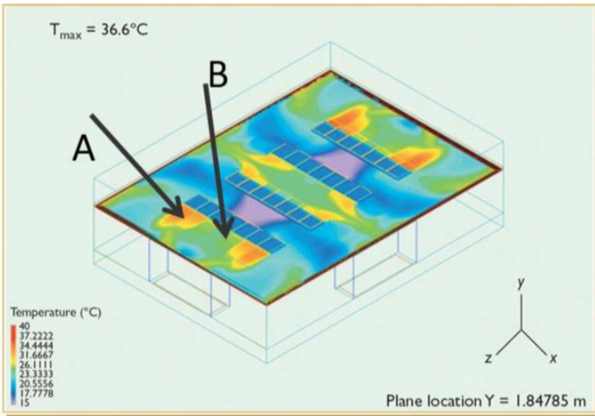
- Sleeping mode
  - Transit into the low-power sleep mode when no transmission is needed, and return back to the active mode when transmission is requested.
  - The transition time overhead of putting a device into and out of the sleep mode may reduce energy efficiency significantly.

- Rate adaptive scheme
  - Observation: Devices consume less energy when the transmission rate is lower
  - Main Idea: Adapt the transmission rate of network operation to the offered workload. Speed negotiation is required in the rate-adaption scheme for both of the transmission ends.
  - IEEE 802.3az standard has been standardized that make use of adaptive link rate technique for energy efficiency

6.9.4 Power Management - Data Center Level

Workload Migration

- Observation: different workloads can result in different power consumption.
- Reason:toomany computation tasks are running in Zone A while only few tasks in Zone B.
- **Q:** How to solve this?
- **Solution**: moving some computation tasks from Zone A to Zone B to have lower overall power consumption



Temperature distribution in data center

Dynamic Component Deactivation
Main idea: Allow servers to be dynamically turned off to save energy and only turn on minimum servers to finish tasks.

6.9.5 Power Management - Inter Data Center Level

- Different geographical distance from service requests
  - Your requests may be directed to the data center in Finland since you are geographically close to Finland
- Different working load
  - Your requests may be directed to the data center where there is low working load
- Different electricity price
  - Your requests may be directed to a data center with low electricity price to reduce energy cost
- Different renewal energy sources availability, greenness, and CO2 emission
  - Your requests may be directed to the data center where renewable energy is readily available

6.10 Energy Cost Problem

Workload Constraint

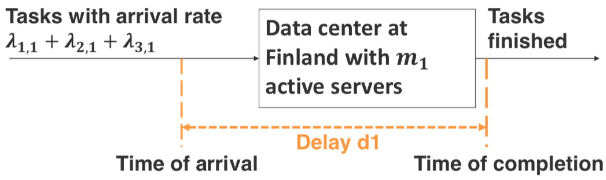
- A client request is first handled by a front-end Web server, then it is forwarded to one server at a specific location of the data center to be processed. The request arrival rate from front-end portal server j to data center i is $\lambda_{j,i}$. We denote the requests demand at each front-end Web server j as $L_j (j = 1, 2, \dots C)$. Therefore we have:

$$\sum_{i=1}^{N} \lambda_{j,i} = L_j; j = 1,2 \dots C$$

- At each data center, there are usually hundreds or even thousands servers to afford large number of requests. However, the number of active servers m_i should not be larger than the total number of servers M_i at each data center i. Therefore, we have

$$m_i \leq M_i; i = 1,2 \dots N$$

Delay requirement



- m active servers in a data center will deal with tasks from all customers.
- When all requests go to Finland with the lowest electricity price, the energy cost may be very low. However, the customers may have to wait for a very long time to get response which is not favorable.
- Hence, there is a constraint related to the allowed latency, e.g., 1 sec. That is, the service delay d_1 should not be longer than the allowed latency D_1.

6.11 Energy Cost Minimization

$$\min_{\lambda} \sum_{h=1}^{24} \underbrace{\left(m_1 P_1^h + m_2 P_2^h + m_3 P_3^h\right)}_{\text{Cost from all data centers in hour } h}$$

Subject to

$$\sum_{i=1}^{N} \lambda_{j,i} = L_j; j = 1,2 \dots C$$

$$m_i \leq M_i; i = 1,2 \dots N$$

$$d_i \leq D_i; i = 1,2 \dots N$$

- where P_i^h denote the price of electricity at datacenter i at hour h.

Q: Is this a linear programming optimization problem?

- No, the delay constraint is not linear.

6.12 More Consideration

Data Centers Use Renewable Energy as Power Sources

- Apple plans to build a 200 MW solar array to power its data center in Reno, Nevada
- **Q:** what are the main challenges data centers face in using solar power?
- **Reliability**: because of solar power's intermittency, data centers must be prepared to store it or supplement it. Centers typically use electricity from solar panels to power daytime operations, and apply any excess to charge UPS (uninterruptible power supply) systems. The challenge isn't in finding alternatives to supplement solar power, but in having the intelligence to optimize the use of each power source.
- **Quality**: The second issue must be addressed in order to protect sensitive IT assets i.e., power quality to stabilize against fluctuations is important.

7 Cloud/Fog Computing for Smart Grid

7.1 Cloud Computing

Evolution of Cloud Computing

- Utility Computing: 1961
- Time Sharing: 1970s
- Large Distributed Data Centers: 1980s-1990s
- Internet Computing: 2000-Present
- What is new in cloud computing today? – Faster data communication
  - Faster data communication
  - Faster and more reliable computing
  - Denser and cheaper storage
  - Newer Programming paradigms
- Comprehensive Computational resource sharing

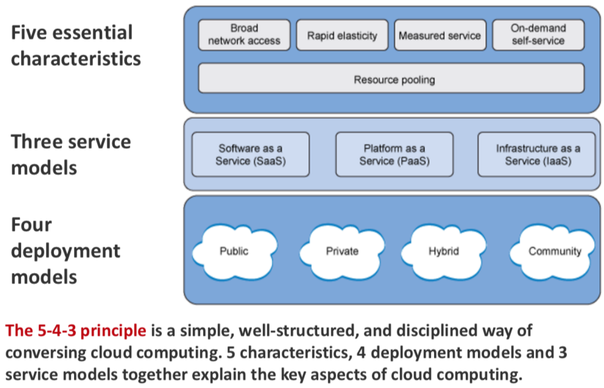
Cloud Computing definition

According to NIST (National Institute of Standard and Technology, USA):

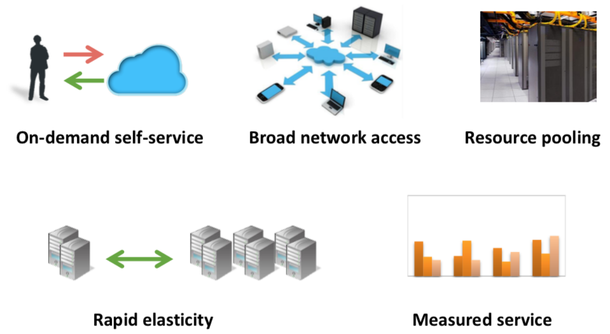
Cloud computing is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction.

Main features: cloud computing is an on demand model; shared pool of computing resources (servers , storage , applications), services rapidly provisioned, rapidly released and minimal management effort of service providers

NIST's 5-4-3 Principle for Cloud Computing



Five Essential Characteristics



7.1.1 Five Essential Characteristics

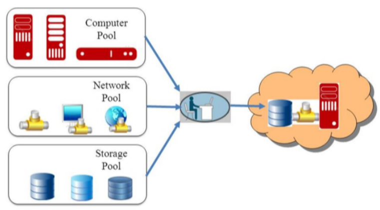
On-demand self-service: A consumer can provision computing capabilities, such as server time and network storage, as needed automatically without requiring human interaction with each service provider.

Plain language: It is there when you need it.

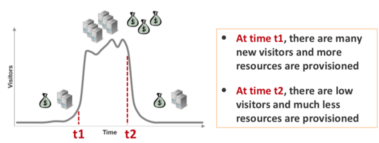
Broad network access: Capabilities are available over the network and accessed through standard mechanisms that are used by heterogeneous client platforms (e.g., mobile phones, tablets, laptops, and workstations).

Plain language: Tons of connectivity options

Resource pooling: The provider's computing resources are pooled to serve multiple consumers with different resources dynamically assigned and reassigned according to consumer demand. Examples of resources include storage, processing, memory, and network bandwidth.
Plain language: Sharing who-knows-where resources

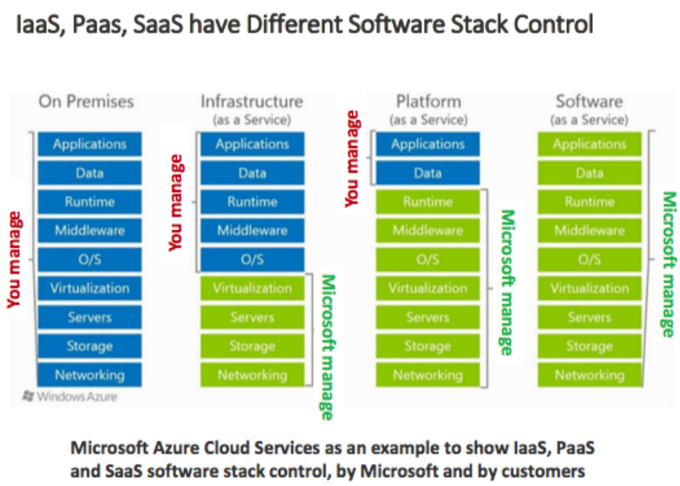


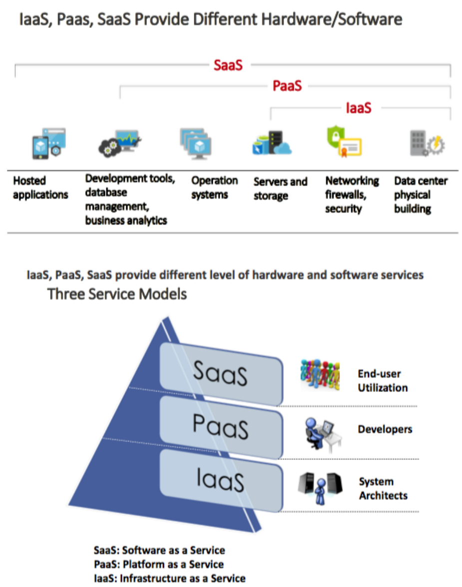
Rapid elasticity. Capabilities can be elastically provisioned and released to scale rapidly with dynamic demand.
Plain language: You get what you need. Ability of a system to increase or decrease its resources to meet the current demand



Measured service: Cloud systems automatically control and optimize resource use by leveraging a metering capability appropriate to the type of service (e.g., storage, processing, bandwidth, active user accounts). Resource usage can be monitored, controlled, audited, and reported, providing transparency for both the provider and consumer.

Plain language: You get what you pay for.





7.1.2 Three Service Models

IaaS: provide the consumer processing, storage, networks, and computing resources. The consumer is able to deploy and run arbitrary software, e.g., operating systems and applications.

The consumer does not manage or control the underlying cloud infrastructure but has control over operating systems, storage, and deployed applications.
Plain language: provision storage, CPU, network and other computing resources. Deploy and run your own OS and software. Example: Amazon Web Service (AWS)

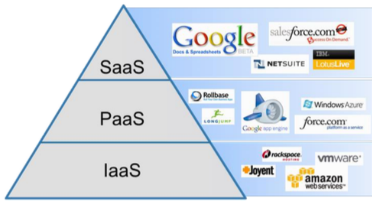
PaaS: provide the consumer to deploy onto the cloud infrastructure. Then, the consumer can create applications using programming languages, libraries, services, and tools supported by the provider.

The consumer does not manage or control the underlying cloud infrastructure including network, servers, operating systems, or storage, but has control over the deployed applications.
Plain language: use the languages, libraries, services and tools supported by the provider to deploy customers-created applications on the provider's network (e.g., "Heroku" which operates on top of the Amazon Web Services IaaS system)

SaaS: provide the consumer to use the provider's applications running on a cloud infrastructure. The applications are accessible from various client devices, e.g., a thin client, or web browser.

The consumer does not manage or control the underlying cloud infrastructure, or even individual application capabilities.
Plain language: the service is typically offered by third party software and web app developers and are hosted on IaaS and PaaS platforms.

i



7.1.3 Four Deployment Models

Public cloud: The cloud infrastructure is provisioned for open use by the general public. It may be owned, managed, and operated by a business, academic, or government organization.

Plain language: the cloud infrastructure is provisioned for open use by the general public.

Example: public cloud service providers like Amazon Web Service (AWS)

Private cloud: The cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers (e.g., business units). It may be owned, managed, and operated by the organization, a third party, or some combination of them

Plain language: the cloud infrastructure is provisioned for exclusive use by a single organization comprising multiple consumers

Example: Amazon AWS GovCloud

Hybrid cloud: The cloud infrastructure is a composition of two or more distinct cloud infrastructures (private, community, or public) that remain unique entities. They are bound together by standardized technology that enables data and application portability.

Example: An enterprise can deploy a private cloud to host sensitive workloads, but use a third-party public cloud provider, e.g., Google Compute Engine, to host less-critical resources

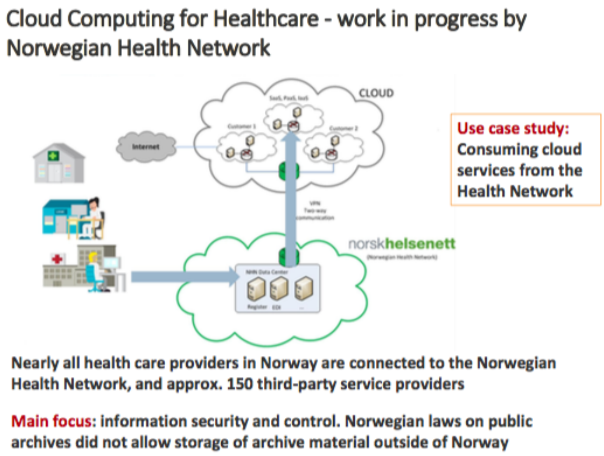
Community cloud: The cloud infrastructure is provisioned for exclusive use by a specific community of consumers from organizations that have shared concerns (e.g., mission, security requirements, policy, and compliance considerations). It may be owned, managed, and operated by one or more of the organizations in the community, a third party, or some combination of them.

Q: community cloud is a public or private cloud?

Alternative understanding: a community cloud is a private cloud of an associated community.

Cloud Computing for Transport in Norway
Background: Driving in slippery condition could be very dangerous. It will be easy to maneuver if you already knew where to look out for on the road. Volvo works with Norwegian Public Roads Administration to launch cloud-based system to share road friction information provided by individual cars.

Approach: 50 test cars on the roads. The Volvo test car detects an icy road, the information is transmitted to Volvo's database via the network ICE and then transmitted to other vehicles that are approaching the slippery area.



7.1.4 Challenges for cloud computing

New Observation 1: people as end-users are changing
An end-user is changing from a pure data consumer to both data consumer and data producer. This change requires more new functions at the end-user side.

Data Consumer: an end-user watches a YouTube video on a smart phone.

Data Producer: people are producing data from their mobile devices.

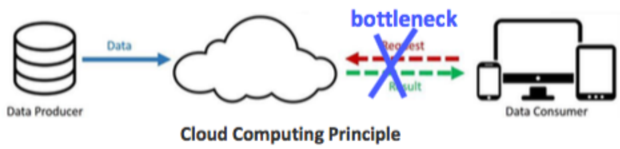
For *example*, every single minute, YouTube users upload 100hours of new video content; Instagram users post nearly 2430000 new photos

New Observation 2: things as end-terminals are changing
50 billion things (devices/terminals/sensors/vehicles) will be connected to the Internet by 2020 – CISCO

Data produced by such massive number of things will reach 500 zettabytes (10^{27} bytes). Comparatively, the global data center IP traffic will reach 10.4 zettabytes by that time - estimated by Cisco Global Cloud Index

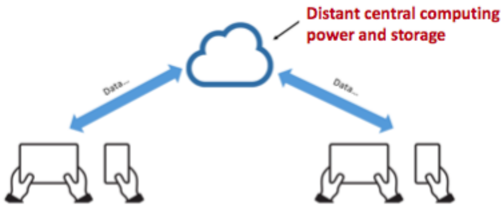
Q: what are new challenges for cloud computing when end-users and end- terminals are changing?

1st New Challenge for Cloud Computing – Data Transmission



With the growing data generation at the end, speed of data transmission is becoming the bottleneck for the cloud-based computing paradigm *For example*: about 5 GB data is generated by a Boeing 787 every second(*), but the bandwidth between the airplane and satellite or base station on the ground is not large enough for data transmission

2nd New Challenge for Cloud Computing –Latency



Cloud may have high latency which does not fit certain applications

An autonomous vehicle generates 1 GB data every second and it requires real-time processing to make correct decisions. If all the data needs to be sent to the cloud for processing, the response time would be too long.

3rd New Challenge for Cloud Computing – Privacy

Storing data and important files on external service providers in cloud always opens up risks

For *example*: for wearable health devices, since the physical data collected by the things is usually private, uploading raw data to cloud has the risk of privacy issues

New challenges motivate to put computation, storages, services close to the end-users to significantly reduce latency and protect private data.

Expectation: by 2019, 45% of data will be stored, processed, analyzed, and acted upon close to, or at the edge of the network

7.2 Fog Computing

According to CISCO

Fog Computing extends the cloud computing paradigm that provides computation, storage, and networking services between end devices and traditional cloud servers. Fog Computing nodes are typically located at the edge of network located away from the main cloud data centers.

Edge computing is interchangeable with fog computing. We define "edge" as any computing and network resources along the path between data sources and cloud data centers.

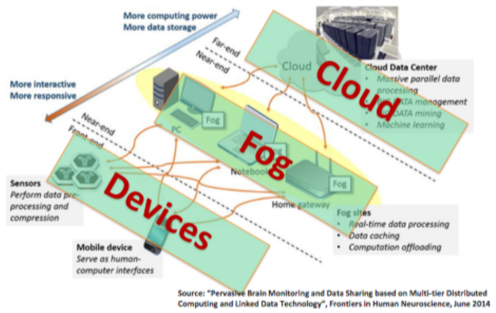
7.2.1 Computing Architecture

The initial definition of fog computing has been expanded. Fog computing now is not a mere extension of cloud computing, but a paradigm of its own.

Fog computing now refers to a three-tier architecture (Clients ↔ Fog nodes ↔ Central Servers)

- Centralized cloud servers coexist with fog nodes but are not essential for the execution of fog services
- Fog nodes can range from resource- poor devices (e.g. end devices) to more powerful cloud servers (e.g. Internet routers, 5G base stations).

Where is Fog or Edge? – an example

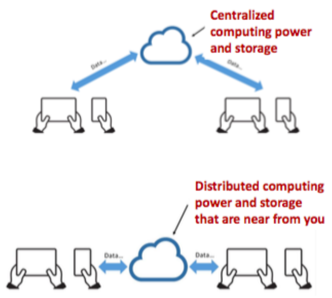


Three-layer architecture: end-devices ➔ Fog layer ➔ Cloud layer. We also call this Fog-to-Cloud (F2C) architecture

Main Difference between Cloud Computing and Fog Computing

Cloud computing: servers, computing power, storage are centralized. You do not know the specific location of the resources; and have no or limited control of the cloud infrastructure.

Fog computing: servers, computing power and storage are close to you. You may know the specific location; manage and control the resources.



Comparison of Fog Computing and Cloud Computing

	Cloud Computing	Fog Computing
Target Users	General Internet users.	Mobile users
Service Type	Global information collected from worldwide	Limited localized information services related to specific deployment locations
Hardware	Ample and scalable storage space and compute power	Limited storage, compute power and wireless interface
Distance to Users	Faraway from users and communicate through IP networks	In the physical proximity and communicate through single-hop wireless connection
Working Environment	Warehouse-size building with air conditioning systems	Outdoor (streets, parklands, etc.) or indoor (restaurants, shopping malls, etc.)
Deployment	Centralized and maintained by Amazon, Google, etc.	Centralized or distributed in regional areas by local business (local telecommunication vendor, shopping mall retailer, etc.)

More Comprehensive Comparison between Cloud Computing and Fog Computing

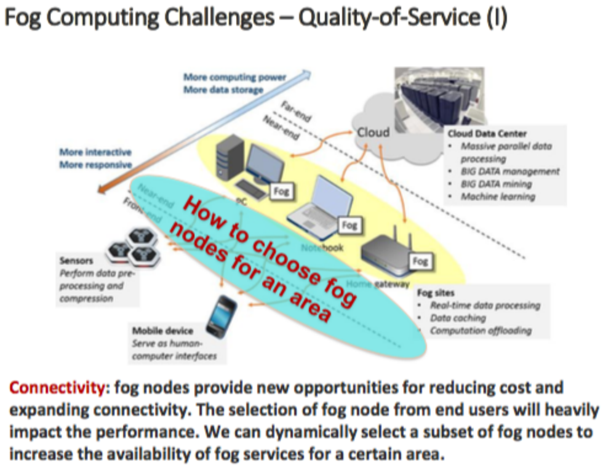
Parameters	Cloud Computing	Fog Computing
Server nodes location	Within the Internet	At the edge of the local network
Client and server distance	Multiple hops	Single/multiple hop
Latency	High	Low
Delay Jitter	High	Low
Security	Non-locally controllable	Locally controllable
Location awareness	No	Yes
Vulnerability	Higher probability	Lower probability
Geographical distribution	Centralized	Dense and Distributed
Number of server nodes	Few	Very large
Real time interactions	Not fully supported	Supported
Usual last mile connectivity	Leased line /wireless	Mainly wireless
Mobility	Limited support	Supported

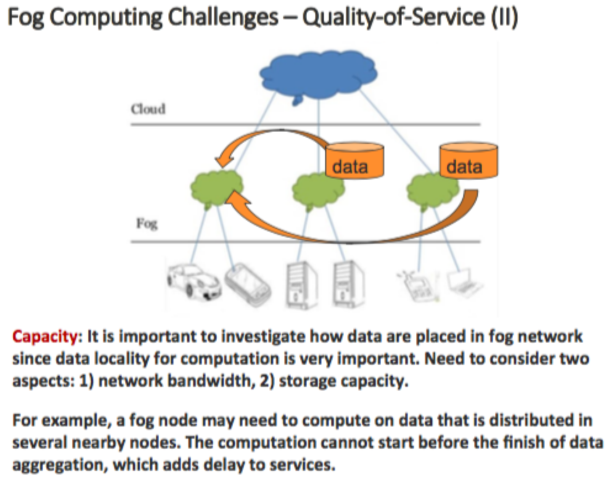
Fog Nodes: two categories

Suppose, in Oslo, bus stops are equipped with fog nodes that comprise a server with processing capabilities.

First category: devices are only data producers. Each bus stop can control a city area. For *example*: sensors measure the CO2 level and forward the data to the fog node which will process all data collected from all sensors in its coverage area

Second category: devices can both sense and compute. Fog node processes sensor's data and aggregates IT capacities of devices. For *example*: sensor connected to a compute board, a mobile phone or a car. The data processing is possible by the device itself, hence producing information to the fog node.

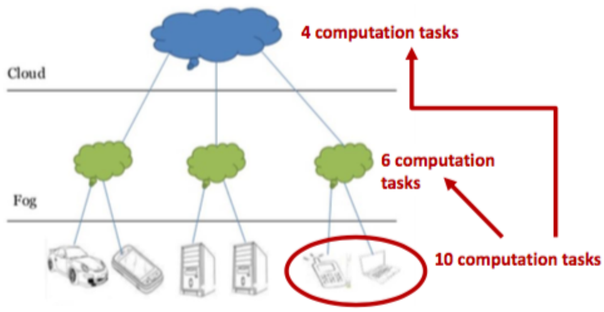




Computation Offloading

Computation offloading: heavy computation tasks can be done in fog or cloud instead of mobile devices. This saves storage and battery lifetime.

The main challenges in offloading in fog computing are how to deal with dynamic. In three-layered architecture: device-fog-cloud. A typical question: how to partition tasks to offload on fog and cloud.



Fog Computing Applications (I) Fog/Edge computing is able to deal with new applications that suffer from the limitations and poor scalability of the centralized cloud paradigm. The applications can be:

Applications susceptible to latency: Typical examples are manufacturing, urban transport and smart energy, where actuation has to be driven in real-time.

Geo-distributed applications: Typical examples include environmental monitoring, which are based on the collection and processing of streams from thousands or even millions of sensors.

Mobile applications: typical applications involving fast moving objects, e.g., autonomous cars. They require moving objects to access local resources (computing, storage) residing at their vicinity.

Distributed multi-user applications with privacy implications and need for fine-grained privacy control. These applications can benefit from a decentralization of the storage and management of private data to the various edge servers, thus alleviating the risk of transferring, aggregating and processing all private datasets at the centralized cloud.

Fog Computing in Industry IT industry leaders (e.g., ARM, Cisco, Dell, Intel, Microsoft, and the Princeton University Edge Laboratory) formed a coalition and have joined forces to create the OpenFog Consortium
Goal: acceleration of Fog computing technology deployment for the good of the industry https://www.openfogconsortium.org/

Mobile Edge Computing (MEC) is developed by ETSI Specification. MEC pushes the cloud computing capabilities close to the Radio Access Networks in 4G. MEC server can be a mini-data center

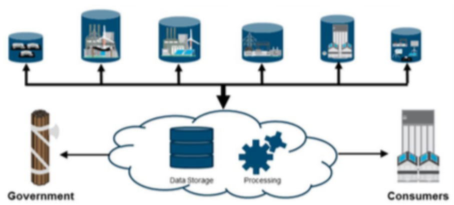
located in either the base station

storage close to the radio access networks

7.3 Cloud/Fog Computing for Smart Grid

Tremendous Smart Meters Data in Smart Grid The amount of data in smart grid may be tremendous since: i) there are a huge number of smart meters/sensors; ii) each meter shall transmit data periodically to the control center.

22million smart meters in Zhejiang Province, China, as required by China State Grid.
A smart meter should be 96 measures/day, which leads to 2.1billion records per day.
Cloud computing is important for smart grid



Dynamic: cloud platform opens up analytics and data to wide groups of users, inside and potentially outside of the utilities, on a dynamic basis

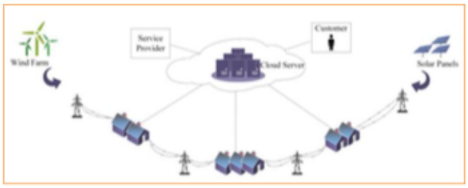
Reliable: access to cloud data streams and ease of data duplication is allowing utilities to provide wide set of reliable data across the enterprise

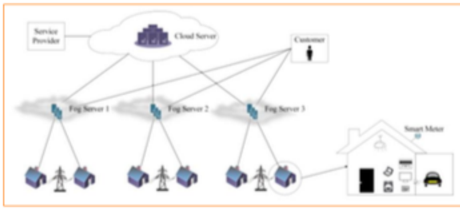
Deregulated: value of data access varies widely in a deregulated territories where utilities are requested to provide usage data to customers

Cloud/Fog Architecture for Smart Grid

Cloud computing: all smart meter data is stored and processed in the cloud.

Cloud/Fog Architecture: Various communication and computational needs a very low latency and improved privacy, which can easily be addressed by fog computing.

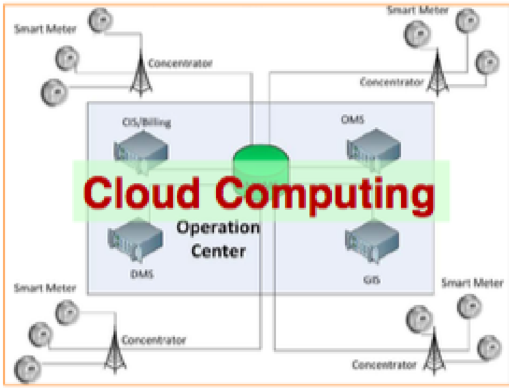




Central Cloud Computing for Data Management

MDMS (Meter Data Management Systems) uses cloud computing to perform long-term data storage and management for the vast quantities of usage data and events. MDMS enables the interaction with operation and management systems that: (i) manages the billing and customer information, (ii) provides power quality report and load forecasting based on meter data

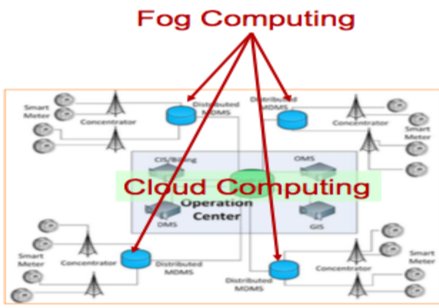
Q: all data from the smart meters need to go through the centralized cloud, it makes the system non-scalable. Why?



One approach: Cloud/fog computing for data management

Several distributed MDMSs using fog computing are close to smart meters. Each MDMS is responsible for the specific area.

Communication distance for data collection is largely reduced, and the corresponding resources needed are also less.



Fog Computing for Smart House

In smart house, the home of the future might be equipped with a variety of power use meters and monitoring devices, adapting behavior to match cost of power, load on the grid, and activities of the residents.

A fog node/server at home can collect, process, analyze, aggregate and transmit a large amount of data. The fog server can also manage the smart house for decision making, e.g., demand response, temperature control.

Fog Computing for Wind Farm

A large wind farm may consist of hundreds of individual wind turbines, and cover an area of hundreds of square miles. Modern wind turbines are very large control structures aimed at improving wind power capture and power quality.

Wide geographical deployment of a large system consists of a number of autonomous yet coordinated turbines. A turbine is supposed to quickly respond to external weather and environment, which gives rise to the need of a fog computing platform.

Wind Farm System Needs Fog Infrastructure

Cloud infrastructure:

Global controller: gathering data, building the global state, determining the policy.

Network Infrastructure: An efficient communication network between sub- systems, system and the internet (cloud)

Fog infrastructure:

The continuous supervisory role of the global controller requires low latency locally in the edge type deployment

Data analytics: This system generates huge amounts of data. Much of this information is actionable in real time. The data can be also used to run analytics over longer periods (months, years). The cloud-fog infrastructure is the natural approach to run to support both long-term and short-term tasks.

7.4 More Considerations

Self-driving Car: use cloud computing or fog computing?

A self-driving car is a complex robotic system equipped with sensors, actuators and ICT capabilities.

Requirement: approximately 1 GB of data will need to be processed each second. This data will need to be analyzed quickly enough so that the vehicle can react to changes in its surroundings in less than a second.

Q: shall we use cloud computing or fog computing for self-driving car?

The requirement mean a self-driving vehicle would require a lot of computing power to minimize application latencies, as well as very low network latencies.

Cloud Computing Future: Anything as a Service (XaaS)

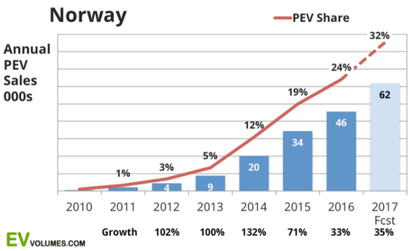
Recent advances: Drones as a Service, Robot as a Service, Blockchain as a Service

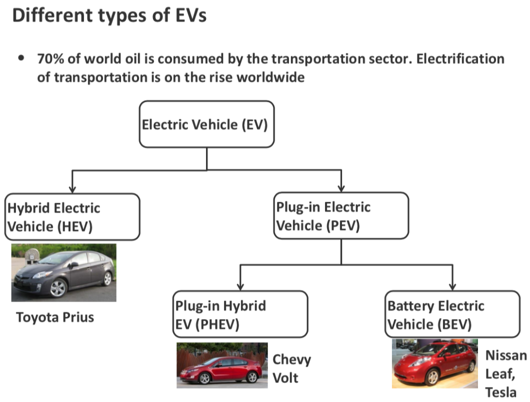
Drones as a *Service*: For example: Amazon delivers packages to your home with their drone delivery system

Drones are used as data mule to deliver data from wind farms to edge server or control center

8 TESLA as Internet of Things

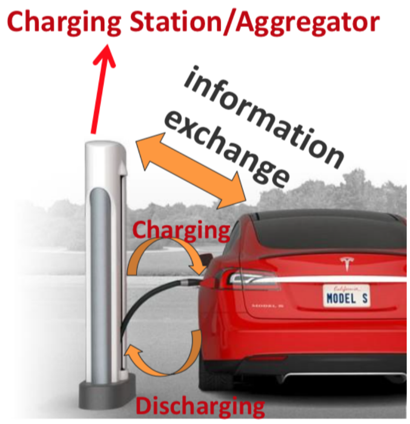
EV facts in Norway
200000 EVs in December 2017; the country with the largest electric cars per capita in the world. Ambition: 400000 EVs by 2020





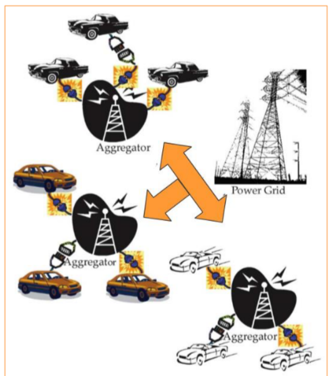
Electric Vehicles (EVs)

- Charging stations/Aggregator
  - Energy domain: power charging and power discharging
  - Information domain: information exchange between EVs and charging station
- Electric vehicles roles
  - Energy consumers
  - Energy storage
  - Energy provider



Vehicle-to-Grid (V2G) Systems: information and energy domains

- Three main components in this system: power grid, several aggregators, and EVs. Each aggregator serves as an interface between the grid and a group of EVs.
- *Information*:The communications among EVs, aggregators, and grid can go through a two-way digital communications (wired or wireless) infrastructure.
- *Energy*: EVs in the system charge or discharge their batteries via an aggregator.



Keywords related to TESLA

- Energy storage
- Battery management
- Energy
- Electronics
- Machine learning
- Computation
- Communications
- Smart devices
- Transport

Then, is TESLA still a car?

Yes, it is a car! But it is more than just a car. It is a super computer, smart device and a mobile robot.

Important: TESLA is NOT a car. TESLA is an IT device that has the car function!

iPhone:

- iPhone is not just phone. It is a smart device that we can make a call.
- Making a call is just one function of the device.

TESLA

- Similarly, TESLA is not just a car. It is a smart device that we can drive. Driving is just one function of the device.
- TESLA is smarter than iPhone, it can even drive by itself.
- TESLA is a learning car by itself.
- **Q**: what more functions can you expect?

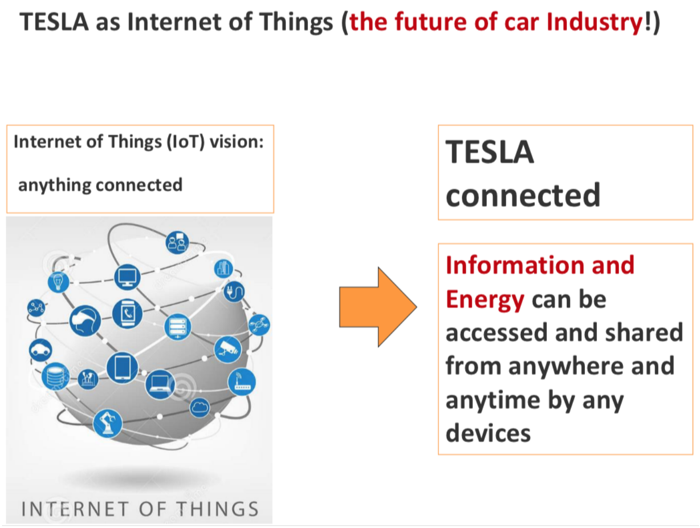
8.0.1 Internet of Things (IoT) Concept

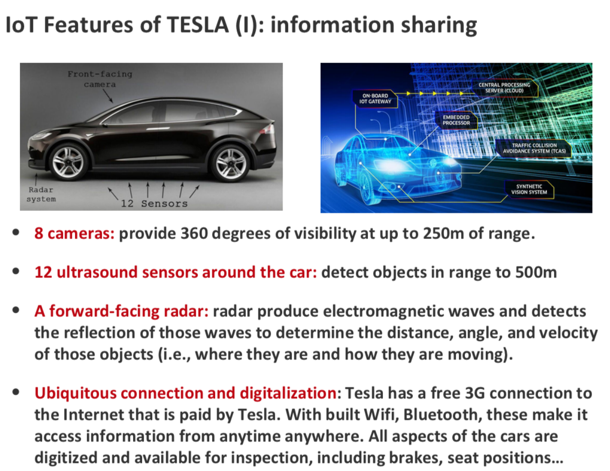
Concept: a network of connected objects that collect and exchange data at anytime anyplace. This includes everything from cellphones, coffee makers, washing machines, wearable devices and almost anything else you can think of. This also applies to components of machines, e.g., a jet engine of an airplane. Gartner forecasted that by 2020 there will be over 26 billion connected devices.

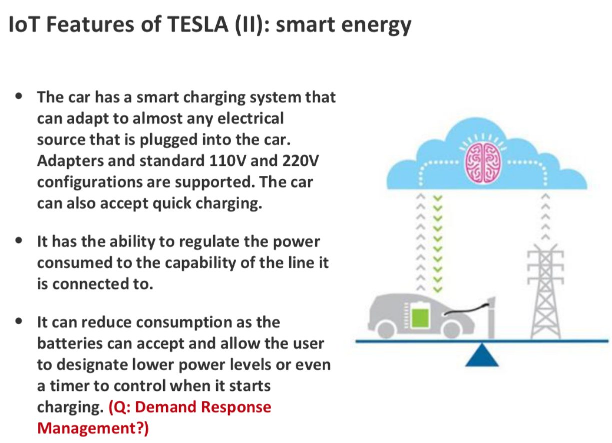
IoT is a giant network of connected "things“ which also includes people.

IoT-enabled cars is the future of auto industry

- Now,thefunctionalityofIoT-enabledcarsis transforming the auto industry:
  - theultra-connectedTesla;Google'sself- driving cars; Uber's hope to send you to your destination in an autonomous vehicle.
- Cars will be cloud-based, artificially intelligent supercomputers.
  - Audi and Mercedes are also looking into similar technology.
  - Companies who fail to recognize this as the future of car, they have a high chance of fading out of existence. This happened before like Blackberry and Nokia







8.1 Information Domain

8.1.1 Intelligent Transport Systems (ITS)

ETSI (European Telecommunications Standards Institute) definition

Intelligent Transport Systems (ITS) include telematics and all types of communications in vehicles, between vehicles (e.g. *vehicle-to-vehicle*), and between vehicles and fixed locations (e.g. *vehicle-to-infrastructure*). However, ITS are not restricted to Road Transport - they also include the use of information and communication technologies (ICT) for rail, water and air transport, including navigation systems.

Automotive systems,Railway systems,Maritime systems,Air transport systems

V2V & V2I for connected vehicles

- *Vehicle-to-Vehicle(V2V) Communications*: communications between vehicles
- *Vehicle-to-Infrastructure(V2I) communications*: communications between vehicles and road infrastructures
- Cars, trucks, buses, and other vehicles can "talk" to each other. They continuously share important safety and mobility information. Connected vehicles can use wireless communications to "talk" to traffic signals, work zones, toll booths, school zones, and other infrastructures
- *Safety applications*: safety packets transmission, emergency alert, driver assistance
- **Non-safety application**: advertisements, map update, entertainment

V2V Communications

- V2V communications systems are composed of devices, installed in vehicles, that use dedicated short-range radio communication (DSRC) to exchange messages
- The messages contain vehicle information (e.g., vehicle's speed, heading, braking status).
- V2V devices use this information from other vehicles and determine if a warning to the vehicle's driver is needed, which could prevent a crash.

8.1.2 Dedicated Short-range Radio Communication (DSRC)

- DSRC-based devices can be installed directly in vehicles, or could be carried into vehicles by drivers in the form of a handheld device (e.g., smartphone).
- In general, safety-related applications are expected to rely on DSRC. DSRC uses spectrum dedicated by the FCC for transportation safety applications.

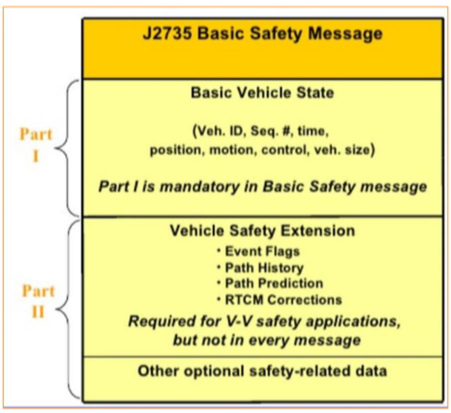
- Non-safety-related applications may use other forms of wireless communications, such as cellular or Wi-Fi.
- **Goal**: A short to medium range communications service that supports both public safety and private operations in V2V and V2I communications modes.
- **Frequency**: 5.9GHz; Bandwidth: 75MHz. This spectrum has been exclusively allocated for vehicle safety and mobility applications.
- Key Benefits
  - Low latency communication (« 50ms). Safety has stringent communications requirements, and future pre-crash and automation requirements may be even more stringent.
  - High data transfer rates (3 – 27 Mbps)
  - Line-of-sight, up to 1500 m and 360^o
- Benefits of Safety Applications:
  - Collision avoidance; Improved mobility; Improved environmental protection

V2V Working Principle

- Each vehicle broadcasts its state information in a "Basic Safety Message" (BSM) up to 10 times per second to surrounding vehicles.
- BSM is sent in 360 degree pattern using DSRC technology. Upon receipt of BSM, vehicle builds model of each neighbor's trajectory, assesses threat and warns driver (or takes control) if threat becomes acute.
- BSM is received by the other vehicles equipped with DSRC devices and processed to determine collision threats.

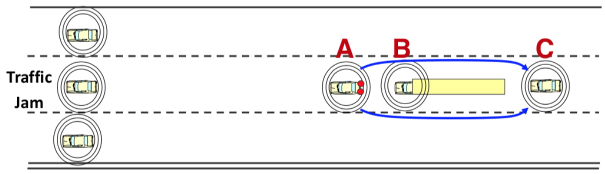
8.1.3 Basic Safety Message (BSM)

- Average message size: 320 ˜ 350 bytes
- Default transmit rate: 10time/second
- BSM is exchanged between vehicles and contains vehicle dynamics information such as heading, speed, and location.
- BSM is transmitted only through the control channel and is tailored for low latency, localized broadcast required by V2V safety applications

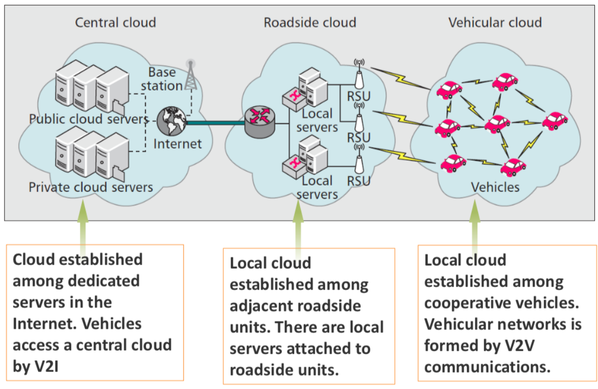


V2V Safety Applications – Emergency Electronic Brake Lights

- Emergency Electronic Brake Lights: approaching a vehicle stopped in roadway, but not visible due to obstructions
- Car A decelerates when it is approaching jam. Car B will inform Car C via DSRC within 100 msec.

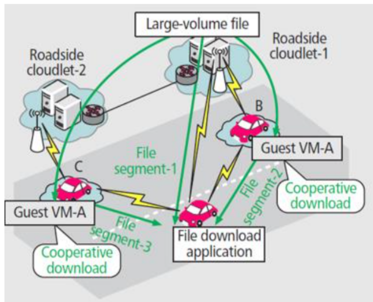


Cloud computing and fog computing are the crucial technologies for intelligent transport



Cloud computing for transport

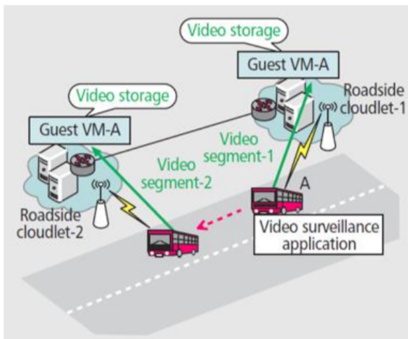
- When car A wants to download a large file, the cars nearby (B and C) can help car A to download segments of the file.
- After B and C get the segments of the file, the two car can forward the data to car A
- **Q**: any challenges for cooperative download?



Cooperative download of a large file

Fog computing for transport: storage resource sharing

- In Oslo,every bus has a camera which records and saves data.
- **A challenge**: the storage of the camera is limited.
- **Solution**: distributed storage in video surveillance. When the bus comes to a bus stop, the bus will transmit the video data to the local cloud (i.e., fog nodes) at the bus stop. Then, all video data will be stored in different bus stops in a distributed manner.



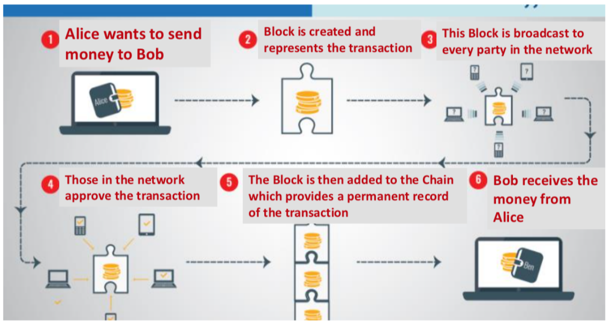
Distributed storage in video surveillance

8.2 Energy Domain

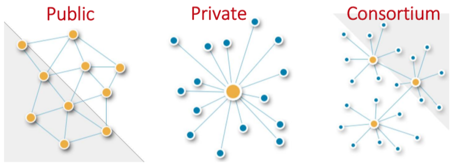
Blockchain and Bitcoin

- Blockchain: a distributed database where anyone can create and complete transactions stored permanently. There are no central nodes and all nodes are equal. Each node stores the whole database.
- Bitcoin: unregulated digital currency designed to bypass currency controls and simplify online transactions by getting rid of third-party payment processing intermediaries.

- Blockchain & Bitcoin relationship: Bitcoin was an application of Blockchain. Blockchain has applications far beyond Bitcoin.



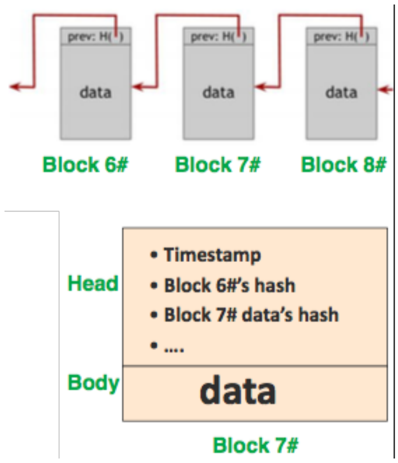
Blockchain types



	Public	Private	Consortium
Administration	No administrator	Single	Multiple
Access	Open read/write	Permission	Permission
Confidence of participation	Low	High	Medium
Computation cost	High	Low	medium

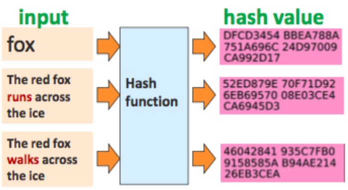
Blockchain data structure

- **Blockchain data structure**: a linked list with hash pointers used to record all transactions. New blocks are added to the end of the chain.
- **Hash pointer**: gives you a way to retrieve data along with the hash of the data. A regular pointer only gives you a way to retrieve data.
- **Block**: blockchain is composed of block. Block refers to a group of transactions at a specific time and hash pointer of the previous block. Each block includes: head and body (i.e., data).
- A simple blockchain in Python: https://github.com/EricAlcaide/pysimplechain



Hash function and tamper-proof mechanism (I)

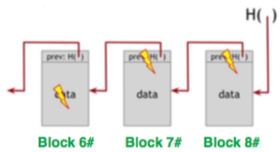
- **Hash function:** takes any size input text (e.g., transactions) and returns a fixed size string (i.e., hash value).



- The ideal hash function has three main properties:
  - Easy to calculate a hash for any given data
  - Hard to calculate the original text that has a given hash
  - Two slightly different messages produce drastically different hash value
- Bitcoin uses a standard SHA-256 hash algorithm which generates a 256 bit hash value. For example: SHA256(123) = a665a45920422f9d417e4867efdc4fb8a04a1f3fff1fa07e998e86f7f7a27ae3

Hash function and tamper-proof mechanism (II)

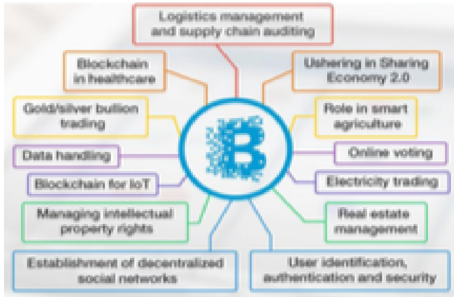
- **Tamper-proof**: an adversary is not able to tamper data in any block without getting detected.



Example:

- An attacker wants to tamper the data in Block 6. Any change in contents of Block 6 will make the hash of this block saved in Block 7 invalid. The attacker has to modify Block 7 which in turn changes the content of Block 7.
- The adversary has to tamper all blocks until the head of the blockchain, which is called the genesis block. Then, the hash of this block doesn't match the hash we hold. The adversary can't tamper with the hash of the genesis block because it is kept in a secure location where no one can alter it.

Blockchain applications in general



- Blockchain: decentralized database that keeps record of all transactions.
- This provides a perfect way for systems to record transactions that should be transparent and permanent.

Blockchain applications: distributed cloud storage

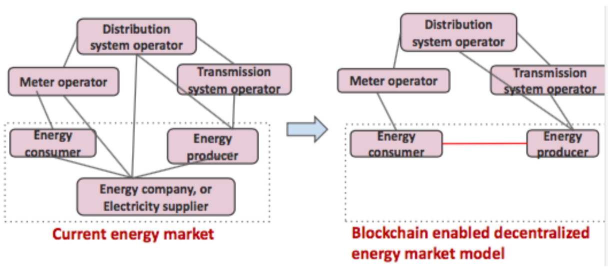
- Current cloud storage services are centralized. The users must place trust in a single storage provider, e.g. dropbox.
- With blockchain enabled distributed cloud storage, such services can be decentralized. Anyone (e.g., individuals, small data storage, large-scale data center), on the Inter
- **Advantages**: *security and reduced dependence on a single node.* Hashing and having the data in multiple locations are the keys to securing it by reducing dependencies on a single node.

Blockchain applications: energy sharing where everyone can contribute/share power

- An electric vehicle may charge or discharge
- Each vehicle can buy and sell its own electricity with other vehicles.
- Each house can generate, store, buy and sell its own electricity with neighbors
- **Advantages**: no need a third-party utility participating energy exchange among houses or electric vehicles $\rightarrow$ low cost, flexible, new business models...
- **Challenge**: new techniques to organize, secure and trace the local energy transaction record $\rightarrow$ blockchain is the key!

Blockchain will transform energy market

- Energy companies or suppliers will not be used in the new energy market.
- Energy consumers and energy producers trade energy directly with each other without a third party. The saved cost serves as the payment to electricity suppliers.

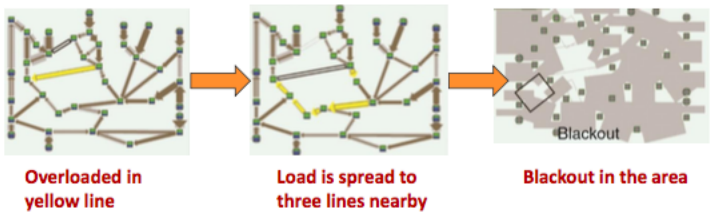


9 Smart Grid Privacy and Security

9.1 Smart Grid Cyber-Security

Security - two concepts

- Physical Security: this is the traditional "security" concept of power systems (cascading failure)



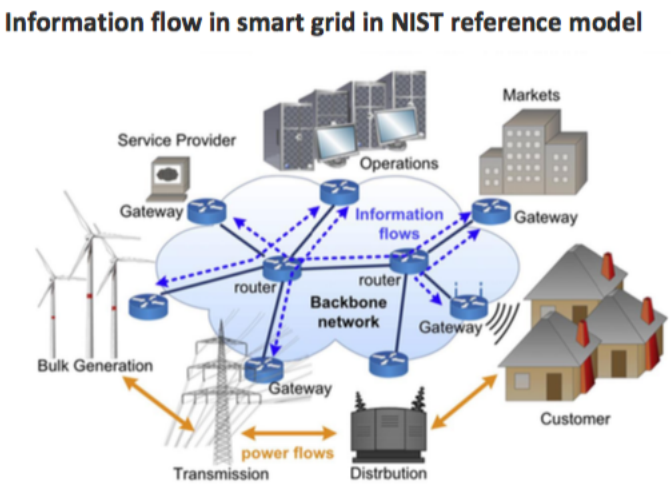
- Cyber security: technologies, processes and practices designed to protect networks, computers, programs and data from attack, damage or unauthorized access → Our focus in this lecture.

Real-World Cyber Attacks in Smart Grid

- Massive cyber attack occurred on Ukraine's power system in December 2015. More than ten thousand homes and facilities experienced a power outage for hours, even days. This attack was enabled by a malware called BlackEnergy installed on the control center computers.
- January 2003, computers infected by the Slammer worm shut down safety display systems at power plants in Ohio
- In 2008, there was evidence of computer intrusions into some European power utilities
- In 2010, Stuxnet worm provides an aggressive attack on control systems

In Norway

- Norway has still not seen any serious cyber attacks that have taken out critical societal functions, as was witnessed in Ukraine during 2015
- But, "Around 80 per cent of all data traffic in Norway passes through Telenor. Every year, our security team deals with and averts thousands of cyber attacks and countless attempts at fraud, but it is impossible to stop them all," says Hanne Tangen Nilsen (Telenor Norway's Chief Security Officer)
- As an owner and operator of civic infrastructure, Telenor Norway has great responsibility. Hospitals, power companies and banks are highly dependent on Telenor infrastructure.

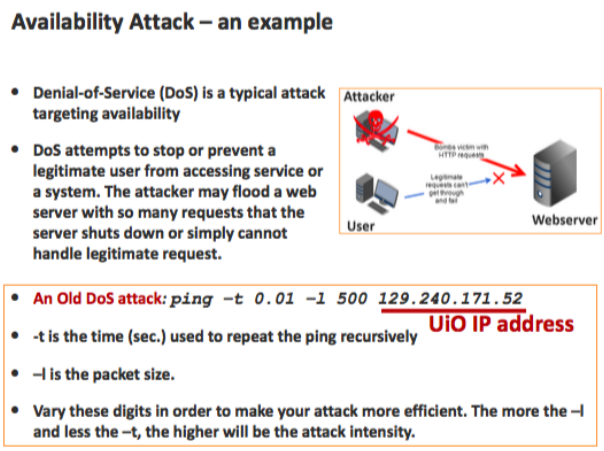


Three Cyber Security Objectives in Smart Grid – NIST guideline

- The cyber security working group in NIST released a comprehensive guideline for Smart Grid cyber security
- CIA (*NOT Central Intelligence Agency*): Confidentiality, Integrity, and Availability
- Loss of ***availability*** — disruption of access to or use of information or a system
- Loss of ***integrity*** — unauthorized modification of information
- Loss of ***confidentiality*** — unauthorized disclosure of information

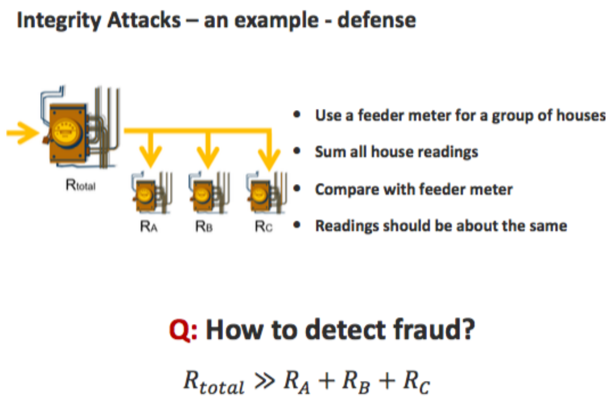
9.1.1 Availability

- **Availability**: Ensuring timely and reliable access to and use of information is of the most importance in the Smart Grid. This is because a loss of availability is the disruption of access to or use of information, which may further undermine the power delivery.
- Availability of critical information is very important
  - Availability of price information: critical due to serious financial and possibly legal implications. Moreover, outdated price information can adversely affect demand.
  - *Availability of commands*: important for turning a meter back on after completing the payment of an electric bill.



9.1.2 Integrity

- Integrity refers to preventing undetected modification of information by unauthorized persons or systems.
- **Target**: either customer's information (e.g.,pricing information)or network operation information (e.g., voltage readings, control commands). Such attacks attempt to deliberately modify the original information in order to corrupt critical data exchange in the smart grid.
- **Consequence**: A loss of integrity is the unauthorized modification or destruction of information and can further induce incorrect decision regarding power management. Violation of integrity may cause safety issues, that is, equipment or people may be harmed.

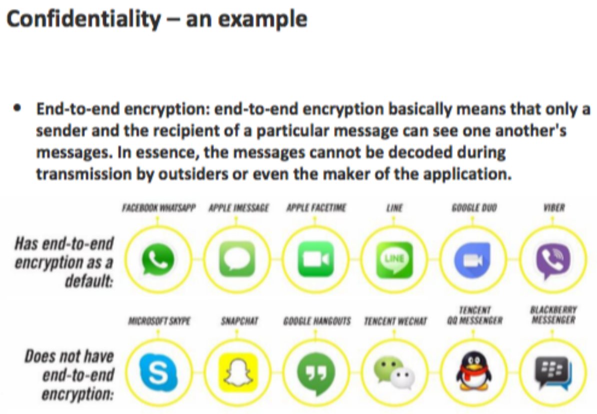


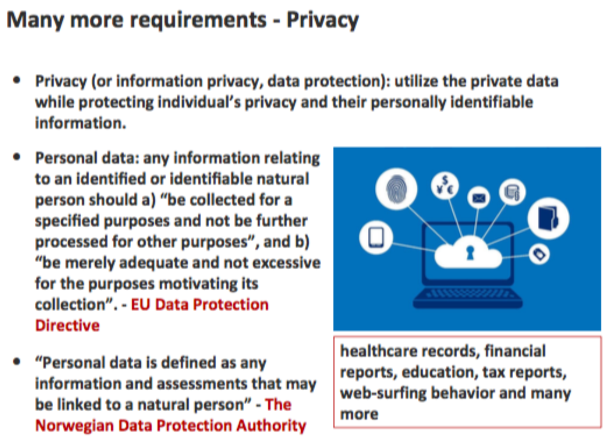
9.1.3 Confidentiality

- **Confidentiality**: set of rules that limit access or place restrictions on disclosure of some information, e.g., by means of encryption. Confidentiality ensures that access to information is restricted to authorized entities, e.g., bank account statements, personal information, credit card numbers, trade secrets, government documents.

- Encryption is a fundamental approach for information confidentiality. Encryption ensures that only the right people can read the information. A typical example is HTTPS (HTTP over SSL/TLS), a security protocol for communications over the Internet.

- *Access control*: another way to ensure information confidentiality to restrict access to sensitive information.





9.1.4 Authentication

- Authentication is concerned with determination of the true identity of a communication system participator and mapping of this identity to a system-internal principal (e.g., valid user account) by which this user is known to the system.

- Most security objectives, most notably authorization, distinguish between legitimate and illegitimate users based on authentication.

- Authentication approaches: username-password; biometric authentication (e.g., fingerprint, face)

Mutual Authentication

- Mutual authentication: network authenticates mobile phones; and at the same time mobile phones will authenticate networks

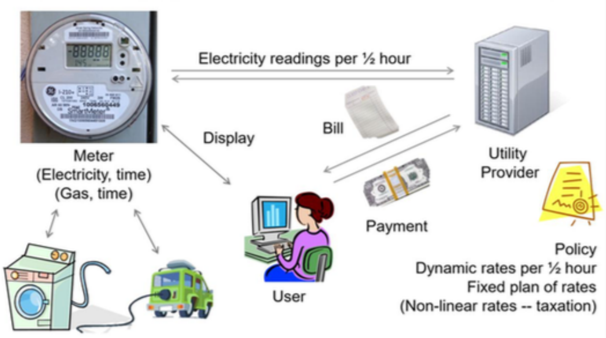
- Mutual authentication provides enhanced protection against false base station attacks by allowing the mobile phone to authenticate the networks

- **Q**: are we using mutual authentication in 4G?

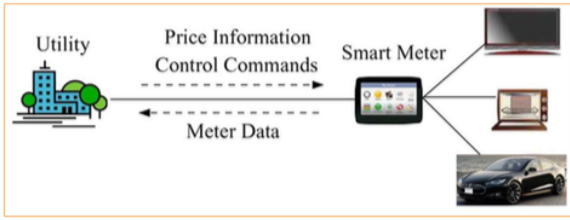
- **Q**: shall we perform mutual authentication when an Electric Vehicle is charging battery?

9.2 Smart Grid Privacy

Privacy is very related to smart metering



Important information exchange through smart meters

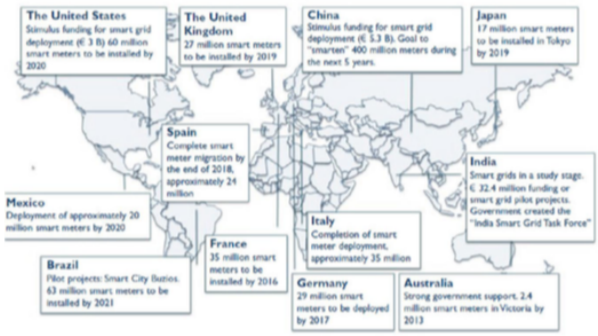


- **Information flows to/from a smart meter including**
  - **price information**
  - **control commands**
  - **meter data**

Smart Meters Deployment Plan in Europe

- Norway: decided that all customers in Norway will receive new smart meters by 1 January 2019.
- European Union: The European Parliament proposed that 80% of electricity customers have smart meters by 2020.
- England: The Government has announced plans to install a smart meter in every home in England by 2020. **However**
- Netherlands: April 2009, First Chamber declined a bill which would require the installation of smart meters in residences. Chamber cited privacy issues as the main concern measure
- Germany: decided against a nation-wide deployment of smart meters due to privacy and cost issues in 2016

Smart Meter Projects around the World



Privacy Concerns – different applications

- At your home
  - Appliances and devices may communicate with one another to use electricity more efficiently.
  - Consumers or others may remotely control temperature and other appliances.
- Remote Connect / Disconnect of Meter
- Demand Response and Pricing
  - Customers will respond to real-time pricing signals.
- Marketing Use of Data
- 3rd Party Application Developers
  - Device-makers may seek access to data to provide in-home tools for better monitoring of electricity use.

Privacy Concerns at End-users

- **Smart Meters**: Information regarding energy use will be collected from the home and may be able to relay information about specific types of appliances being used at specific moments in time.
- **Fine-grained meter reading**: one essential technology of smart grids is fine- grained meter reading within a very short period (15min or fewer) per household. However, meter readings of a household reveal detailed information about daily activities of the household and used appliances during a specific time period. The smart metering data is normally saved in centralized data center. Fine-grained meter readings may cause serious privacy issues.
- **Load signatures**: could potentially indicate when you are home and whether you are cooking, watching television, or using other electronic devices.

What Could Power Usage Information Reveal?

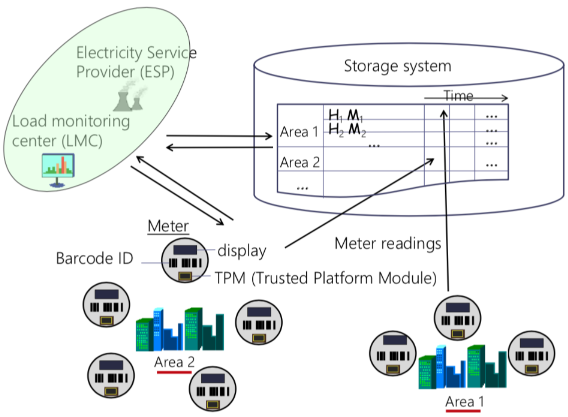
Questions	Pattern	Purpose
Are you at home after you claim sick to your boss?	Yes: Power activities during the day No: Low power usage during the day	?
Did you get a good night's sleep?	Yes: No power events overnight for at least 6 hours No: Random power events overnight	Health services
Are you on vacation?	Yes: No power events for some days No: random power events in a day	Useful information for malicious
Do you have children? Did you leave your child home alone?	Yes: Single person activity pattern No: Simultaneous power events in distinct areas of the house	Safety services
How many people live there? When you have a party!	Need data for some time and analyze High power usage one day	Your landlord have great interest to know

Potential Risks

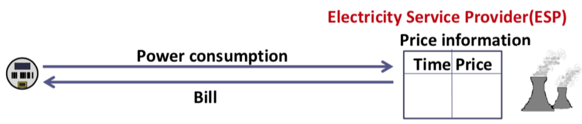
Who wants smart meter data?	How could the data be used?
Utilities	To monitor electricity usage and load; to determine bills
Advisory companies	To promote energy conservation and awareness
Insurance companies	To determine premiums based on unusual behaviors that might indicate illness
Marketers	To profile customers for targeted advertisements
Law enforcers	To identify suspicious or illegal activity
Civil litigators	To identify property boundaries and activities on premises
Landlords	To verify lease compliance
Private investigators	To monitor specific events
The press	To get information about famous people
Creditors	To determine behavior that might indicate creditworthiness
Criminals	To identify best times for a burglary, or valuable appliances to steal

- Q: do we really need minute-scale energy monitoring?

Bill Application and Load Monitoring Application



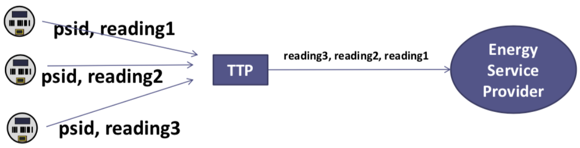
Automatic Bill – we do not need minute granularity of meter readings



- Automatic Bill: In billing applications, the electricity service provider only needs the amount of power consumption per hour to compute a bill.

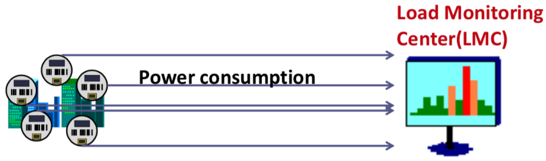
Privacy-preserving Automatic Bill

- Trusted third party (TTP) computes the bill, e.g., the grid operator
- Anonymization – Instead of a user's real ID, a pseudonym (a fake ID, denoted by psid in the illustration below) is attached to the data. Thus, the receiver do not know to which individual the received data belong. Daily (or hourly) power consumption remains anonymous, user identity is only used at the end of each billing period when presenting all consumed credentials together.
- Q: any disadvantages of this scheme?



- May have large volume of data; Complex accountability service

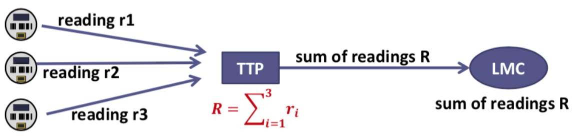
Load Monitoring - we do not need minute granularity of meter readings



- Load Monitoring: the load monitoring center (LMC) collects the amount of electricity usage over a local area in order to monitor current activities of the power grid.
- LMC only needs the total power consumption over the area at recent time units.

Privacy-preserving Load Monitoring

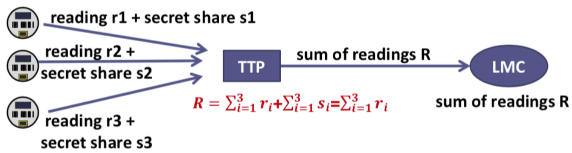
- Aggregation: The data from individual users are aggregated and only the aggregated data are sent to the receiver.
- Privacy-friendly aggregation: the aim is for a utility to reveal the sum of readings from multiple meters without learning the readings themselves.
- Q: any disadvantages of using this scheme?



- May not be scalable when there are too many users in an area
- Need a Trusted Third Party (TTP) to aggregate data

Privacy-preserving Load Monitoring – secret shares

- Secret shares of 0 among meters
- Meters generate readings that are added by shares (s1, s2, s3) summing to zero (i.e. s1+s2+s3 = 0). When the readings are summed, shares cancel out and the output of LMC is the sum of metering reading
- Q: any disadvantages of using this scheme?



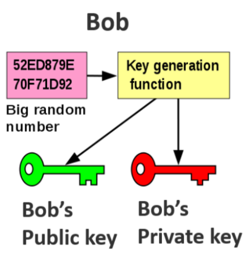
- These protocols are very efficient but suffer from inflexibility: the groups of meters that can be aggregated are static, and missing readings prevent the computation of the aggregate. Need a trusted third party (TTP) to aggregate data (Q: what if TTP is not trusted?)

Untrusted TTP: We Need Public Key and Private Key

- Bob has two keys: public key and private key (basically long random numbers). An example of a Public Key:

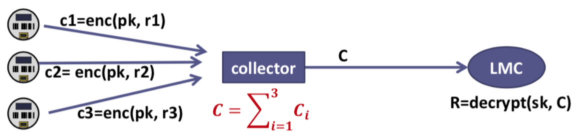
3048 0241 00C9 18FA CF8D EB2D EFD5 FD37 89B9 E069 EA97 FC20 5E35 F577 EE31 C4FB C6E4 4811 7D86 BC8F BAFA 362F 922B F01B 2F40 C744 2654 C0DD 2881 D673 CA2B 4003 C266 E2CD CB02 0301 0001

- Public Key: available to everyone via a publicly accessible directory
- Private Key: confidential to its owner, i.e., Bob
- The key pair is mathematically related, a message encrypted with a Bob's public key cannot be decrypted by anyone, except Bob possessing with his private key.



Aggregation with collector which may not be trusted

- Load monitor center (LMC) has two keys: public key and private key. Each smart meter encrypts (enc(.) in the illustration) its data with the public key (pk in the illustration) and sends the ciphertext to the collector.
- The collector aggregates the received ciphertexts and sends the aggregated ciphertext to load monitor ccenter.
- The load monitor center decrypts the aggregated ciphertext with its private key (sk in the illustration) to obtain the aggregated data of all smart meter readings.



9.3 Smart Grid Security

We focus on four attacks

- **Replay Attack**
- **Man-in-the-Middle Attack**
- **Integrity Attack**
- **False Data Injection attack**
- Virus
- Backdoor
- Logic bomb
- Trojans

9.3.1 Replay Attack

- Replay attack is based on interception of smart grid's usage pattern. The replay attacks can easily be realized since the utility company cannot manage every smart meter distributed in a very large area.
- The attacker can intercept the smart meters to observe the reported data for a certain amount of time.
- The attacker replays (or retransmits) the data to smart grid control center.
- Attacker can make customers' smart meters out of order by injecting incorrect data to the system, which may lead to incorrect energy price or inaccurate prediction.
- In order to launch such an attack the adversary needs to
  - capture and analyze the data transmitted between appliances and smart meters to gain the customer's characteristics of power usage
  - inject false control signals into the system; and lead to power system instability
- The aim of the replay attack is:
  - to steal energy by charging idle equipment's status to busy state in order to reroute the power to another place
  - cause physical damage to the system

Replay Attack – an example

Stuxnet is 500KB computer worm and allegedly designed in a joint effort by USA and Israel as a cyberweapon against Iranian nuclear facilities. Stuxnet targets power supplies used to control the speed of a motor in nuclear centrifuges. The malware intercepts commands sent to the power supply from the Siemens industry control software, and replaces them with malicious commands to control the speed of a motor, varying it wildly.

Defending Replay Attack - adding timestamp and sequence number

Put a time stamp in each message to ensure that the message is "fresh"

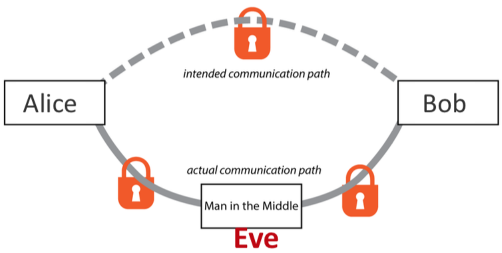
- Do not accept a message that is to old
- place a sequence number in each message
- Do not accept a duplicated message

Defending Replay Attack - adding random noise

Add random noise in request-response

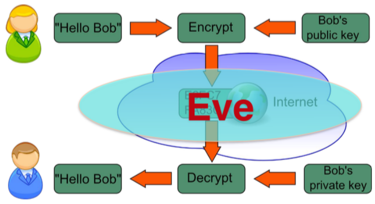
- Sender of request generates a nonce (random number)
- Places the nonce in the request
- Server places the nonce in the response
- Neither party accepts duplicate nonce

9.3.2 Man-in-the-Middle Attack



- *Alice* thinks her communication with *Bob* is secure, but actually a man-in- the-middle Eve is eavesdropping on the conversation. *Eve* impersonates two communication nodes and makes them believe that they are talking together.
- Alice and Bob think their communications are end-to-end encrypted when actually there is an eavesdropper Eve who is removing the encryption, examining (or modifying) the unencrypted content then re-encrypting it before passing it to.

Man-in-the-Middle Attack – technically for Public key Encryption

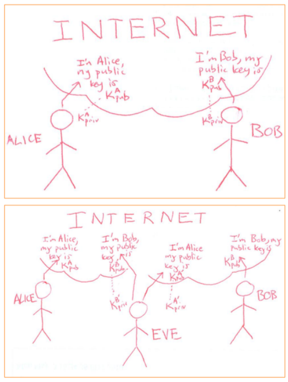


Q: what may happen if there is Eve between?

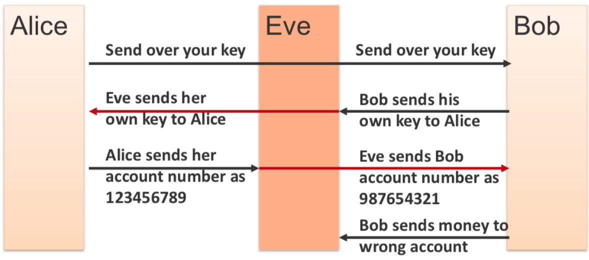
- Bob's public key is on the Internet. Bob "owns" his public key because he alone has the matching private key. Alice sends Bob the plaintext message by:
  - *"Hello Bob!"* — encrypt with Bob's public key → *ciphertext*
  - Alice sends Bob *ciphertext*. Bob then use his own private key to decrypt *ciphertext*.

Man in the middle attack

- A malicious user Eve intercepts a public key on its way to one of the parties involved
- Alice thinks she's encrypting her message with Bob's public key, but the public key she uses really belongs to Eve.
- Eve then decrypts and reads the message, perhaps modifying it a bit, and then encrypts it using the real Bob's public key. The real Bob receives the message which may have been modified.

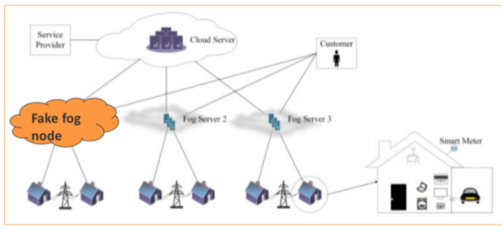


Man-in-the-Middle Attack may result in financial loss



- Eve impersonates both sides of the conversation to gain access to funds. Eve intercepts a public key and can transpose her own credentials to trick Alice/Bob into believing they are talking to one another securely.
- Very common in wireless communications. One typical solution is to enforce a secure wireless authentication protocol, e.g., WPA2 (Wi-FI Protected Access 2)

Man-in-the-Middle Attack in Smart Grid with Fog Computing



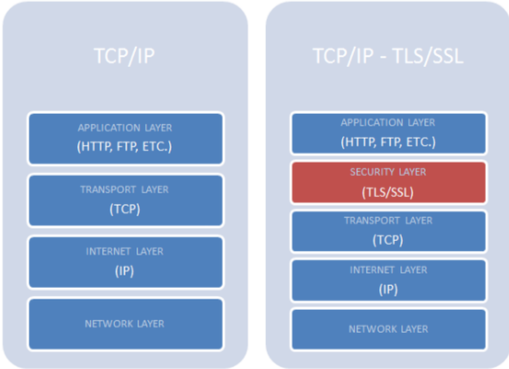
- Gateways serving as fog nodes may be compromised or replaced by fake ones. Private communications of victims (individual users, block of houses) will be hijacked once the attackers take the control of gateways.
- This attack happens when the adversary intercepts network data (e.g., switch states) and meter data, fabricates part of these, and forwards the modified version to the control center.

Man-in-the-Middle: Consequences

- The adversary can mislead the control center that the grid is operating under a topology different from that in reality.
- Such an attack, if launched successfully and undetected by the control center, will have serious implications:
  - A grid that is under stress may appear to be normal to the operator. This delays the deployment of necessary measures to ensure stability.
  - Grid operating normally may appear to be under stress to the operator. This causes load shedding and other costly remedial actions by the operator.

Man-in-the-middle: Defense

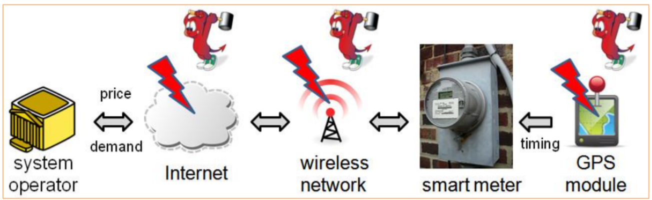
- Man-in-the-middle attack can succeed only when the attacker can impersonate each endpoint to the satisfaction of the other.
- Two crucial points in defending against the attack are *authentication* and *encryption*. We need to ensure that information from our devices will indeed go to the intended recipient.
- SSL (Secure Socket Layer) and HTTPS (HTTP over SSL) is the standard for creating this virtual trust and establishing secure communication between devices.



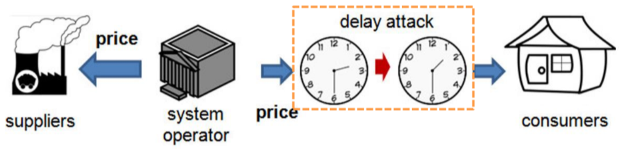
9.3.3 Integrity Attacks

Integrity Attacks on Real Time Pricing (RTP)

- We consider integrity attacks on the price received by consumers. Small malicious modifications to the price signals can be iteratively amplified, causing inefficiency and even severe failures such as blackouts.
- Delay attacks: The compromised price is an old price.



Integrity Attack - delay attack



- The delay attack can be launched by modifying the smart meters' internal clocks. Attacks on the clocks can be realized by compromising the vulnerable time synchronization services in smart grids. Current commercial smart meters synchronize their clocks by either built-in Global Positioning System (GPS) receivers or a network time protocol (NTP) supported by time servers. Both approaches have been shown to be vulnerable to realistic attack methods.
- Smart meters typically assign a memory buffer to store received prices. When a smart meter's clock has a lag, it will store newly received prices in the buffer and apply an old price for the present.

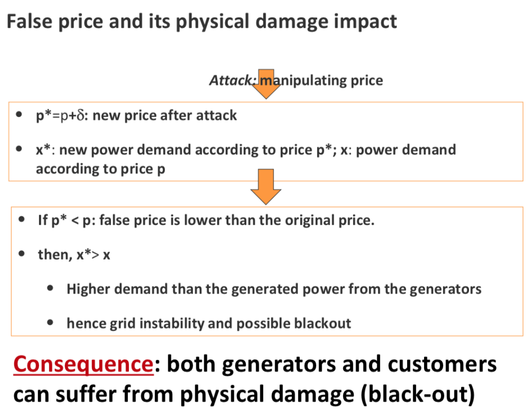
Attack on price information and its physical damage impact

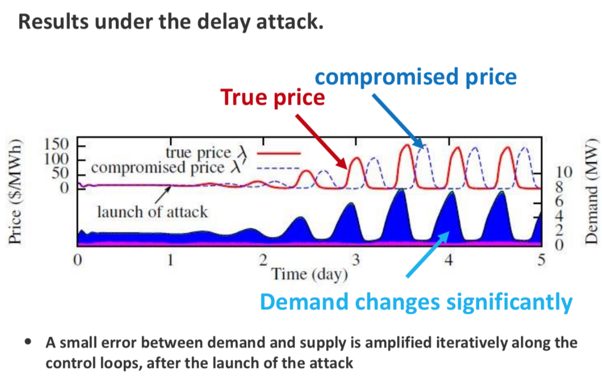
False or diverged price:

- Cyber attack with false price information injection or manipulation
- Overload power networks; Equipment damage

Consequence:

- New price leads to new power demand from users





False Data Injection Attacks, similar as integrity attack, here change the amount of energy demand amount

- The adversary compromises power demand nodes and injects forged quantity of demanded energy, e.g., requesting a large amount of energy.
- **Consequence**: This will mislead the electric utility into making incorrect decision about local usage. In addition, with the increase of forged demanded energy, the energy transmitted on links would be increase.
- **Requirement to implement this attack**: Requires the attacker to have knowledge of the power system. Adversaries have to manipulate some meters (or meter measurements) before they are used for state estimation.

Defending Integrity Attack: *rule-based*

- **Main idea**: false data injection attacks would result in abnormal state estimations (e.g. energy demands/supplies) , we can apply anomaly detection techniques.
- **Rule-based mechanisms**: typically exploit static thresholds to identify anomalies. When the value of a raw data input exceeds the thresholds, this value is regarded anomalous.
- *Advantage*: these schemes introduce very low overhead to the system
- *Disadvantage*: they fail to adapt to valid changes in the environment. Furthermore, a high level of professionalism is required to manually maintain detection thresholds.

Defending Integrity Attack: *machine learning based approaches*

- Statistical models (e.g. Bayesian, k-nearest neighbor) and artificial intelligence (e.g. neural networks) are utilized to distinguish anomalies from norms.
  - Bayesian network based approach: Uses the historical data of normal users to determine the conditional probabilities of data anomalies. Based on the conditional probabilities, a Bayesian network can be established to determine conditional probabilities of false data and limits its impacts on grid operations.
- Spatiotemporal correlation can be exploited since it is a natural property in smart grid since they are typically continuous over both the time and spatial domains.
  - Power suppliers generate energy based on the needs of consumers
  - Wind energy suppliers are typically correlated to nearby suppliers.
  - Energy consumers within the same area behave in similar patterns

9.4 More Considerations

Charging Electric Vehicles (EV)

- When connected to a charging station, the EV's charging profile must always be monitored. EV itself must also properly respond to command signals coming from the charging station.
- It is imperative that such commands come from an authentic source (the charging station) and reach the required host (authenticated EV) unaltered.
- In a typical charging scenario, an EV arrives at a charging station and requests a charging session. As stated by IEC 15118, each EV must have a secret key stored in its ECU (*Electronic Control Unit*).
- If the control station verifies the car's key through wireless communication, the charging process can be easily compromised by a Man-in-the-Middle attack. **(Q: are you able to see Man-in-the-Middle attack here?)**

Man-in-the-Middle for Electric Vehicles

- It was shown that a registered car with a valid key performs the authentication process and the charging cable is connected to a stolen car.
- Not only does Main-in-the-Middle attacks allow charging of unregistered cars, but also the firmware on the stolen car's ECU can be modified as not to respond to commands from the charging station.

10 Machine Learning for Renewable Energy Forecasting

10.0.1 Renewable Energy Forecasting

- Wind energy
  - Depend on wind and weather in general
  - **Main focus in this lecture**
- Solar energy
  - Depend on sun, cloud, and weather in general
- Wave energy
  - Wave energy converter are floating on the ocean surface waves
  - Wave power is the energy from ocean surface waves. The capacity depends on wave

Offshore & Onshore

Wind turbines can be installed:

- Onshore: on land
  - Cheaper installation
  - Cheaper integration
  - Cheaper maintenance
- Offshore: on sea
  - Less obstruction
  - Higher and more steady wind speed

Q: what is the advantage of offshore wind farm?

10.1 Wind Energy in Norway

Wind speed in Norway

- Wind power production depends on wind speed and varies considerably from year to year
- **Reference period:** the period 2000 - 2015 is chosen as the reference period, and the average wind conditions during this period are used to set default values for mean wind speed

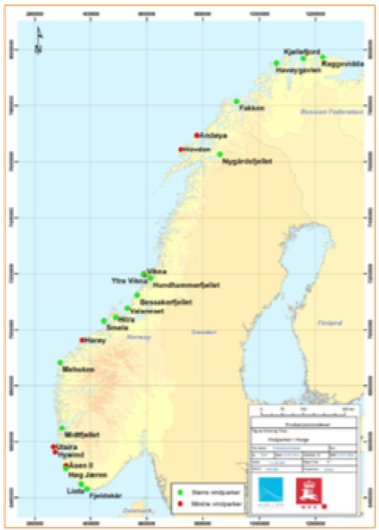
- **The figure right-side:** shows the deviations in average wind given in percent for 2015 compared to the reference period 2000-2015.

Main Observations

- Norway has a very good wind with wind level well above normal. The highest wind is in the area just north of Stadt (near Kristiansund).
- Power annual production depends on the power strength and it also depends on the wind regime, i.e., how the wind is distributed over the different wind strengths over year.
- *Reason:* the power curves are not linear and that power will not be able to utilize the energy in extreme winds fully (Wind speed of e.g., 25 m/s).

Wind farms in Norway

- Total power generation from wind energy of 2.5 TWh, corresponding to 1.7% of Norway's total electricity production
- **Green dot**: Large wind farms where reserve power curves are useful in the calculation of production time series
- **Red dot**: Small wind farms where a simple methodology is useful in calculation of production time series



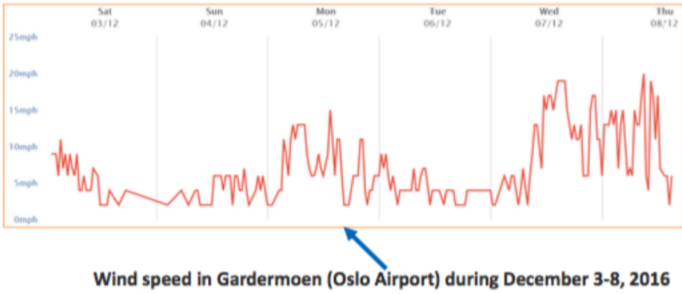
10.2 Wind Energy Forecasting

Wind forecasting and wind energy forecasting are important

- **Energy cost of intermittent wind:** unforecasted spin-regulation waste 2.5- 7.5% of total energy
- **System reliability:** unforecasted ramp events may compromise reliability
- **Economic value:** good forecasts may lead to high economic value
- **Effective power grid management:** forecasts are essential for effective grid management with high wind penetrations (>5%)

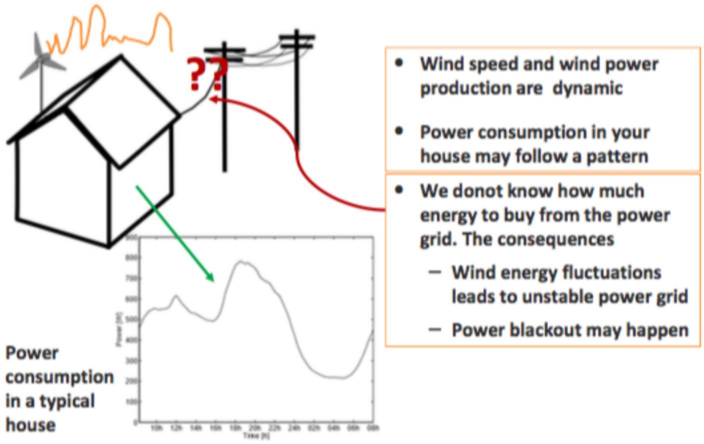
Wind forecasting is also challenging

- The key problem is the intermittency. Changes in wind speed will result in changes in wind power.



Wind speed in Gardermoen (Oslo Airport) during December 3-8, 2016

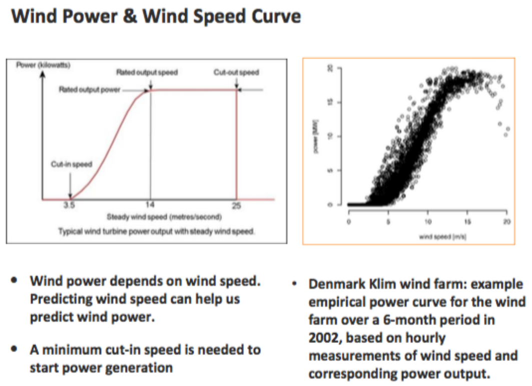
Consequence of unforecasted wind power – your house as an example

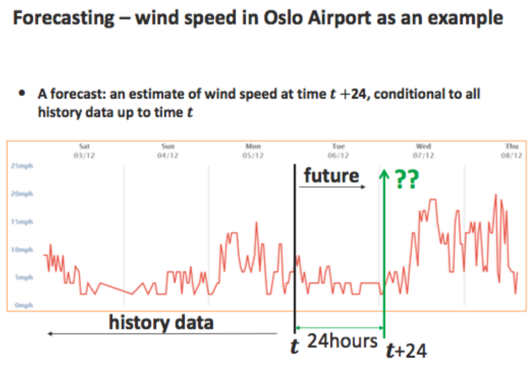


Wind energy forecasting is challenging

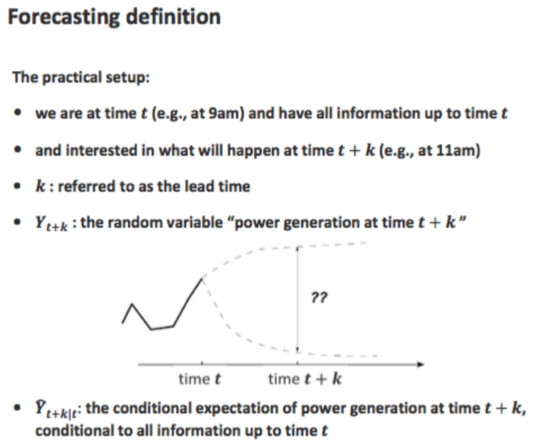
- Small pressure gradients over large distances: hard to forecast accurately
- Turbulent & chaotic processes are important: even harder to forecast
- Local topography can have strong influence: not in standard weather models
- Wind-power curves are highly nonlinear: small errors in wind = big errors in power
- Abnormal plants activities: malfunctions, downtime and sub-optimality

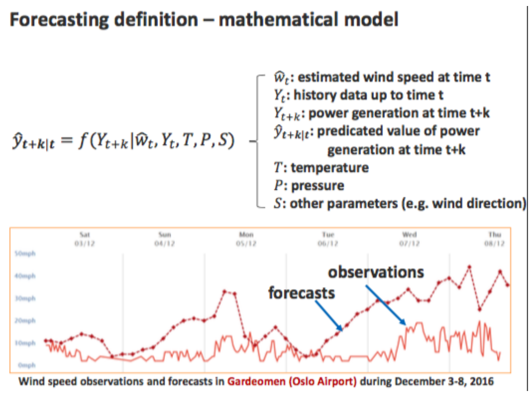
10.2.1 Wind power and wind speed curve



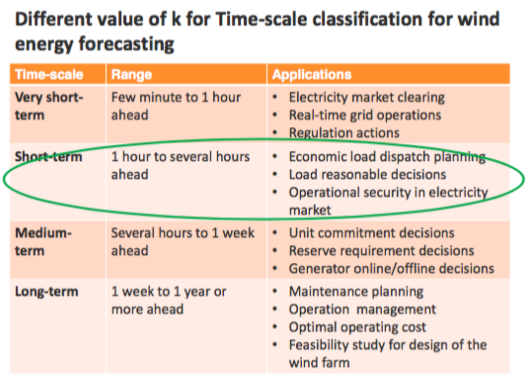


10.2.2 Forecasting definition

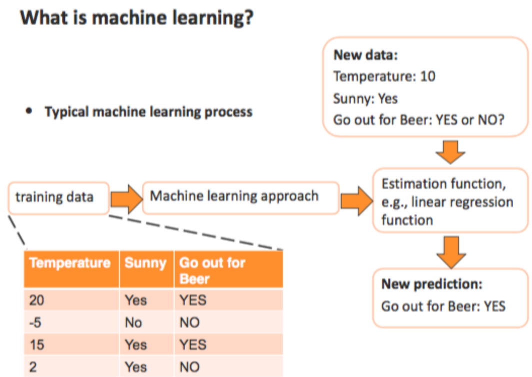


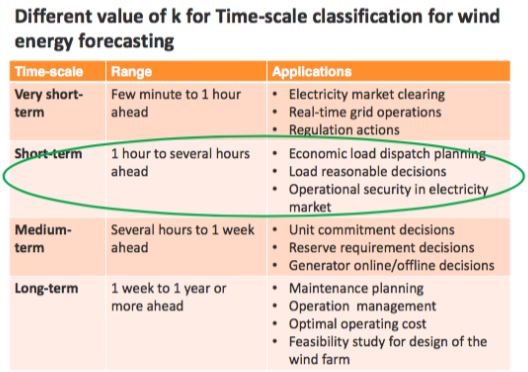


10.3 Machine Learning Techniques and Their Applications in Wind Energy Forecasting



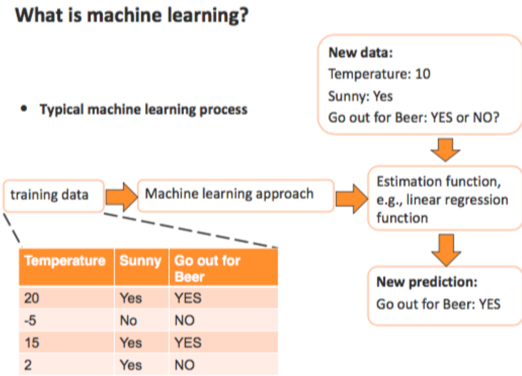
Time-scale	Range	Applications
Very short-term	Few minute to 1 hour ahead	• Electricity market clearing • Real-time grid operations • Regulation actions
Short-term	1 hour to several hours ahead	• Economic load dispatch planning • Load reasonable decisions • Operational security in electricity market
Medium-term	Several hours to 1 week ahead	• Unit commitment decisions • Reserve requirement decisions • Generator online/offline decisions
Long-term	1 week to 1 year or more ahead	• Maintenance planning • Operation management • Optimal operating cost • Feasibility study for design of the wind farm





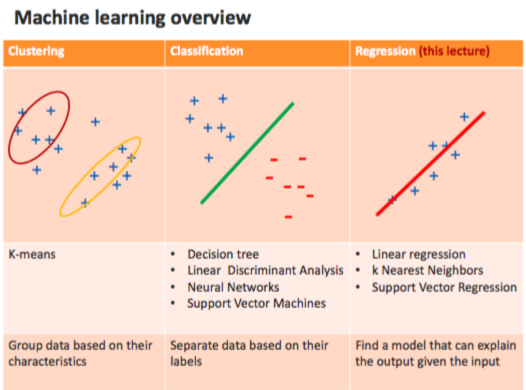
Different value of k for Time-scale classification for wind energy forecasting

Time-scale	Range	Applications
Very short-term	Few minute to 1 hour ahead	• Electricity market clearing • Real-time grid operations • Regulation actions
Short-term	1 hour to several hours ahead	• Economic load dispatch planning • Load reasonable decisions • Operational security in electricity market
Medium-term	Several hours to 1 week ahead	• Unit commitment decisions • Reserve requirement decisions • Generator online/offline decisions
Long-term	1 week to 1 year or more ahead	• Maintenance planning • Operation management • Optimal operating cost • Feasibility study for design of the wind farm



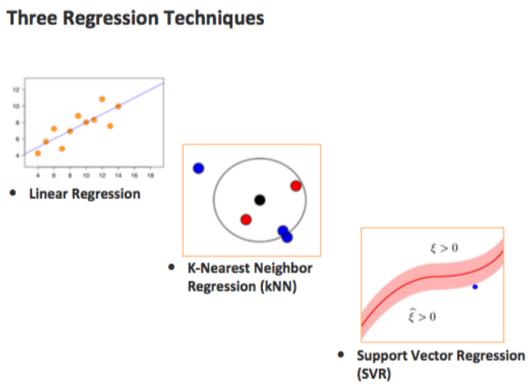
What is machine learning?

Temperature	Sunny	Go out for Beer
20	Yes	YES
-5	No	NO
15	Yes	YES
2	Yes	NO

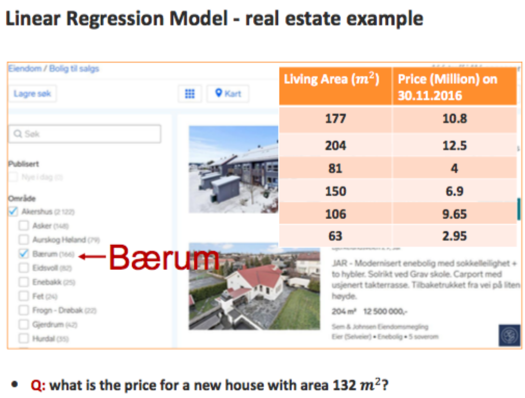


Machine learning overview

Clustering	Classification	Regression (this lecture)
K-means	• Decision tree • Linear Discriminant Analysis • Neural Networks • Support Vector Machines	• Linear regression • k Nearest Neighbors • Support Vector Regression
Group data based on their characteristics	Separate data based on their labels	Find a model that can explain the output given the input

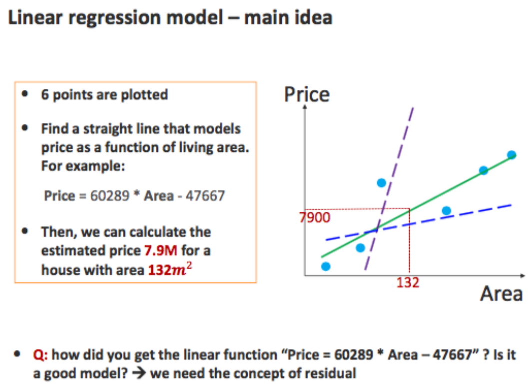


10.3.1 Linear Regression(LR)



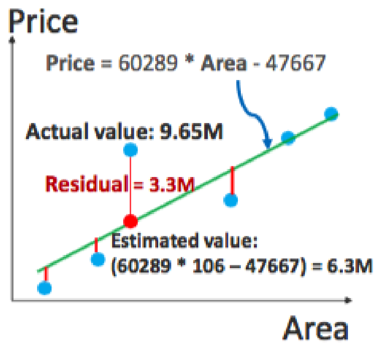
Linear Regression Model - real estate example

Living Area (m^2)	Price (Million) on 30.11.2016
177	10.8
204	12.5
81	4
150	6.9
106	9.65
63	2.95



Linear regression model - residual

- Not all observations are perfectly on the line. There is error between the actual value and the estimated value by the linear model
- *A residual*: is defined as the difference between the actual value and the estimated value by the linear approximation.
- We have interest in minimizing the overall error between the linear model and the actual observations. Here, overall error refers to: **the sum of residual squares.**



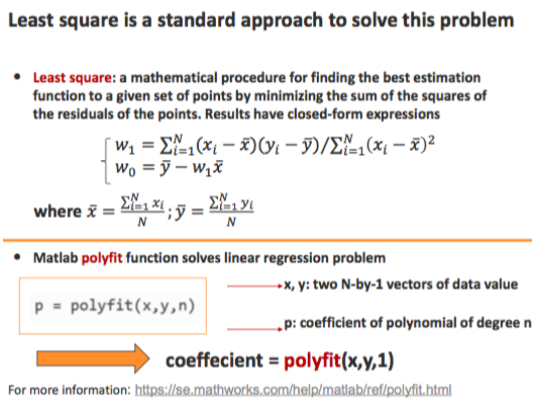
Linear regression model

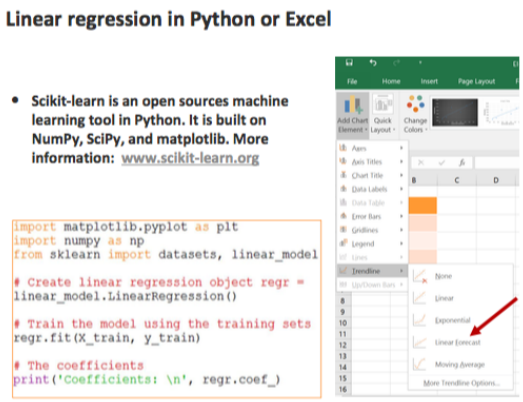
- Given ℕ observations (x_i, y_i); $(\mathbf{i} = \mathbf{1},\ \mathbf{2},\ \ldots\ \mathbb{N})$.
- *Goal*: to find the linear model (or estimation function) to minimize the overall error between the model and the actual observations.
- The linear model is defined as: $\mathbf{Y} = w_0 + w_1\mathbf{X}$
- The residual for observation $\mathbf{i}$ is given by

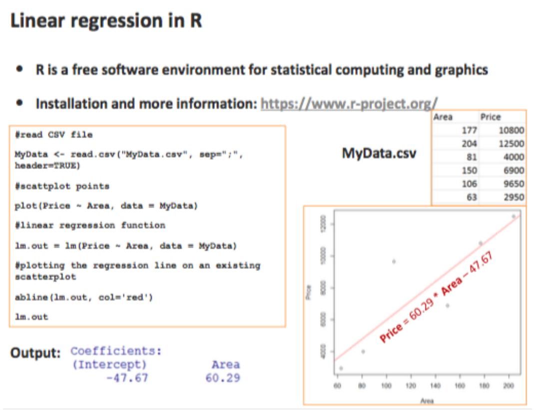
$$e_i = \underbrace{y_i}_{\text{Actual value}} - \underbrace{(w_0 + w_1 x_i)}_{\text{Estimated value}}$$

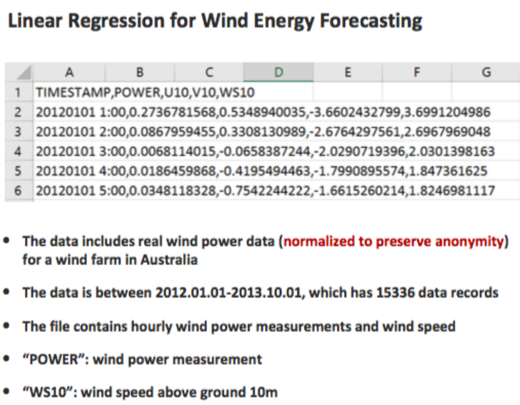
- The overall error between the linear model and the actual observations is defined as the sum of residual squares. Hence, we need to find coefficients w_0 and w_1 that minimize

$$\min_{w} \sum_{i=1}^{N} e_i^2$$









Find the Training Model, i.e., linear function

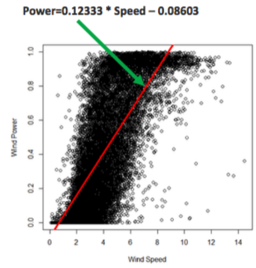
- The wind speed in the 336 test data records are used in the linear function to find the predicted wind power generation.
- The true wind power in the test data are then used to compare with the forecasted power generation. This can show if the training model (i.e., linear function) can provide accurate prediction.

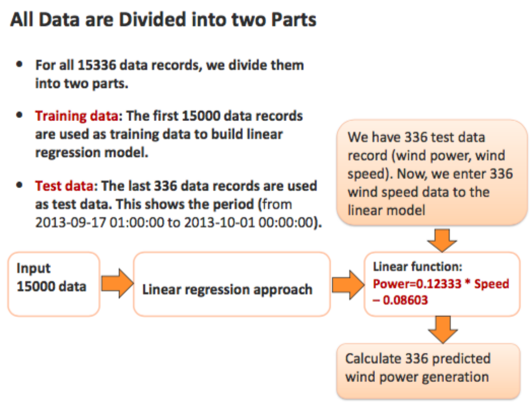
```
#read data from CSV file
data <- read.csv("WindPowerData.csv)

#the first 15000 data are training data
powertrain = data$POWER[1:15000]
wsTrain = data$WS10[1:15000]

#linear regression
lmOut = lm(powerTrain ~ wsTrain)
#get the coefficients in the linear model and save in coeffs
coeffs = coefficients(lmOut)
```

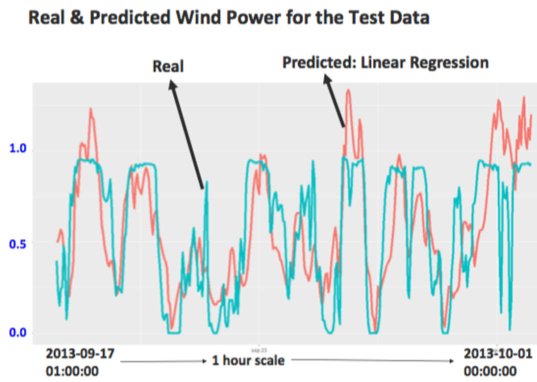
linear regression

Linear Regression for Wind Energy Forecasting

- Black circle "O": the scatter plotting of (wind speed, wind power) pairs for the training data
- We assume linear relationship between wind power and wind speed. Then, we use linear regression to demonstrate their relationship.
- **Q:** is linear relationship a good model?

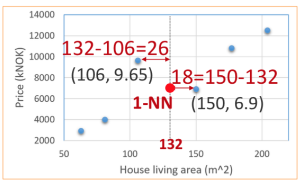






10.3.2 K-Nearest Neighbor (kNN)

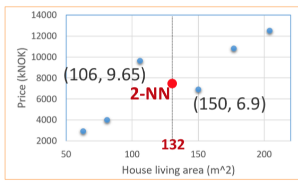
k-Nearest Neighbor (kNN) Regression. Here, 'k' is a variable, k=1,2,3....



- **Q:** what is the price for a house with area 132 m^2?

- 1-NN (one nearest neighbor)
  - the observation (150, 6.9) is the nearest point from the area 132 m^2
  - then, the predicted price has the same price as the house with area 150 m^2, which is 6.9M

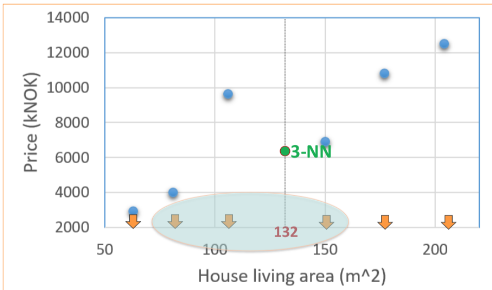
2-Nearest Neighbor (2NN)



- **Q:** what is the price for a house with area 132 m^2?

- 2-NN (two nearest neighbors)
  - the observations (106, 9.65) and (150, 6.9) are the two nearest node from the area 132 m^2
  - Then, the predicted price is the average price of these two houses, i.e., (9.65+6.9)/2= 7.825M

K-Nearest Neighbor Regression (kNN)



- 3-NN (three nearest neighbors): the observations (106, 9.65), (150, 6.9) and (81, 4) are the three nearest node from the area 132 m^2, so, the prediction is the average of 9.65M, 6.9M and 4M, which is 6.55M

kNN Algorithm

Given:

- Training set $(x_i, y_i), (i = 1, 2 \dots N)$
  - x_i: attribute-value representation of examples
  - y_i: real-valued target (e.g., price, rating on YouTube, profit etc)
- Testing point x that we want to predict the target

Algorithm

- Compute the distance $D(x, x_i)$ to every training example x_i: $D(x, x_i) = |x - x_i|$
- Select k closest points $(x_{i1}, x_{i2}, \dots, x_{ik})$ and their values $(y_{i1}, y_{i2}, \dots, y_{ik})$
- Output the average value of $y_{i1}, y_{i2}, \dots, y_{ik}$: $$\overline{y} = \frac{\sum_{i=1}^{k} y_{ik}}{k}$$

kNN for Wind Energy Forecasting

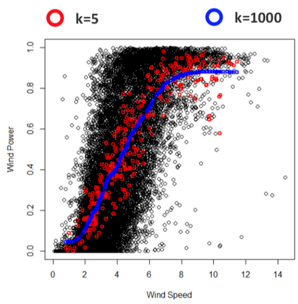
- Same real wind power data for a wind farm in Australia
- For 15336 data records, we divide them into two parts. The first 15000 data records are used as training data to build kNN model. The wind speed in the rest 336 records are used as test data to find the wind power generation

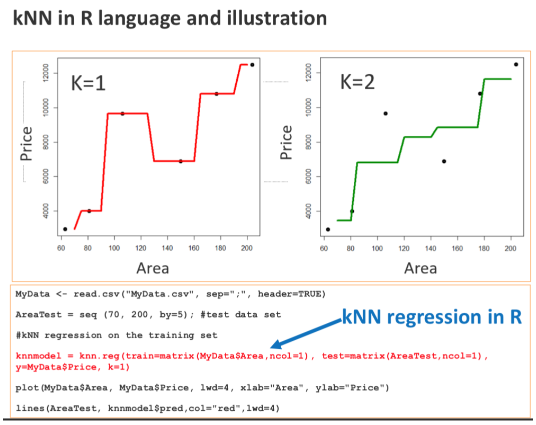
```
powerAll = data$POWER #all wind power data set
wsAll = data$WS10 #all wind speed data set
trainLength = 15000 #number of data records to be used in training set
testLength = length(wsAll) - trainLength
wsTrain = wsAll[1:trainLength] #wind speed - training data set
powerTrain = powerAll[1:trainLength] #wind power - training data set
wsTest = wsAll[trainLength+1 : length(wsAll) ] #wind speed - test data set
wsTest = wsTest[1:testLength] #wind speed - training data set
knnmodel =
knn.reg(train=matrix(wsTrain,ncol=1),test=matrix(wsTest,ncol=1),y=powerTrain, k=10)
```

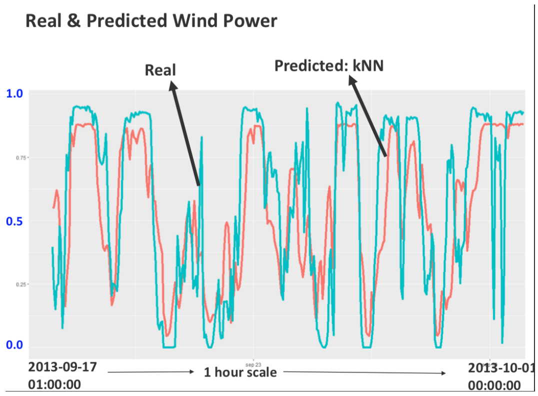
kNN regression

kNN for Wind Energy Forecasting

- **Black circle** “O”: the scatter plotting of (wind speed, wind power) pairs in the training set
- **Red circle** “O”: (wind speed, predicted wind power) pairs when k=5
- **Blue circle** “O”:(wind speed, predicted wind power) pairs when k=1000
- In this kNN model, the curve is not linear
- **Q:** what is an appropriate k?



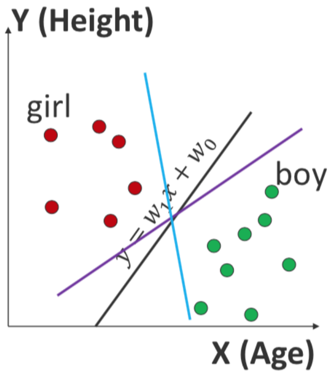


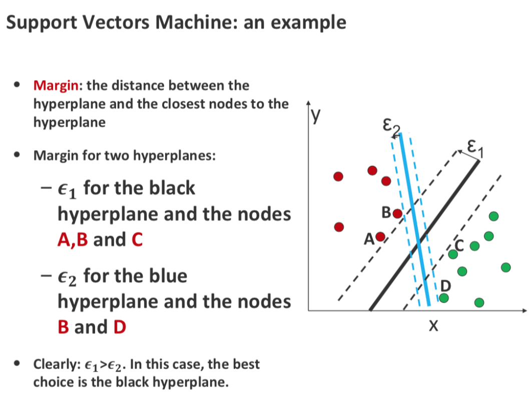


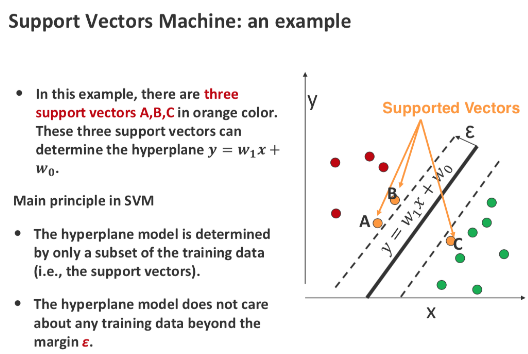
10.3.3 Support Vector Regression(SVR)

Support Vectors Machine: main idea

- Supported Vectors Machine (SVM) can be used for classification and regression. Let's see an example of classification to understand the main idea.
- Two sets: green circle and red circle.
- Goal: draw a line (called as hyperplane) that classifies all data in two classes.
- **Q:** Three lines in the right, which line shall we choose?
- **Best choice:** line that leaves the maximum margin from both classes, i.e., the black line





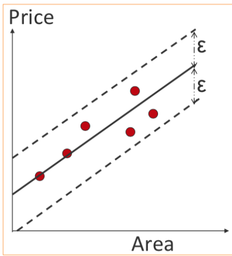


SVM for Regression Problem: Supported Vector Regression

- Wind energy forecasting is a regression problem. SVM is used for regression and called as Supported Vector Regression (SVR).

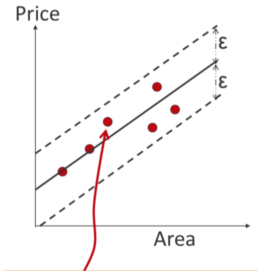
Main idea in SVR:

- **Margin**: we set a margin ε which is similar as the concept in SVM.
- We ignore the errors as long as they are less than ε, but will not accept any deviation larger than this.



ε-intensive error: concept

- **ε-intensive error:** error is zero if the points are inside the band. Only the points out of the band will be considered.
- In SVR, we draw a line such that the distance between the points and the line is not be higher than ε.
- The line is in the middle of the set of points such that the line is as close to the points as possible.
- In R programming, SVR uses default value ε=0.1



We donot care about errors as long as they are less than ε.

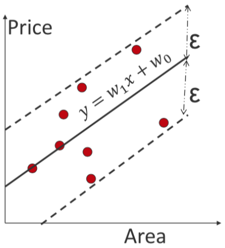
Support Vectors Regression (SVR)

- *Goal*: to find a function with at most ε-intensive error from the actual value y_i for all training data; and at the same time is as flat as possible.
- For function $y = w_1 x + w_0$, flatness means that one seeks a small w_1

$$\min \frac{1}{2} {w_1}^2$$

$$s.t.\, y_i - (w_1 x_i + w_0) \le \varepsilon$$

$$(w_1 x_i + w_0) - y_i \le \varepsilon$$



Q: why flatness? To avoid overfitting problem, we want the function to be as flat as possible. Then, the function is less sensitive to x or the change of x. (Overfitting: the model does a perfect job of fitting the training data, but will do a bad job of predicting new data)

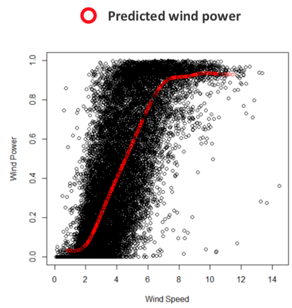
SVR for Wind Energy Forecasting

- We use the same real wind power data for a wind farm in Australia
- For all 15336 data records, we divide them into two parts. The first 15000 data records are used as training data to build SVR model. The wind speed in the rest 336 records are used as test data to find the wind power generation

```
#use the same source code to get wsTrain, powerTrain, wsTest
plot(wsTrain, powerTrain, xlab="Wind Speed", ylab="Wind Power")
#Support vector machine/regression
df <- data.frame(x = wsTrain, y = powerTrain)
svrmodel <- svm(y ~ x, data = df)
powerPredicted = predict(svrmodel, newdata = data.frame(x = wsTest))
# Add points for fitted svrmodel
points(wsTest, powerPredicted , col = "red" )
```

SVR for Wind Energy Forecasting

- Black circle "O": the scatter plotting of (wind speed, wind power) pairs in the training set
- Red circle "O": (wind speed, predicted wind power) according to the wind speed in the test dataset
- Similar as kNN model, SVR-model based predicted wind power also follow a non- linear curve



Metrics measures the model goodness in prediction

- **Q:** we discussed three models: linear regression, kNN and SVR. Which model is the best?
- For each data point x_i, its real value is y_i. A model makes a prediction $\hat{y}_i$. In order to measure how good a model is we will compute how much error it makes. We compare each real value y_i with its predicted value $\hat{y}_i$ and calculate the difference.
- The expression $\hat{y}_i - y_i$ is the error. For each data point, we have the sum of the squared errors, and then take the mean. We have the Mean Squared Error (MSE)

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$$

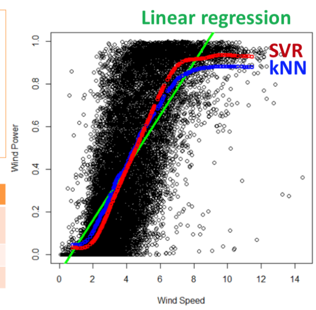
- A common way to measure error uses the Root Mean Squared Error (RMSE), we take the square root of MSE

$$RMSE = \sqrt{MSE}$$

An illustration to show the RMSE between these models

```
rmse <- function(error)
{    sqrt(mean(error^2))
}
# Calculate error
errorSVR <- powerTest - powerPredicted
rmse(errorSVR)
```

Model	RMSE
Linear regression	0.2545
kNN (k=1000)	0.2276
SVR	0.2343



10.4 More Considerations

10.4.1 Interested to forecast something new?

- What else may you need to forecast, in an electricity market context?
  - solar power generation
  - energy prices
  - electricity load
  - battery health prediction
- What else may you need to forecast, in general?
  - how many travelers in Oslo this summer?
  - traffic flow prediction in Opera tunnel
  - Norway's wealth fund hit \$1trillion

More parameters

- House price is not only dependent on the area. It is also related to location, public transport, crime statistics, school rankings, etc. How to use machine learning to model and solve this problem?
- For example, if we consider more factors affecting house price, we have **Multiple Linear Regression** model

$$\text{Price} = w_0 + w_1 * \text{Area} + w_2 * \text{Location} + w_3 * \text{SchoolQuality}$$

- Similarly, wind power generation is not only dependent on wind speed. It is also related to wind direction, temperature, and pressure.

$$\text{Wind Power} = w_0 + w_1 * \text{WindSpeed} + w_2 * \text{WindDirection} + w_3 * \text{Temperature} + w_4 * \text{Pressure}$$

Code: `lm(powerTrain ~ wsTrain + winddirection + temperature + pressure)`

Different time horizon prediction

- Normally we need short-term forecasting for solar or wind energy power generation
- **Q:** in what cases, we need very short-term forecasting, e.g., 5mins?

Solar power plant in desert

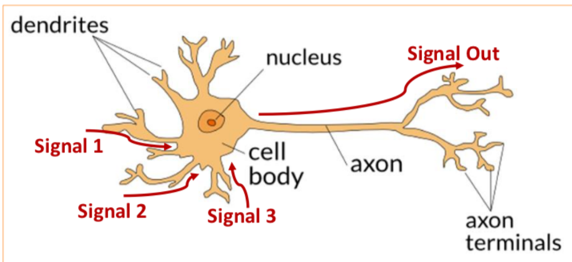
11 Deep Learning for Renewable Energy Forecasting

11.1 Neural Network

11.1.1 Neural Network (ANN)

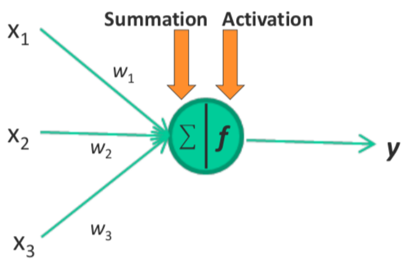
Computers are great at solving algorithmic and math problems, but many problems can't easily be defined with a mathematical algorithm. Facial recognition is a typical such example. Computers believe that the two photos show the same person. However, this is apparently incorrect and such tasks are trivial to humans.
Artificial Neural Networks main idea: enables computers to process information in a similar way as our own biological brains and our own nervous system.

Biological Model for a Single Neuron



Incoming signals are received by the cell's *dendrites* through a biochemical process. The process allows the impulse to be weighted according to its relative importance or frequency. As *the cell body* accumulating the incoming signals, a threshold is reached at which the cell fires and the output signal is transmitted via an electrochemical process down the *axon*

11.1.2 From Biological Model to a Single Artificial Neuron



- This figure defines the relationship between the input signals (**x variables**) received by the dendrites and the output signal (**y variable**).
- Each dendrite's signal is weighted (**w values**) according to importance.

- The input signals are summed by the cell body and the signal is passed on according to an **activation function denoted by f**.
- **Model for a single artificial neuron**: A typical artificial neuron has $\boldsymbol{n}$ input dendrites. The $\boldsymbol{w}$ weights allows each of the $\boldsymbol{n}$ inputs (denoted by x_i to contribute a greater or less amount to the sum of input signals. The total is used by the activation function $f(x)$, and the resulting signal $\boldsymbol{y}$, is the output axon:

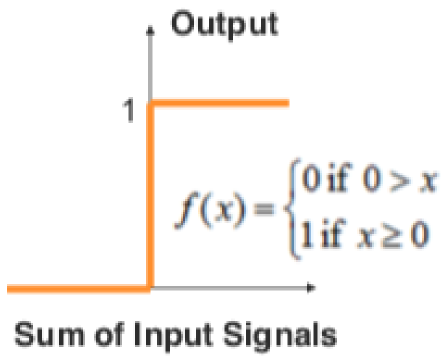
$$y(x) = f(\sum_{i=1}^{n} w_i x_i)$$

- **Activation function f(.)**: can be threshold activation function, sigmoid function or other different functions.

11.1.3 Activation Function

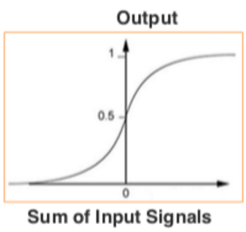
Threshold activation function

- In the biological case, the activation function is the process that involves summing the total input signal and determining whether it meets the firing threshold. If so, the neuron passes on the signal; otherwise it does nothing
- Threshold Activation Function: it results in an output signal only once a specified input threshold has been attained.
- The neuron fires when the sum of the input signals is at zero.
- The threshold activation function parallels with biology, but it is rarely used in artificial neural networks. (**Q: why?**)



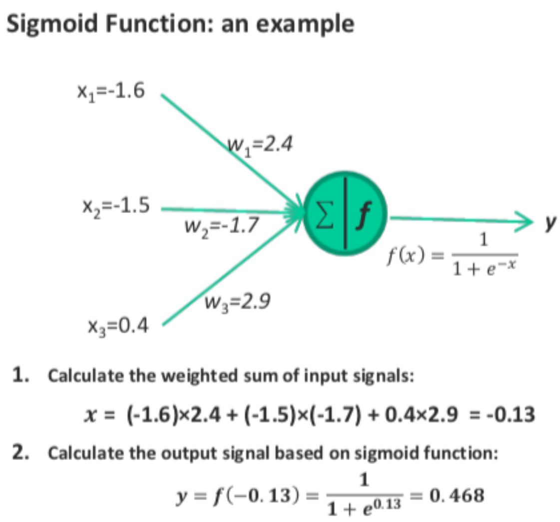
Sigmoid activation function

- Sigmoid activation function: the most commonly used activation function
- The output signal is no longer binary; the output values can be any value in the range from 0 to 1.
- The sigmoid is **differentiable**, and it is possible to calculate the derivative across the entire range of inputs. In addition, the derivative function can be easily computed and can significantly reduce the computation cost during training.



$$f(x) = \frac{1}{1+e^{-x}}$$

$$f'(x) = f(x)[1 - f(x)]$$

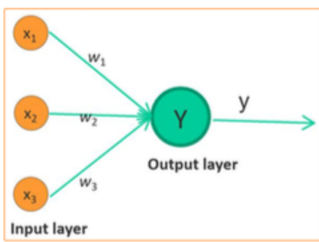


11.1.4 From a Single Artificial Neuron To Neural Networks

- Neural Networks: we use many neurons as building blocks to construct complex models of data.
- A neural network has three characteristics
- **Activation function** $f(x)$: transforms a neuron's combined input signals into a single output signal to be broadcasted further in the networks.
- **Network topology**: describe the number of neurons in the model, the number of layers, and the manner in which the layer are connected
- **Training algorithm**: specifies how weights $\boldsymbol{w}$ are set in order to excite neurons in proportion to the input signal.

11.1.5 Network Topology

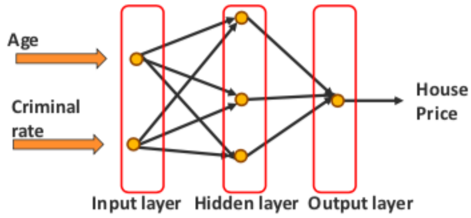
A single layer neural network



- **Input layer**: A set of neurons receives the input data. Each input node is responsible for processing a single feature in the dataset. The feature's value will be transformed by the corresponding node's activation function.

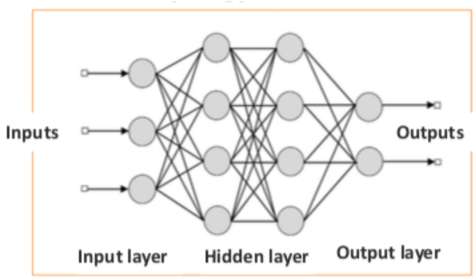
- **Output layer**: The signals sent by the input nodes are received by the output layer, which uses its activation function to generate a final prediction **y**.
- This is a single-layer neural network: the input nodes process the incoming data, the network has only one set of weights (**w1, w2, w3**).

A typical neural network with one hidden layer

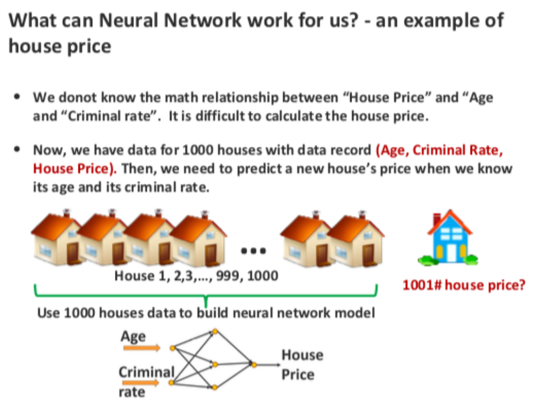


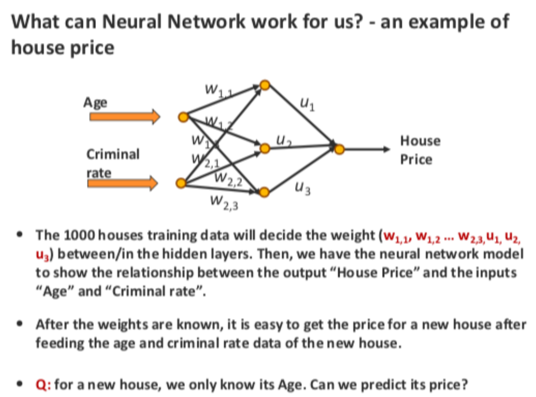
- A typical neural network topology used to predict the house price. It has 2 input nodes, 1 hidden layer with 3 nodes, and 1 output node.
- The input nodes feed the attributes (Age, Criminal rate) into the network. There is no input node for each attribute.
- The information flows unidirectional to the hidden layer; and then to the output layer. The output calculates a weighted sum on the received data to predict house price.

11.1.6 De facto standard topology



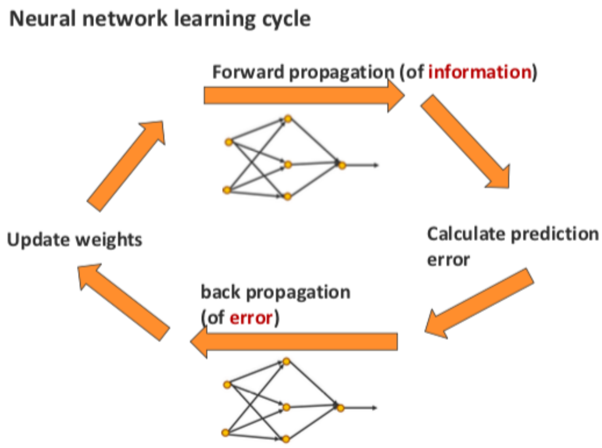
- An artificial neural network has multiple layers and it is an interconnected group of nodes. One or more hidden layers are added that process the signals from the input nodes prior to it reaching the output node.
- This topology is also called **Feedforward neural network** since the information flows unidirectional from the input layer to the hidden layer; and then to the output layer.
- This is the de facto standard neural network topology.





11.1.7 We need to decide the weight: Backpropagation Algorithm

- **Need to decide the weights**: There are many learning algorithms. They will train the network by iteratively modifying the weights until the error between the output produced by the network and the desired output falls below a specified threshold.
- **Backpropagation (BP) algorithm**: the first popular learning algorithm and is still widely used. It uses **gradient descent** as the learning mechanism. Starting from random weights, the backpropagation algorithm calculates the weights and gradually makes adjustments determined by the error between the result produced by the neural network and the real outcome.
- The algorithm applies error propagation from outputs to inputs, and gradually fine tunes the network weights to minimize the sum of error.



11.1.8 Training Neural Networks with Backpropagation

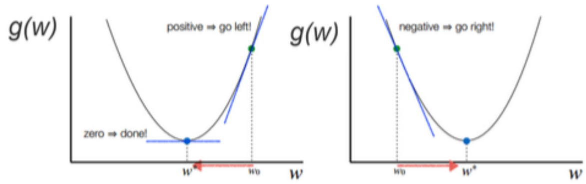
- **Forward phase**: neurons are activated in sequence from the input layer to the output layer, applying each neuron's weights and activation function along the way. Upon reaching the final layer, an output signal is produced.
  - **Forward phase focuses on information forward propagation**
- **Backward phase**: network's output signal resulting from the forward phase is compared to the true value in the training data. The difference between the output signal and the true value results in an error. This error is propagated backwards in the network to adjust the weights and reduce future errors.
  - **Backward phase focuses on error backward propagation**

Gradient Descent Technique for Adjusting Weight (I): general idea

- **Gradient Descent**: the algorithm iteratively updates the weights in the direction of the gradient of the error function until a minimum is reached. We need to adjust weights in the direction of reducing error.
- This is analogous to going down a hill. When we are standing on top of a hill. We look around and decide the direction to go. Then, we can reach the bottom of the hill fastest, clearly the route with red color instead the route with white color.

The steepest stretch of road from Geiranger towards Eidsdal on road no. 63.

Gradient Descent Technique for Adjusting Weight (II): how to adjust?



- w: weight; $g(w)$: error function; $g'(w)$: derivative function
  ➔ importance of having a differentiable activation function
- $g'(w_0) > 0$: the weight is decreased
- $g'(w_0) < 0$: the weight is increased
- $g'(w_0) = 0$: we arrive at the optimal weight w^*

Neural networks package **neuralnet** in R

- You need to install it by typing install.packages("neuralnet") and load it with the library(neuralnet) command.
- Build the model:
  - m <- neuralnet(target~predictors, data = MyData, hidden = 1)
  - target: outcome in the MyData data frame to be modeled
  - predictors: features in the MyData data frame to use for prediction
  - data: data set where the target and predictors variables can be found
  - hidden: the number of neurons in the hidden layer
- Making predictions:
  - p <- compute(m, test)
  - m: trained model by the neuralnet() function
  - test: test data with the same features as the training data
- More info: https://cran.r-project.org/web/packages/neuralnet/index.html

Boston House Price (we do not have Oslo house price)

1	CRIM,ZN,INDUS,CHAS,NOX,RM,AGE,DIS,RAD,TAX,PTRATIO,LSTAT,MEDV
2	0.00632,18,2.31,0,0.538,6.575,65.2,4.09,1,296,15.3,4.98,24
3	0.02731,0,7.07,0,0.469,6.421,78.9,4.9671,2,242,17.8,9.14,21.6
4	0.02729,0,7.07,0,0.469,7.185,61.1,4.9671,2,242,17.8,4.03,34.7

....

505	0.06076,0,11.93,0,0.573,6.976,91,2.1675,1,273,21,5.64,23.9
506	0.10959,0,11.93,0,0.573,6.794,89.3,2.3889,1,273,21,6.48,22
507	0.04741,0,11.93,0,0.573,6.03,80.8,2.505,1,273,21,7.88,11.9

- The Boston data frame has 506 rows and 13 columns. For each column data,
- CRIM: per capita crime rate by town; ZN: proportion of residential land zoned for lots over 25,000 sq.ft; INDUS: proportion of non-retail business acres per town; CHAS: Charles River dummy variable (= 1 if tract bounds river; 0 otherwise); NOX: Nitrogen oxides concentration (parts per 10 million); RM: average number of rooms per dwelling; AGE: proportion of owner-occupied units built prior to 1940; DIS: weighted mean of distances to five Boston employment centres; RAD: index of accessibility to radial highways; TAX: full-value property-tax rate per \$10,000; PTRATIO: pupil-teacher ratio by town; LSTAT: lower status of the population (percent);
- MEDV: median value of owner-occupied homes in $1000s.

Training Data, Test Data, and Experiment

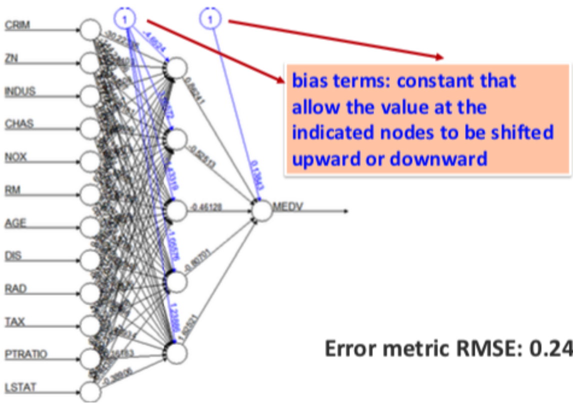
- For 506 data records, we divide them into two parts. We use 75% first data record (from 1 to 380 data records) are used as training data to build neural network model. The last 25% data records (from 381 to 506 data records) are used as test data.
- We try neural network with one hidden layer that has 5 nodes

```
bostonModel <- neuralnet(MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM
+ AGE + DIS + RAD + TAX + PTRATIO + LSTAT, data = trainingData,
hidden = 5)
```

- We try neural network with three hidden layers with
  - 5 nodes in the first hidden layer
  - 4 nodes in the second hidden layer
  - 3 nodes in the third hidden layer

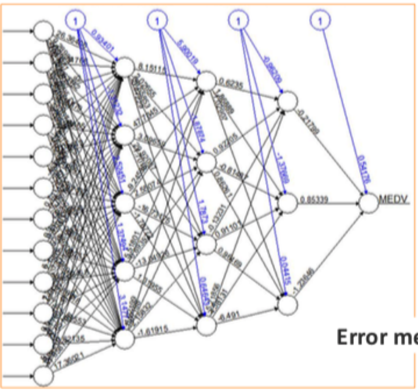
```
bostonModel <- neuralnet(MEDV ~ CRIM + ZN + INDUS + CHAS + NOX + RM
+ AGE + DIS + RAD + TAX + PTRATIO + LSTAT, data = trainingData,
hidden = c(5,4,3))
```

Neural network with one hidden layer and 5 hidden nodes in the hidden layer



Error metric RMSE: 0.246

Neural network with three hidden layers (5 nodes in the 1st layer, 4 nodes in the 2nd layer, and 3 nodes in 3rd layer)



Error metric RMSE: 0.137

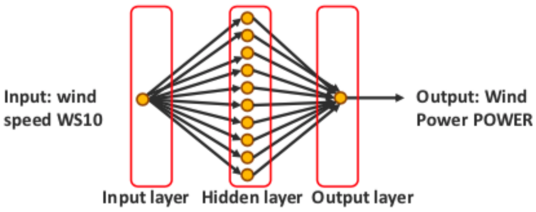
Neural Network for Wind Energy Forecasting

	A	B	C	D	E	F	G
1	TIMESTAMP,POWER,U10,V10,WS10						
2	20120101 1:00,0.2736781568,0.5348940035,-3.6602432799,3.6991204986						
3	20120101 2:00,0.0867959455,0.3308130989,-2.6764297561,2.6967969048						
4	20120101 3:00,0.0068114015,-0.0658387244,-2.0290719396,2.0301398163						
5	20120101 4:00,0.0186459868,-0.4195494463,-1.7990895574,1.847361625						
6	20120101 5:00,0.0348118328,-0.7542244222,-1.6615260214,1.8246981117						

- The data is between 2012.01.01-2013.10.01, which has 15336 data records. The file contains hourly wind power measurements and wind speed
- The data includes real wind power data (normalized to preserve anonymity) for a wind farm in Australia

Neural Network for Wind Energy Forecasting

- One input node in the input layer
- One hidden layer, 10 hidden nodes in the hidden layer
- One output node



R code to build neural network model

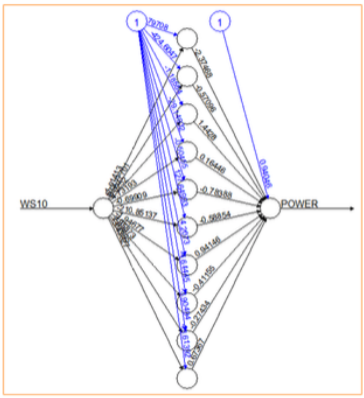
- For 15336 data records, we divide them into two parts. We use 3000 data record (from 12001 to 15000 data records) are used as training data to build neural network model. The wind speed in the last 336 records (from 15001 to 15336 data records) are used as test data

```
#calling neuralnet function to build the model.
#The function will return a neural network objec that can make predictions

powerModel <- neuralnet(POWER ~ WS10, data = trainingData, hidden = 10)

#generate the prediction on the test data
modelResults <- compute(powerModel, testData$WS10)
powerPredictedNN = modelResults$net.result

# plot the neural networks
plot(powerModel)

# Calculate error
errorNN <- powerPredictedNN - testData$POWER
rmse(errorNN)
```

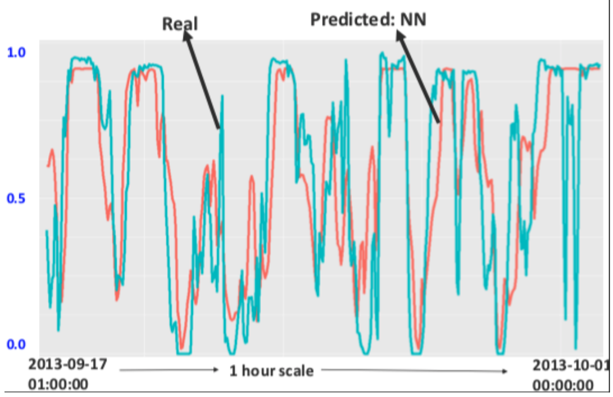
Neural network training

Neural network model with one hidden layer (10 hidden nodes)

- Q: can we use all 15000 data records as training data?



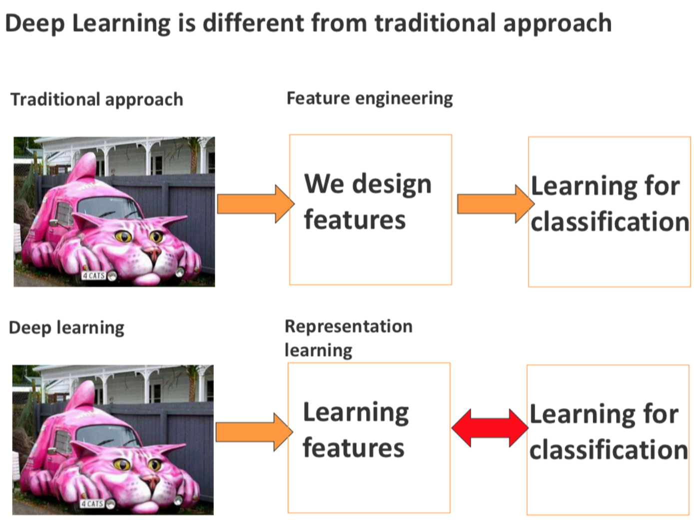
Real & Predicted Wind Power. RMSE=0.2277



11.2 Recurrent Neutral Networks (RNNs)

11.2.1 Deep learning: definition

- Deep learning: not an $\boldsymbol{n}$-layered neural network where n is a very big number. Even a simple two layered ANN which consists of input and output layers only is classified as a Deep Learning model.
- Deep learning breakthrough came in 2006, when methods were developed to overcome the difficulties in training deep neural networks. Deep learning is essentially multiple layer neural network with layer-by-layer training.
  - For example, for image recognition: Pixel (hidden layer 1) $\rightarrow$ edge (hidden layer 2) $\rightarrow$ texton (hidden layer 3) $\rightarrow$ motif (hidden layer 4) $\rightarrow$ part (hidden layer 5) $\rightarrow$ object (hidden layer 6)

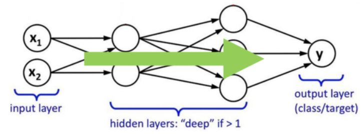


11.2.2 Challenges of traditional feedforward neural networks

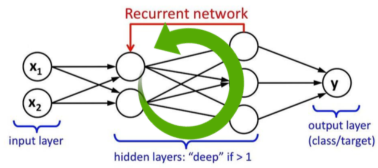
- Feedforward neural networks
  - input signals unidirectional from the input layer to the output layer
  - all input and outputs are independent of each other.
- Main challenges
  - If we want to predict the next word in a sentence, we'd better know which word came before it. $\rightarrow$ **Traditional neural networks have no history information!**
  - When we read a book, we understand each word on our understanding of previous words. We don't throw everything away and start thinking from scratch again. We have a memory. $\rightarrow$ **Traditional neural networks has no memory!**

11.2.3 Traditional neural network and Recurrent neural network

- Traditional artificial neural networks: input signals unidirectional from the input layer to the output layer.



- Recurrent neural networks: a kind of neural network which will send current information back to itself. As a result, it has memory and can "remember" the history information.



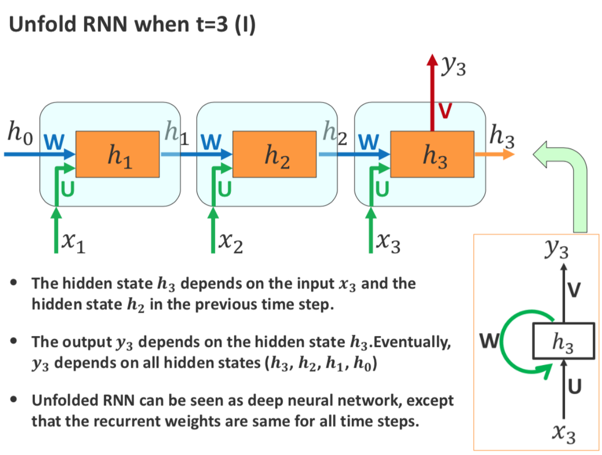
11.2.4 Recurrent Neural Networks (RNN)

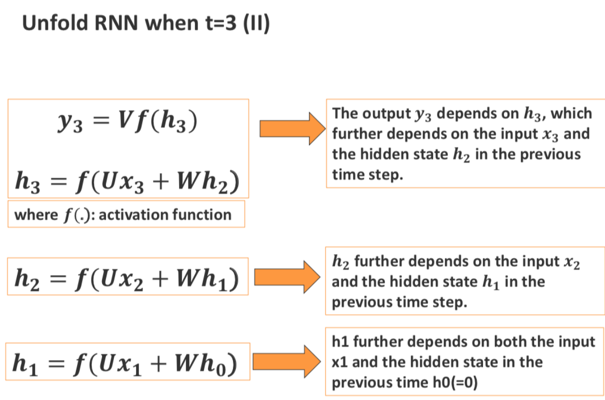
Main Features:

- RNN allows signals to travel in both directions. This property mirrors more closely how a biological neural network works.
- RNN has loops, allowing information to persist.
- RNN has a "memory" which captures information about what has been calculated so far.
- RNN can have a very long memory with variants of RNN

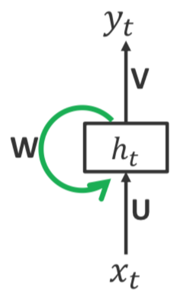
Application:

- RNN is usually very good at predicting sequences or time series data. If your task is to predict a sequence or periodic signal, then using a RNN might be a good starting point.



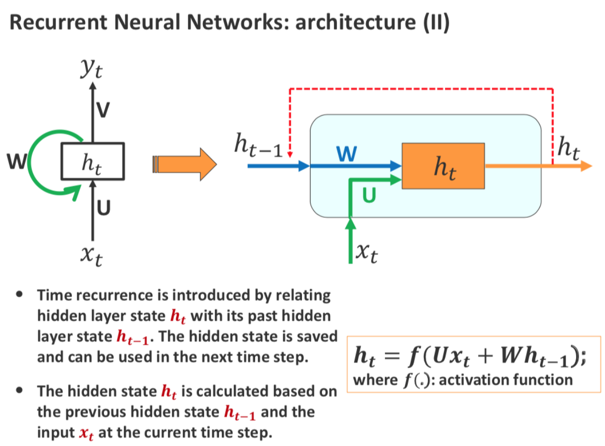


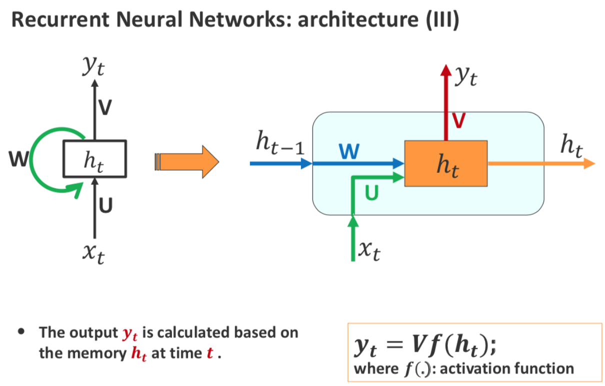
11.2.5 Reccurent Neural Networks: architecture



Recurrent neural networks have

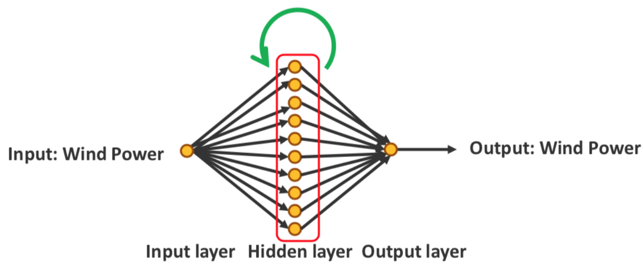
- two inputs at time steps **t (t=1,2,3,..):**
  - the present input x_t
  - the hidden state h_t: here, we use "**state**" to refer to a set of values that summarizes all the information about the past behavior of the system. The hidden state h_t captures information about what happened in a*all* the previous time steps. It is the memory of the network.
- Output y_t: two sources of input are combined to decide the output
- **U, V, W** weights: U for the input layer, W for the hidden layer and V for the output layer.



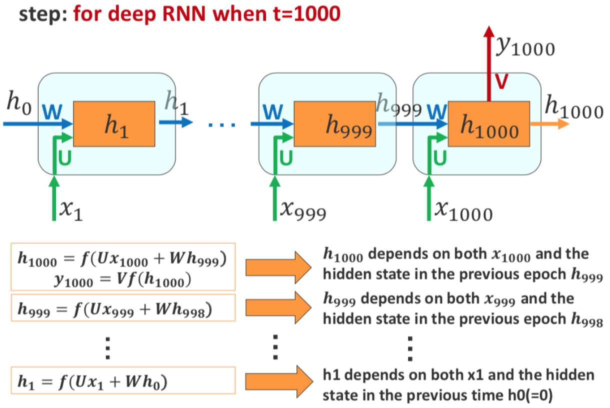


11.2.6 Recurrent Neural Network for Wind Energy Forecasting

- One input node in the input layer
- One recurrent hidden layer, 10 hidden nodes in the hidden layer
- One output node
- Input: wind power; output: wind power → **time series forecasting**



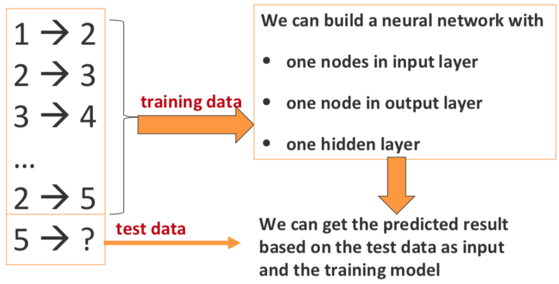
RNNs connect previous information to the present time step: for deep RNN when t=1000



Prediction for time-series: a simple example

1,2,3,4,3,2,1,2,3,4,3,2,5,?

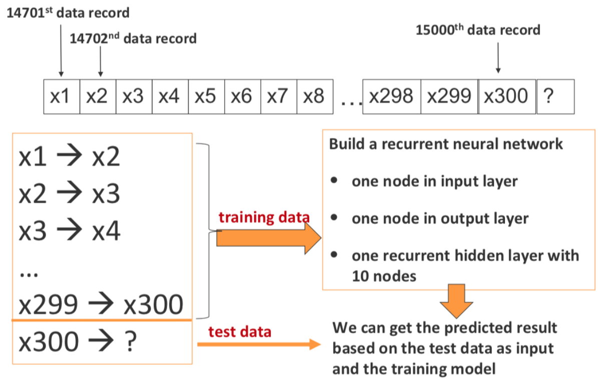
- Q: we have a series of numbers, what is the number after 5?
- We might use a number to predict the immediate next number, which means: one single input neuron and one single output neuron.



Encoding for Wind Power Data

- For 15336 data records, we only use the wind power data POWER
- We use 300 data record as the training data to build recurrent neural network model. We use the last 336 records (from 15001 to 15336 data records) as test data to predict the wind power generation

Coding window for 300 training data



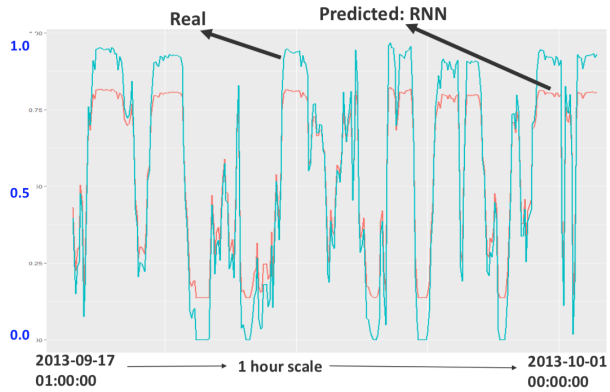
R code to build recurrent neural network model

```r
#read CSV file
data <- read.csv("WindPowerDataAU.csv")
Power <- data$POWER

TotalDataLength <- length(Power)
TotalTrainLength <- 15000
ActualTrainLength <- 300

#use x_i to predict x_(i+1)
X <- Power[(TotalTrainLength - ActualTrainLength) : (TotalTrainLength -1)]
X <- array(X,dim=c(NROW(X),NCOL(X),1))

Y <- Power[(TotalTrainLength - ActualTrainLength + 1) : TotalTrainLength]
Y <- array(Y,dim=c(NROW(Y),NCOL(Y),1))

# Train model. Keep out the last 30 sequences.
model <- trainr(Y = Y, X = X, learningrate = 0.01, hidden_dim = 10, numepochs = 150)

# Predicted values, Xpred will be all training data and the test data
Xpred <- Power[(TotalTrainLength - ActualTrainLength + 1): TotalDataLength]
Xpred <- array(Xpred,dim=c(NROW(Xpred),NCOL(Xpred),1))

# use Xpred as the input to make prediction
Ypred <- predictr(model, Xpred)
```

Building training model

Calculate the predicted wind power

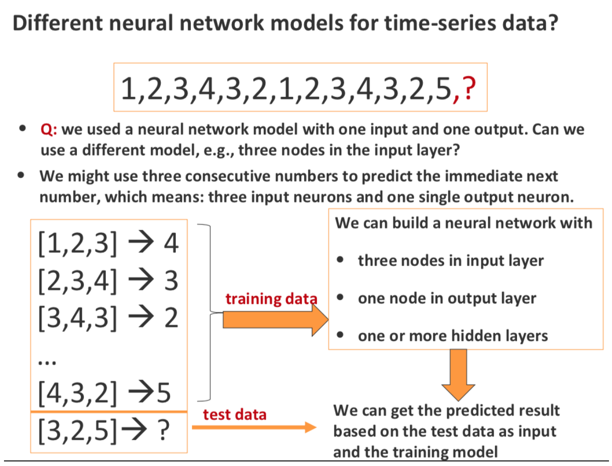
Real & Predicted Wind Power (RMSE = 0.09)

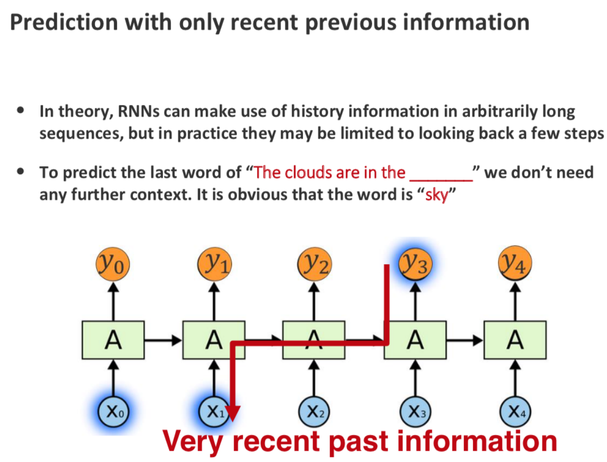


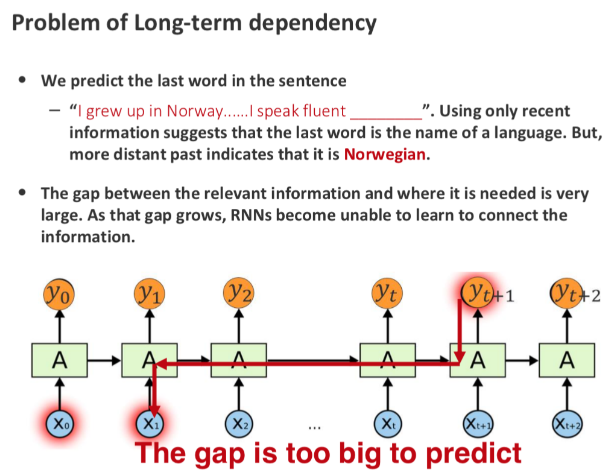
11.3 More Considerations

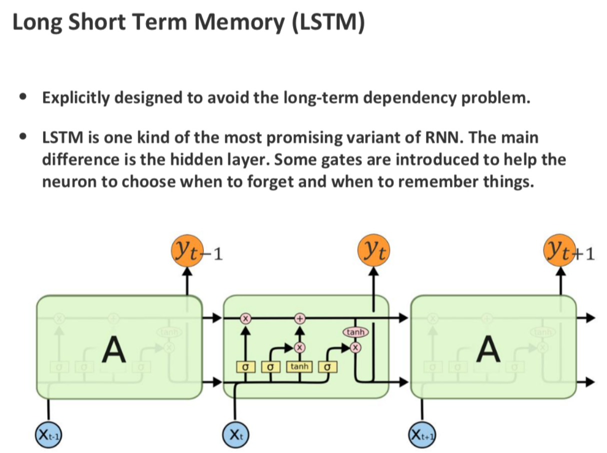
11.3.1 TensorFlow

- TensorFlow: An open-source software library for Machine Intelligence, developed by Google. The primary language in which TensorFlow machine learning models are created and trained is Python.
- It supports Python, R, and (experimental API) for java
- Tensor: The central unit of data in TensorFlow. A tensor consists of a set of primitive values shaped into an array of any number of dimensions.
- Nodes in the graph represent mathematical operations, while the graph edges represent the multidimensional data arrays (tensors) communicated between them.
- More information: https://www.tensorflow.org/









12 Research Directions in Energy Informatics

Our Energy Informatics Agenda: Putting Bits to Energy (Guest Lecture)

12.1 What is Energy Informatics

- Emerging **interdisciplinary area**
  - Interdisciplinary research involving among others electrical, mechanical, civil and computer engineering
  - Acquisition of new competences (energy conversion, power systems, etc.)
- Two perspectives
  - Develop systems that **manage energy more sustainably**
  - Develop more sustainable computer systems
- Evaluate systems based on **realistic models and data**

12.1.1 Why is Data an Issue?

- Real grid data sets are **not** openly available
- If available, data is incomplete or inaccurate
- Gathering and maintaining this data is expensive