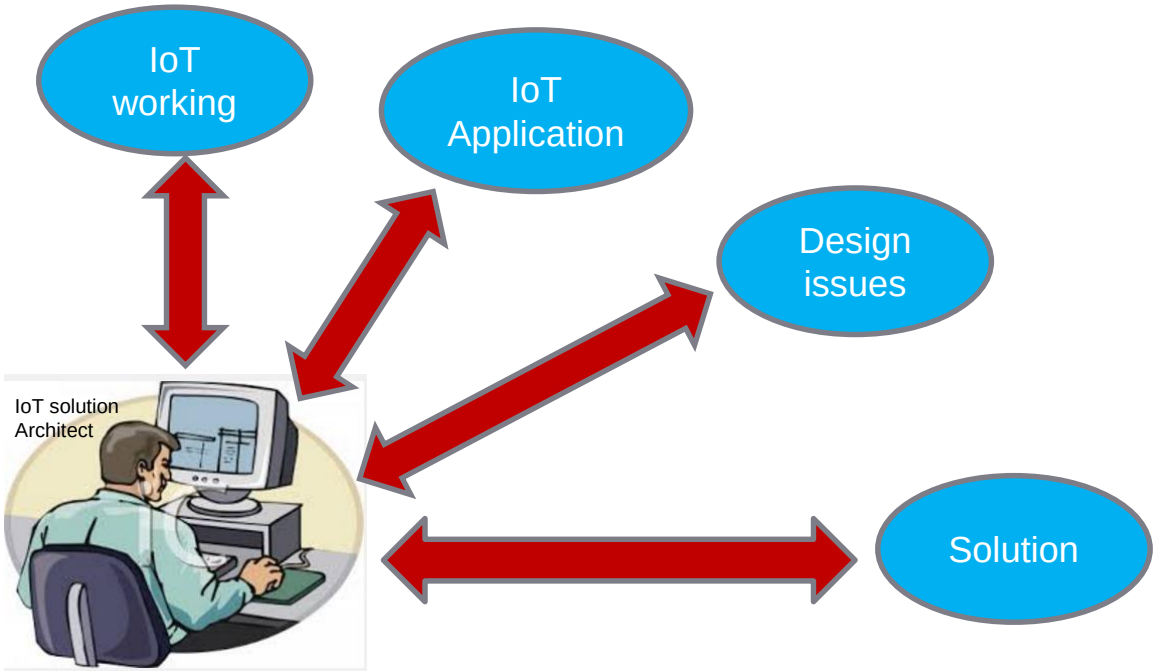
Emerging trends in the design of cognitive IoTs over embedded systems

Dr.Manjunath Ramachandra

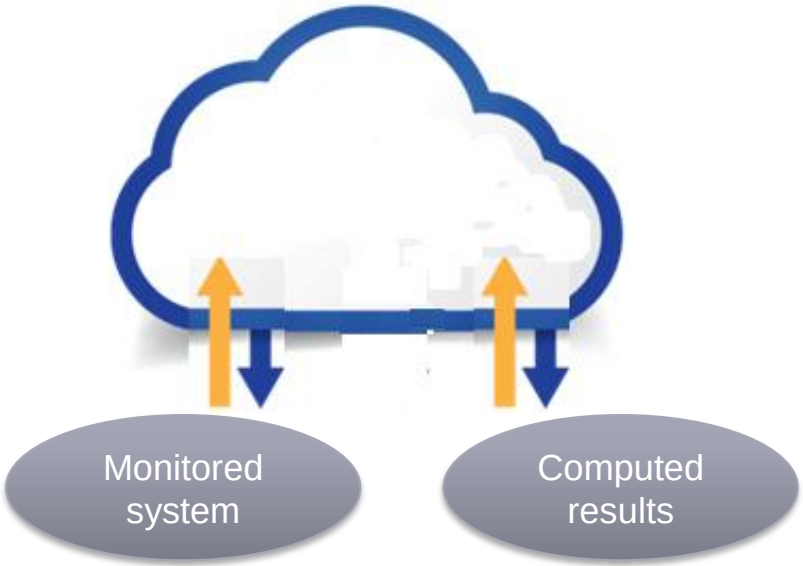
Objectives

- A fresh look at the IoT
- The issues
- Prediction based solution
- case study

Agenda

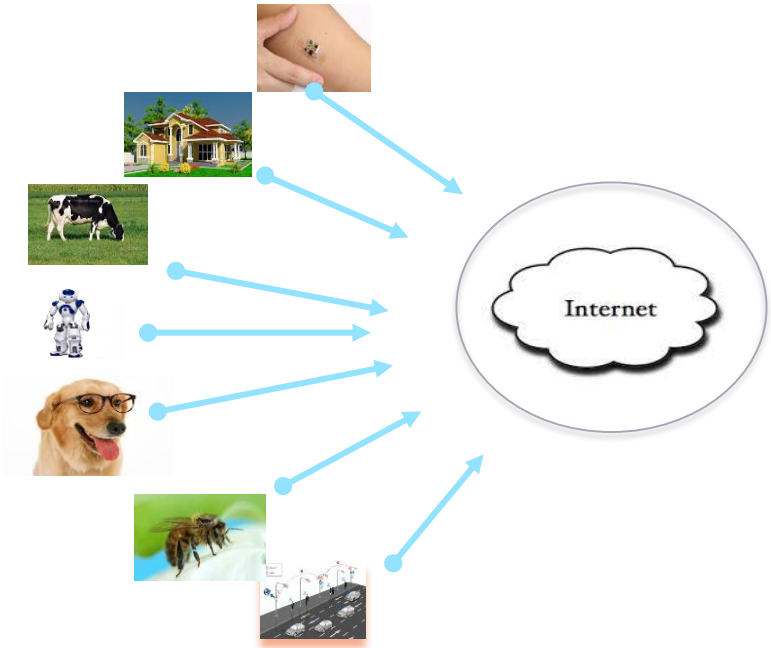

IoT
working
IoT
Application
Design
issues
IoT solution
Architect
Solution

IoT architecture

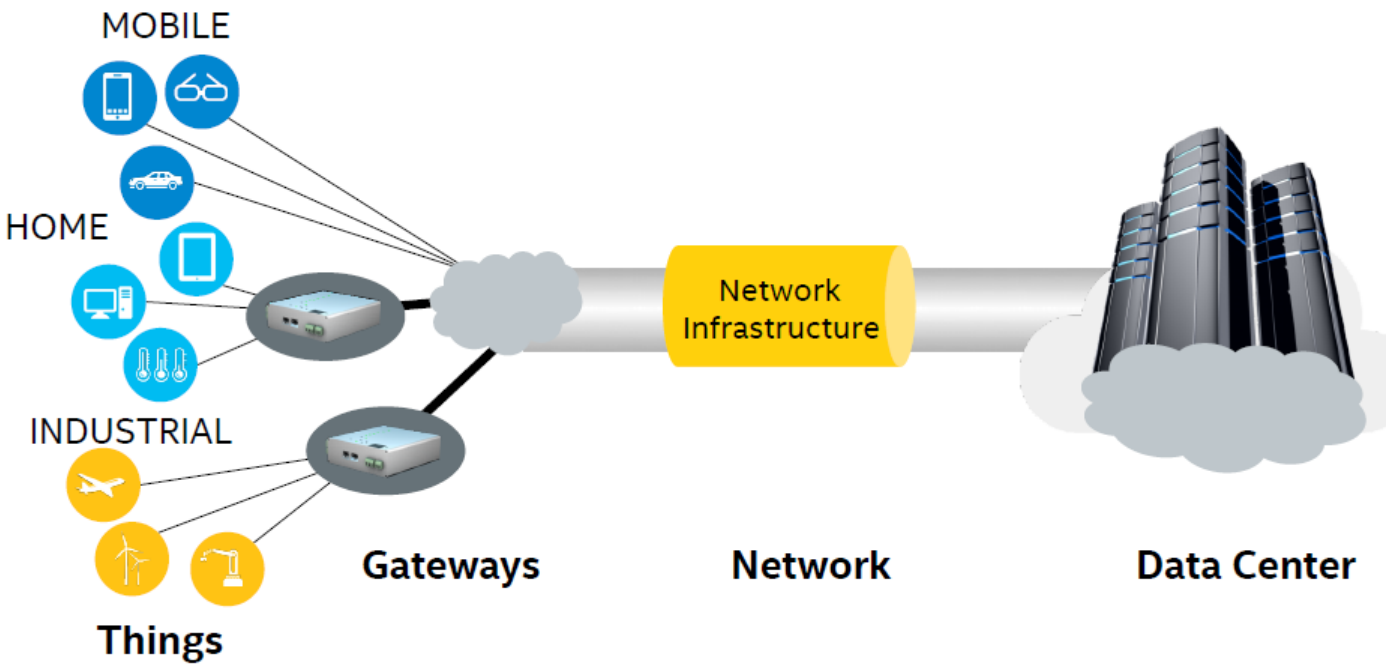


Ack: B B Smart work

The IoT: Brain dump

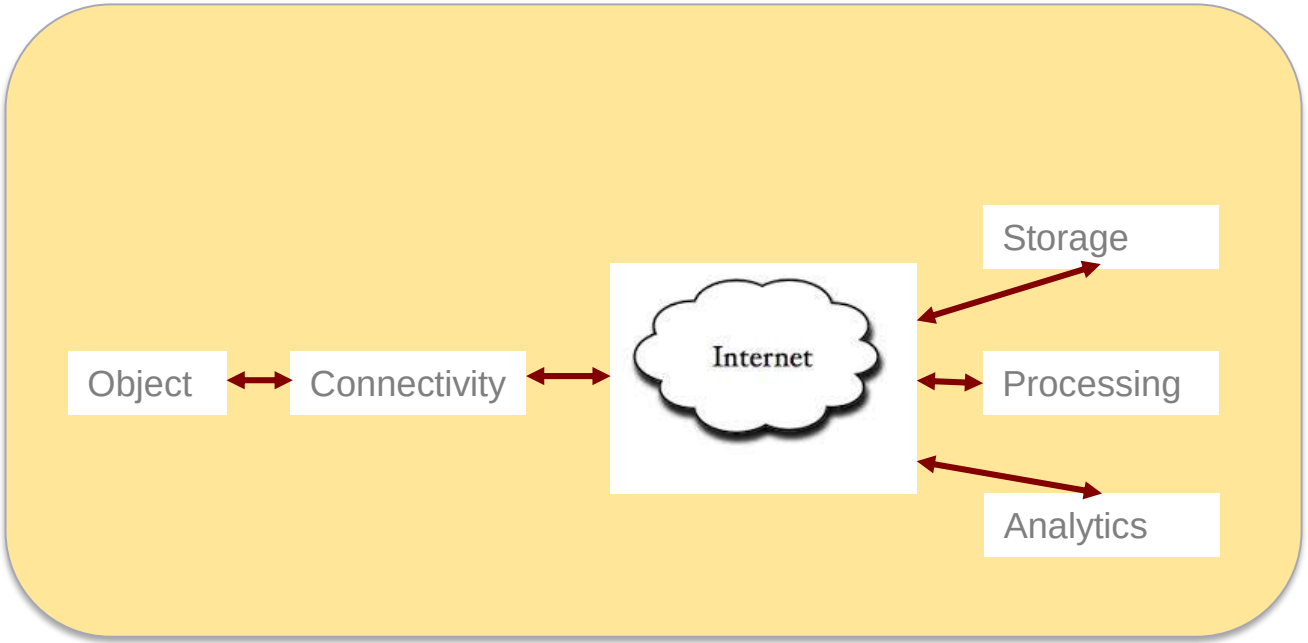


IoT components

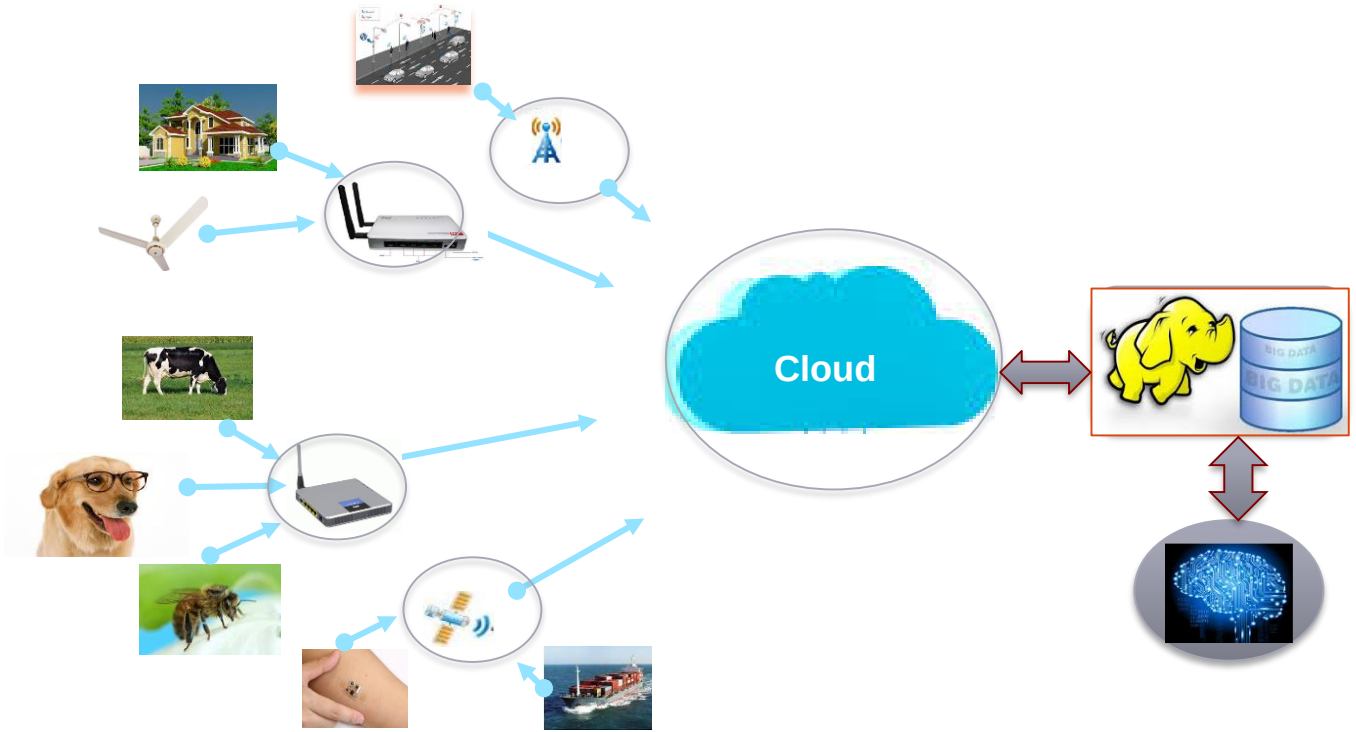


Ack:intel

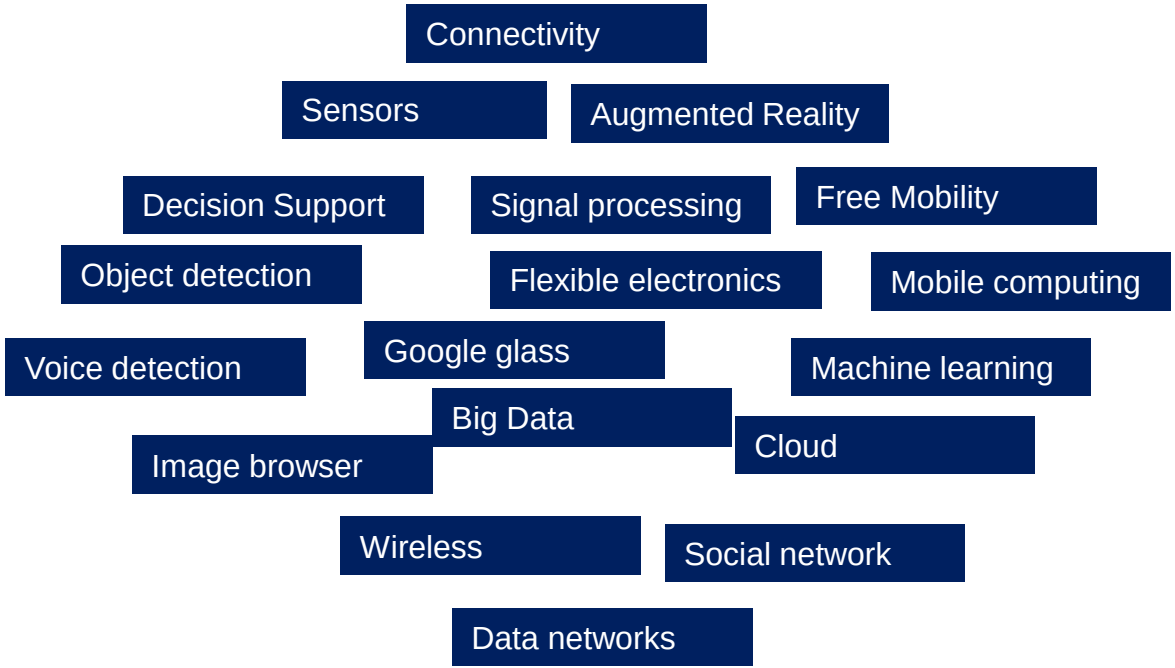
IoT components



The IoT solution : Component diversity

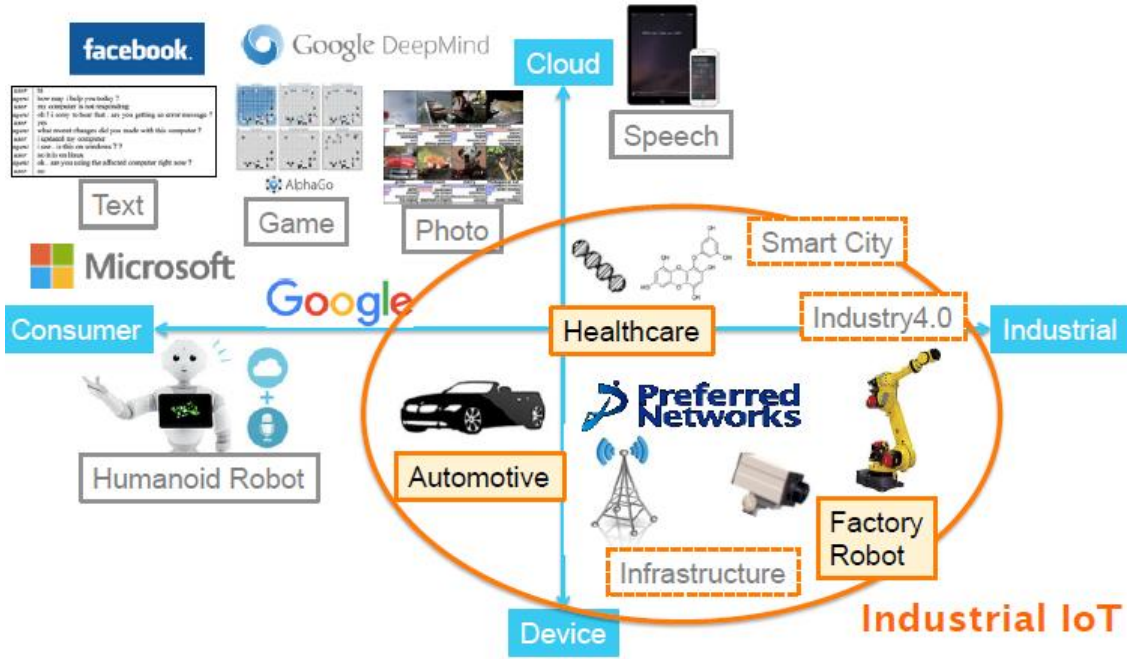


IOT Solution :Technologies Diversity

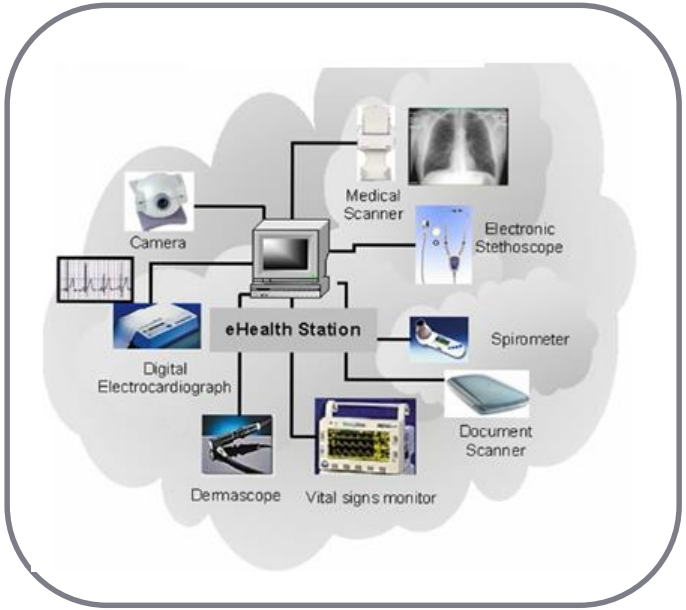


IoT solution: Application diversity

Industrial IoT



E health & M health

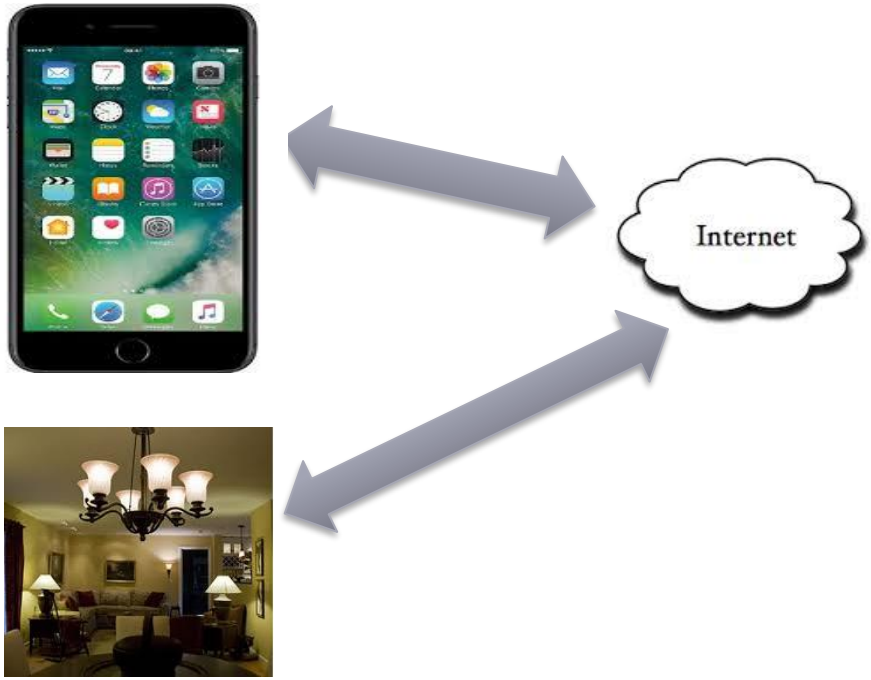


- global e Health products and services market stands at $130 billion
- grows to $160 billion by 2015.

- More than 7 Billion mobile network subscriptions worldwide
- m Health market account for $9 Billion in 2014, grows to $23 billion by 2017
- m Health could save 99 billion EUR in healthcare costs in the EU

IoT based solution

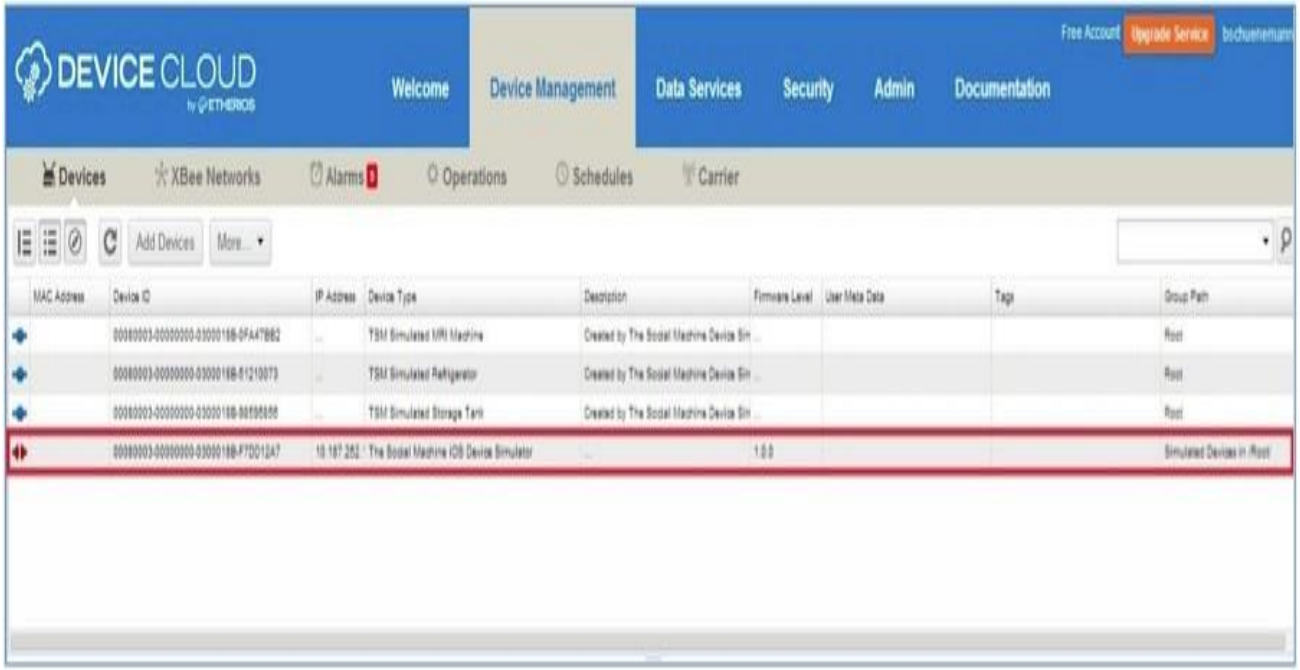
Step 0: Application Selection



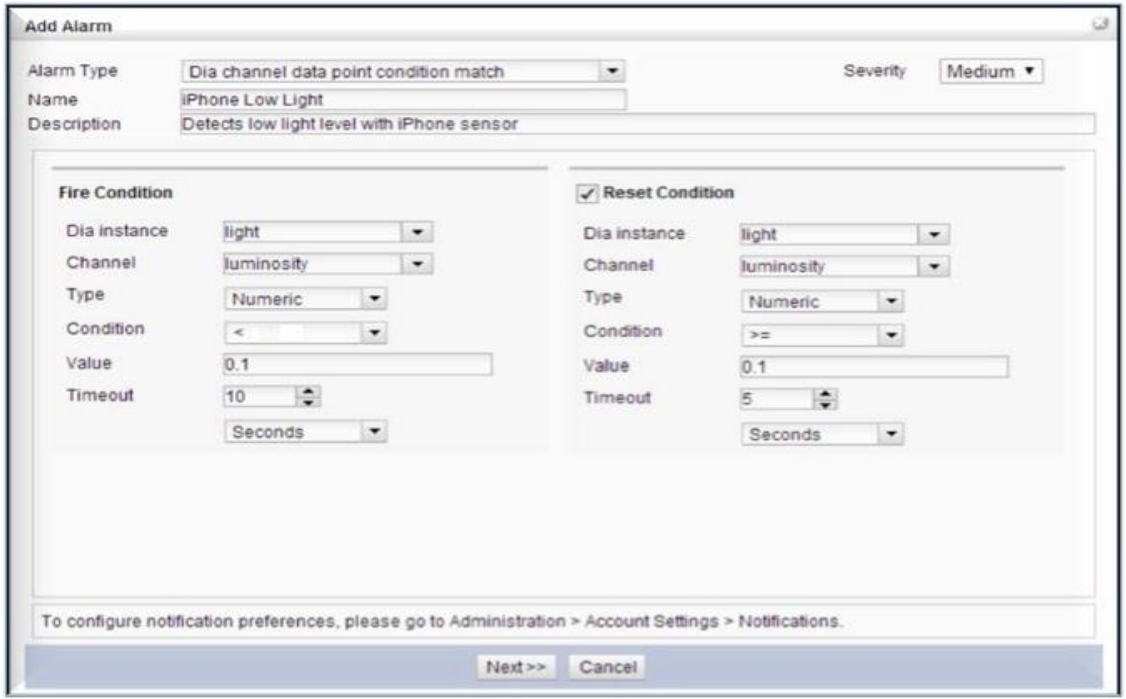
Step 1: authentication

Ref: Reid Carlberg

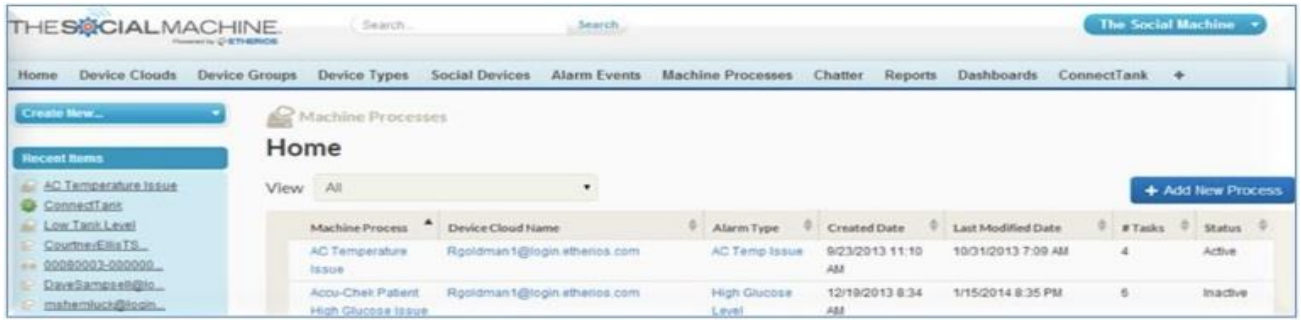
Step 2: Device identification



Step 3: Event set up

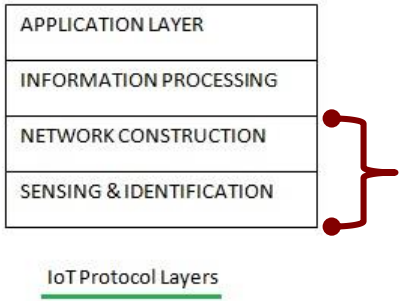


Step 4: Response set up

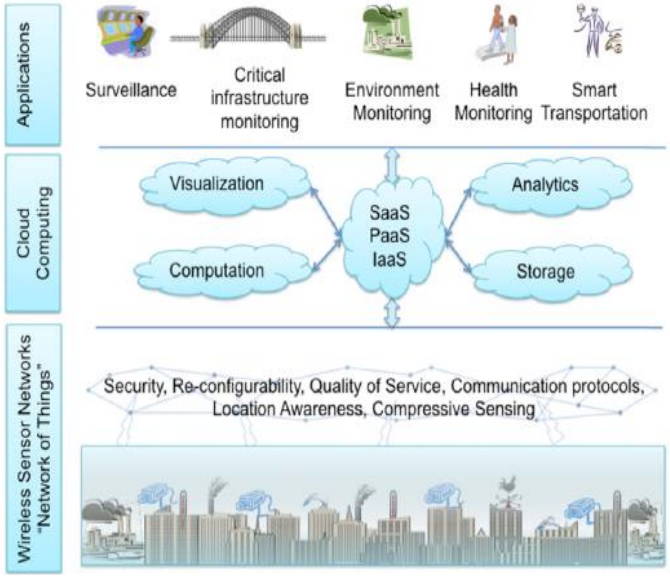


IOT functional architecture

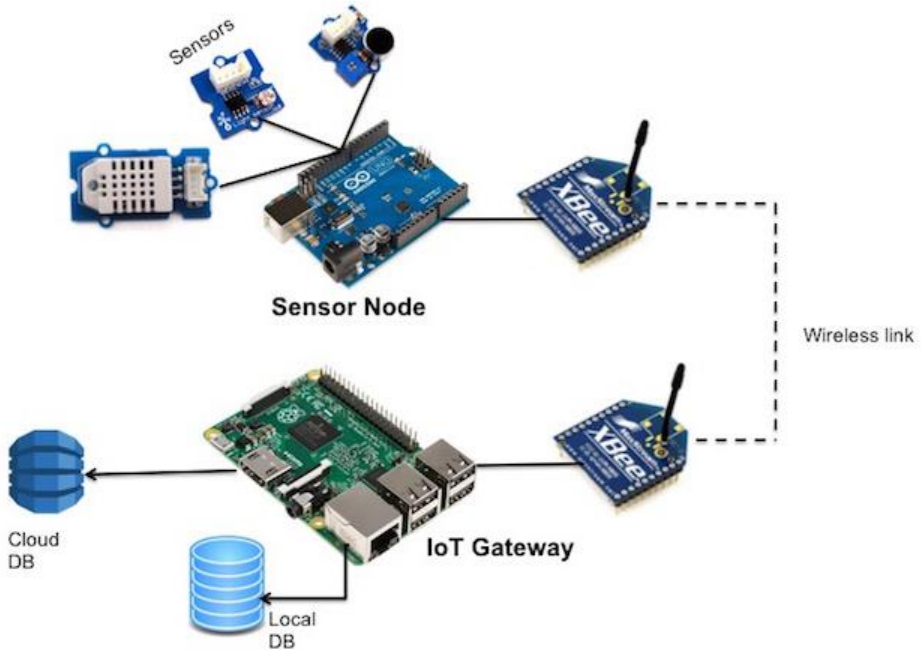
IOT protocol layers



Middleware



The Hardware



IoT hardware

IO devices

- Sensors
- actuators:
- LED
- Relays
- Motors
- Linear actuators
- Lasers
- Solenoids
- Speakers
- LCD
- Plasma displays
- Robots

Network devices

- Modem
- Gateway
- Router
- Satellite
- Tower

Processing devices

- Grid
- Cloud
- Embedded processor
- Quantum computer
- PC\Laptop

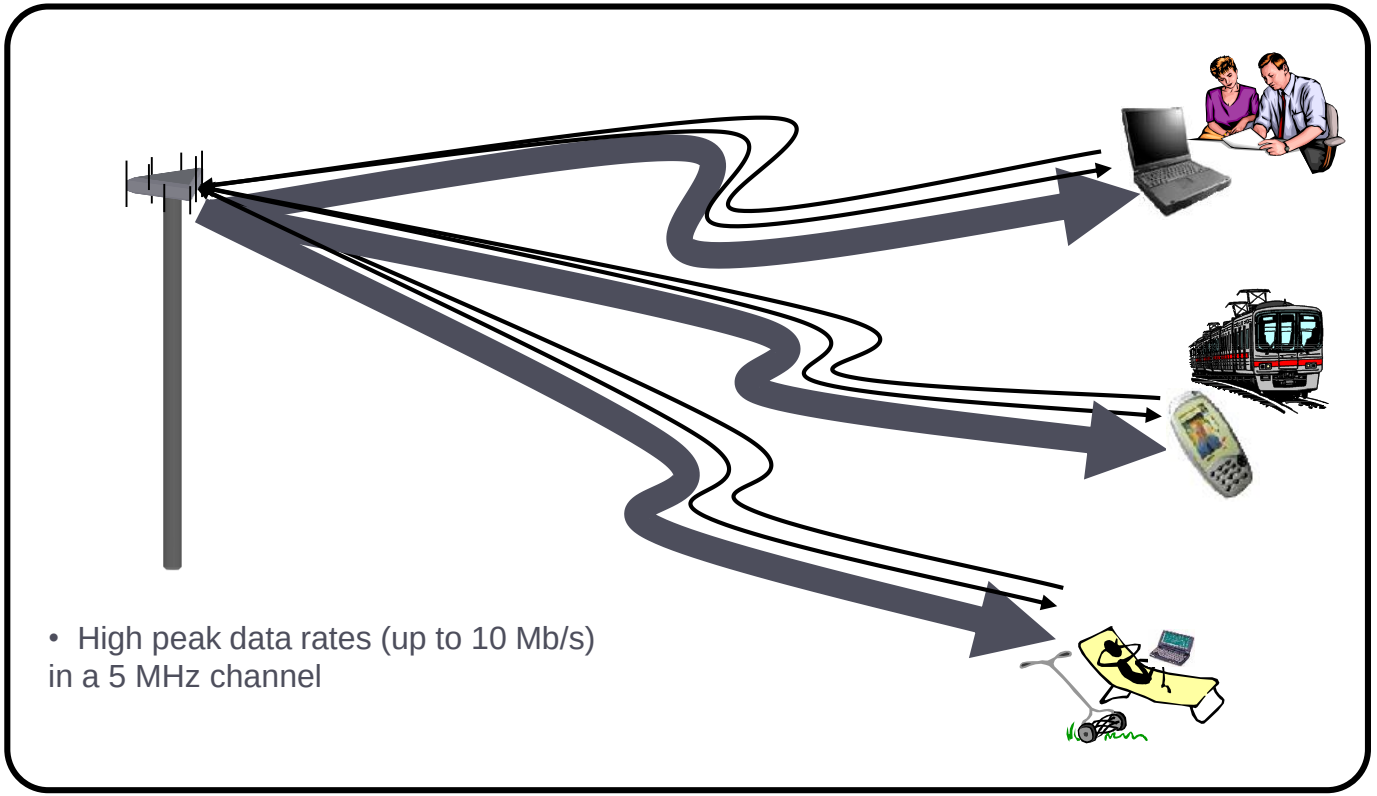
Memory devices

- Cloud memory
- Flash storage
- Quantum memory

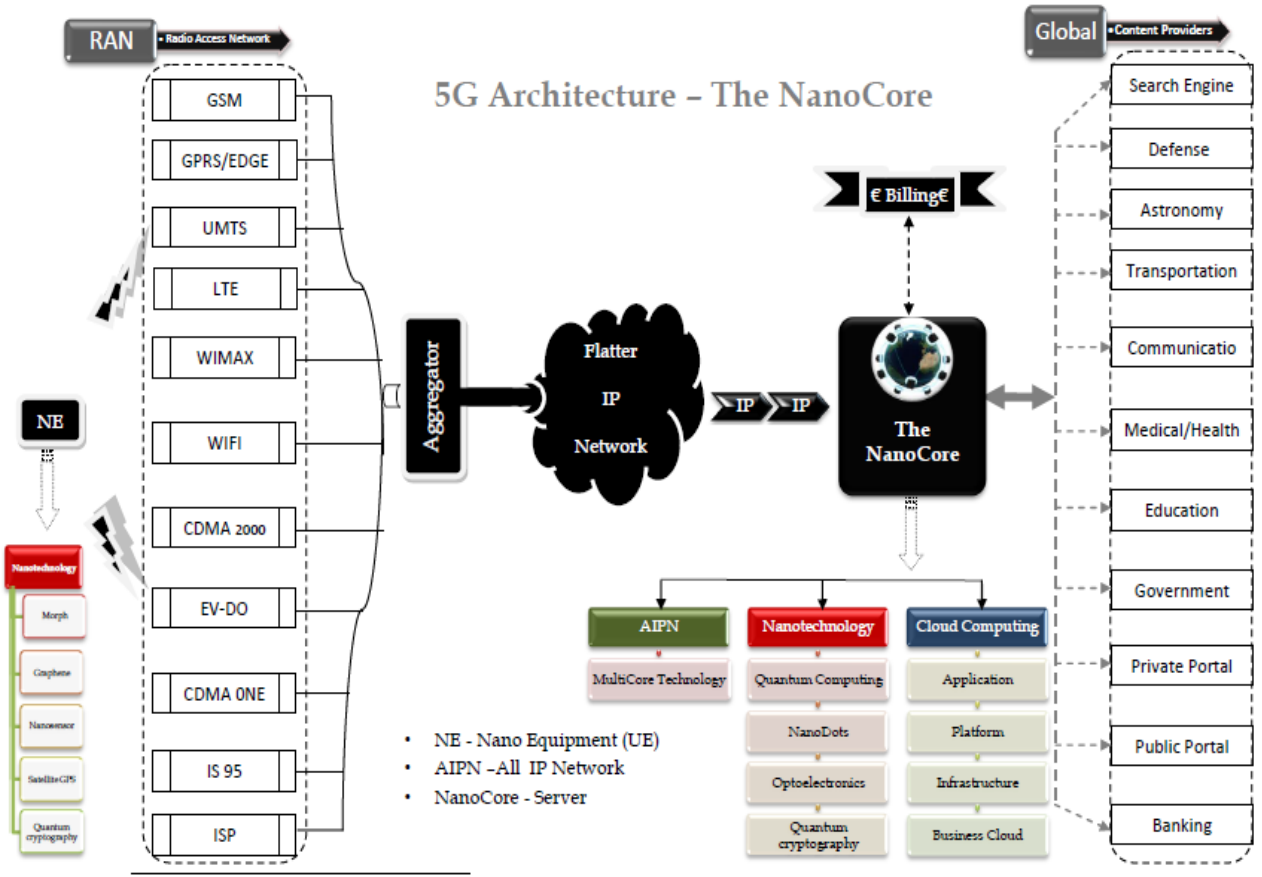
Device Connectivity Protocols

- WiFi
- Bluetooth
- RFID
- ZIGBEE
- NFC
- Ethernet
- LTE
- 3G
- GSM
- CDMA

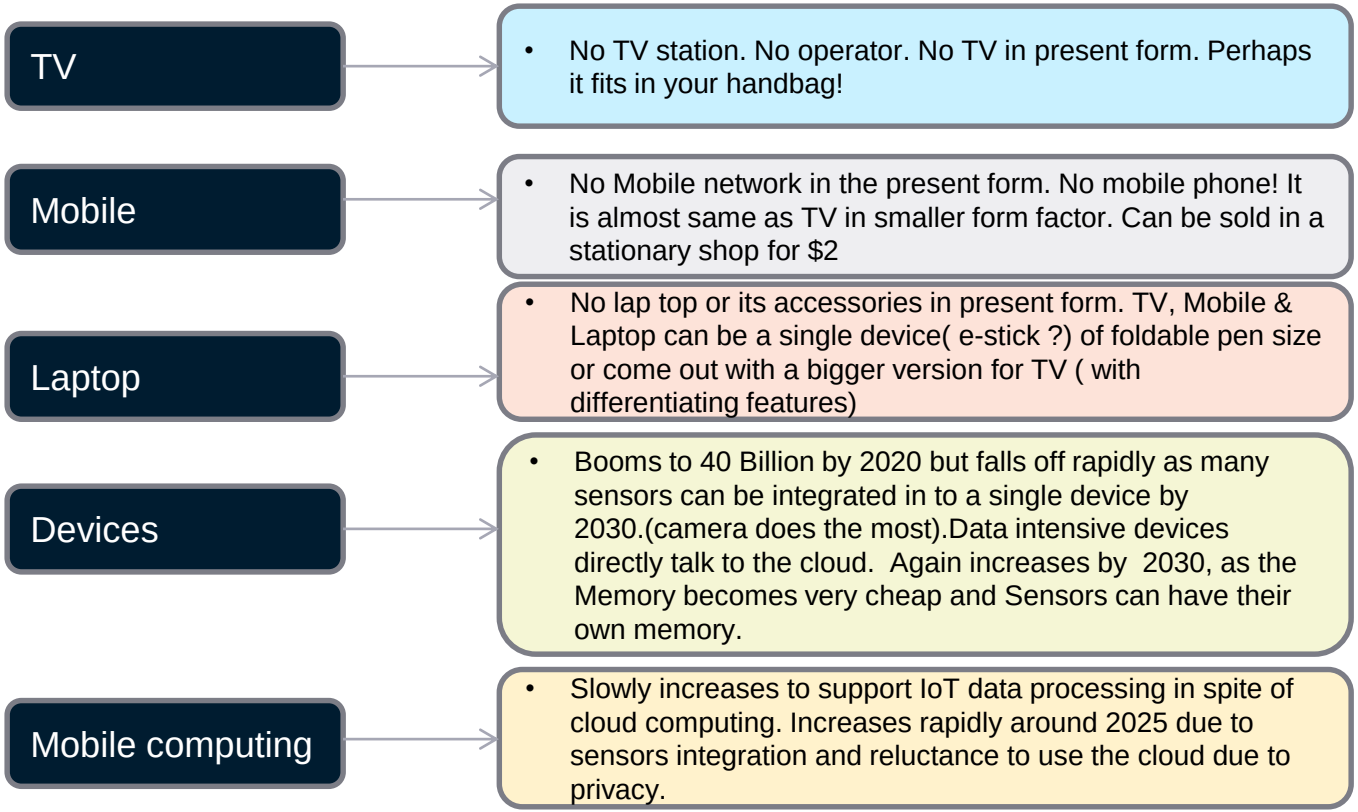
4G Wireless



- High peak data rates (up to 10 Mb/s) in a 5 MHz channel


5G Architecture – The NanoCore
RAN
• Radio Access Network
GSM
GPRS/EDGE
UMTS
LTE
WIMAX
WIFI
CDMA 2000
EV-DO
CDMA ONE
IS 95
ISP
NE
Nanotechnology
Morph
Graphene
Nanowear
SatelliteGPS
Quantum cryptography
Aggregator
Flatter IP Network
IP
IP
€ Billing€
The NanoCore
AIPN
MultiCore Technology
Nanotechnology
Quantum Computing
NanoDots
Optoelectronics
Quantum cryptography
Cloud Computing
Application
Platform
Infrastructure
Business Cloud
Global
• Content Providers
Search Engine
Defense
Astronomy
Transportation
Communicatio
Medical/Health
Education
Government
Private Portal
Public Portal
Banking
• NE - Nano Equipment (UE)
• AIPN –All IP Network
• NanoCore - Server

Interesting Projections



Author's perceptions

Connectivity: example

	Home gateway	3G/4G	Satellite
Google glass	√	√	√
Temperature sensor	….		
camera			
Water/liquid gauge			
TV			
Mobile phone			

IoT software

Requirements of WSN (software)

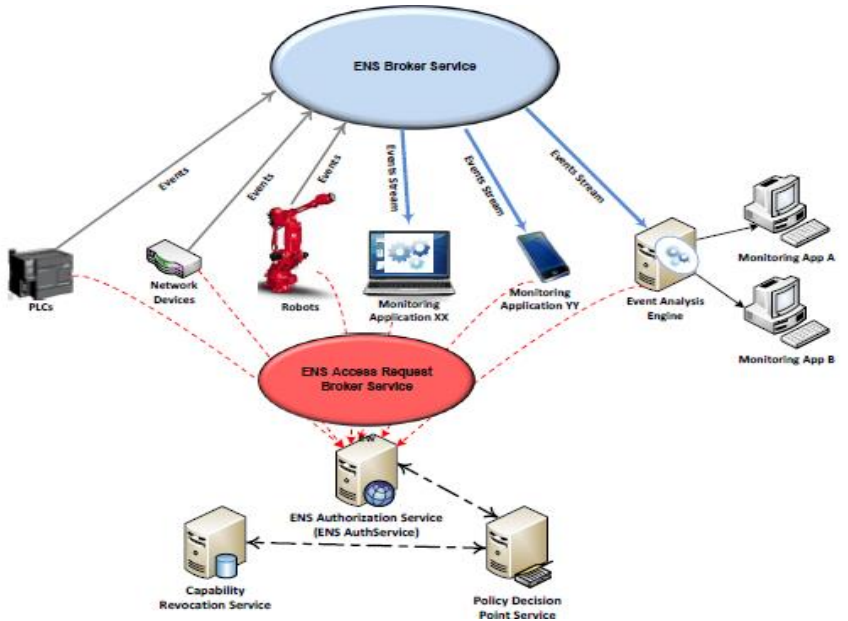
- Smart and autonomous
- Auto-configuration
- Self monitoring and self-healing
- Anomaly detection and tracking

29

Communication protocols of IoT

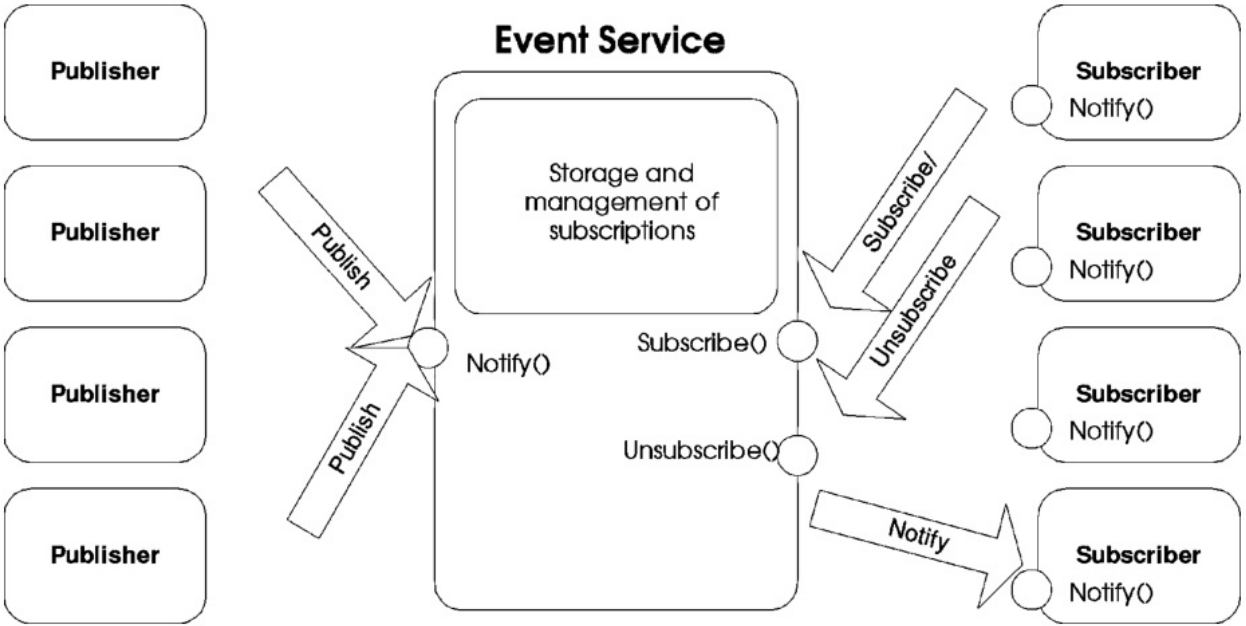
PROTOCOLS	MQTT	CoAP	XMPP	SOAP	UPnP
XML Based	No	No	Communications protocol for message-oriented middleware based on XML	Relies on XML information set for its message format	No
TCP/UDP	TCP	UDP	TCP	Both TCP and UDP	UDP
IPv4/IPv6	IPv4,IPv6	IPv4,IPv6	IPv4,IPv6	IPv4	IPv4,IPv6
M2M support	OASIS Standard	ETSI Standard	---	---	---
USAGE	From Pervasive devices to a server/small message broker.	Simple electronic devices, Resource constrained devices	Video, File transfer, gaming, IoT apps such as smart grid & social networking services	Implementation of web services In computer networks	Permits networked devices to seamlessly discover each other presence on network

Event notification service of IoT @work



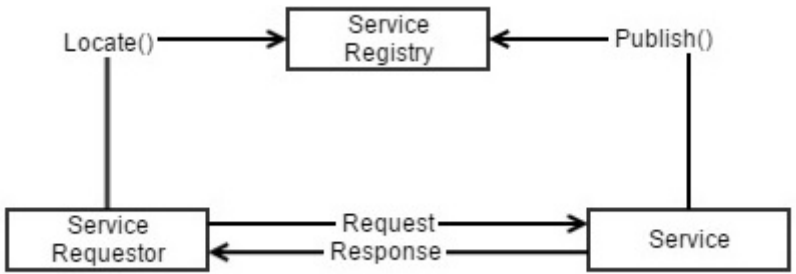
- ENS acts as a common collector and distributor of events
- Events come from multiple heterogeneous sources
- Devices can subscribe to specific subsets of events

Publish-Subscribe software architectural style



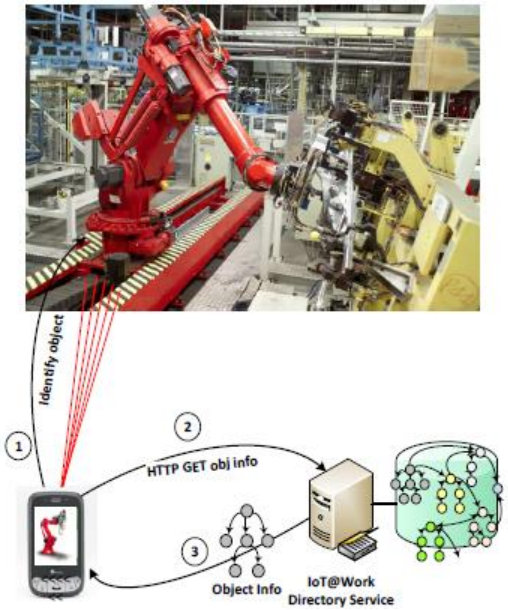
Middleware

- Lies between the technology and the application.
- SOA (Service Oriented Architecture) middleware



33

Object as a (web)service

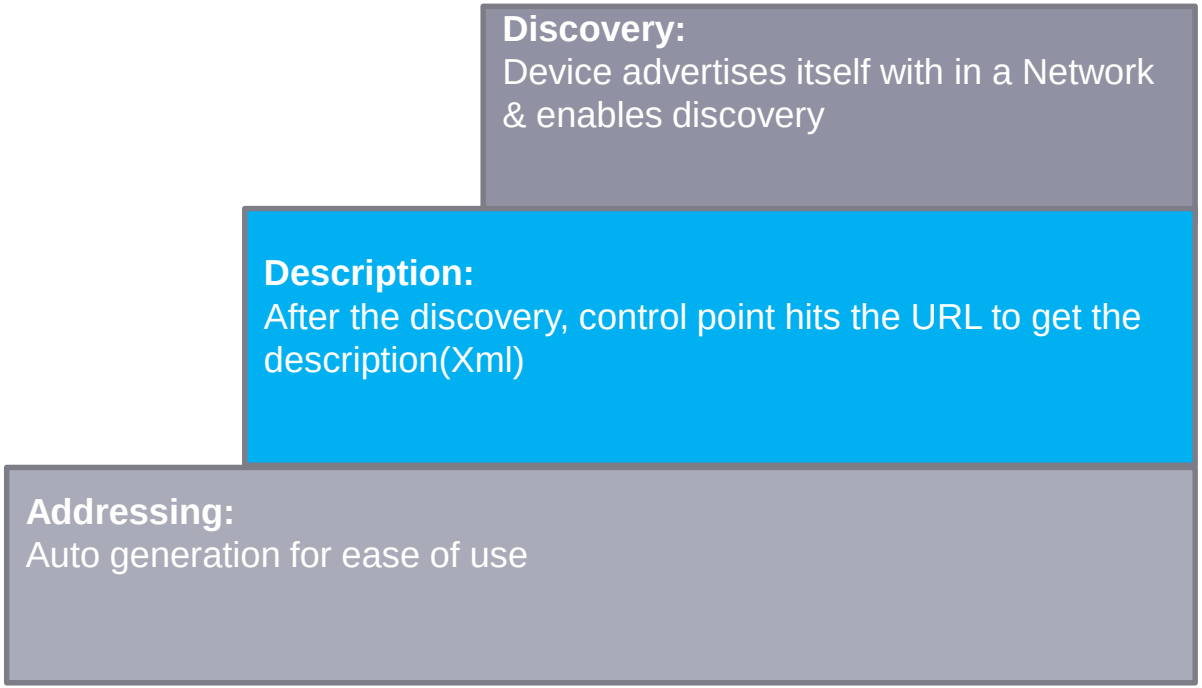


Service Management

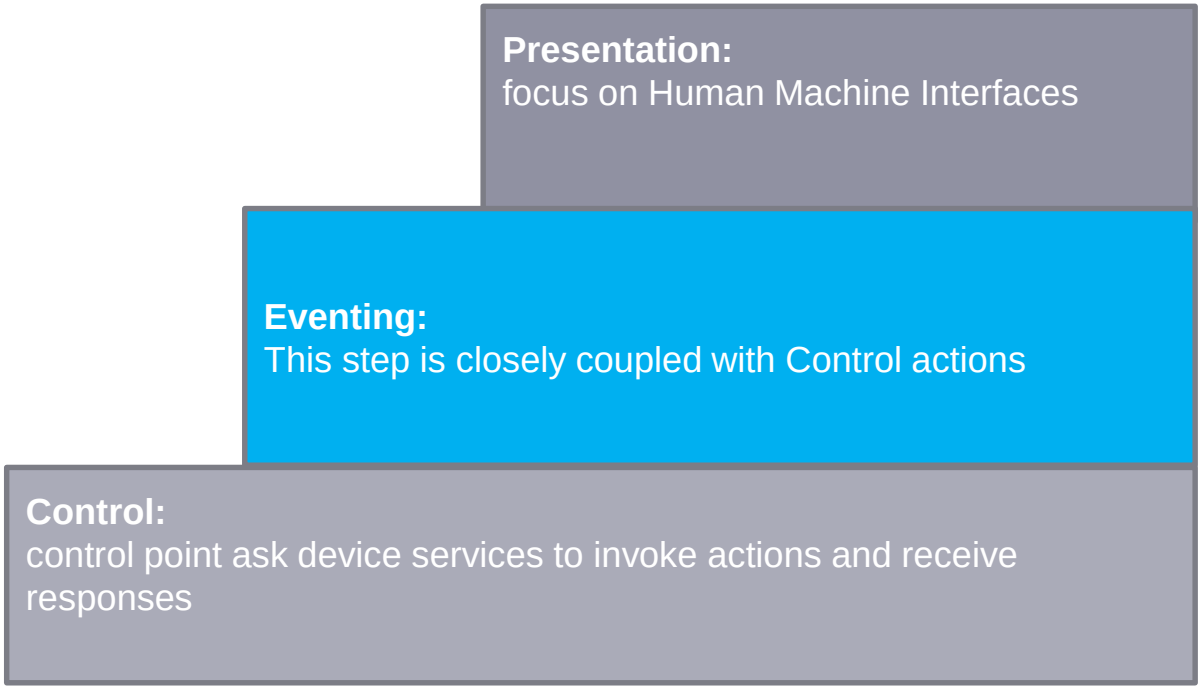
- Basic set of services :
  - Object dynamic discovery
  - Status monitoring
  - Service configuration
- Functionalities related to QoS

35

Upnp+ Protocol for IoT



Upnp+ Protocol for IoT



The Subscription Setup Protocol

- Subscription Setup :

  Role Publisher, Subscriber

  Subscriber -> Publisher : request [out ID , out metadata]

  Publisher-> Subscriber : accept [in ID , in metadata , out ID]

  Publisher -> Subscriber : reject [in ID , in metadata , out ID]

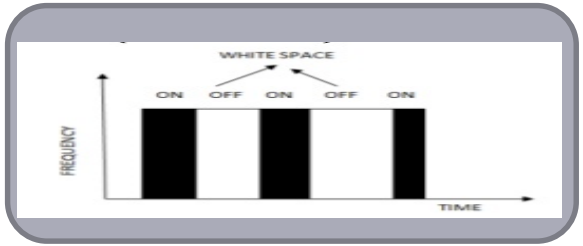
Actions on the nodes

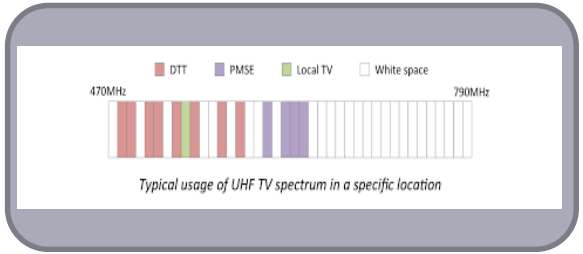
- Nodes can be:
  - Created CreateInstance()
  - Read GetValues()
  - Updated SetValues()
  - Deleted DeleteInstance()
  - Notified Alarming Feature: UPnP state variable event including the node & value of the node

Factors affecting SLA over the IoT

- Variations in platform of application sw (speed of execution)
- connectivity
- Granularity of software (dictate inter-process communication overhead)
- Database organization (hierarchy and granularity)
- Data retrieve mechanism and policies
- QoS protocol
- Sla policies
- Client Platform
- middle ware

TV white space





- Bursty requests
- Prioritized service
- Flooding

Ack: Nominet

Agenda

Objective

The IoT

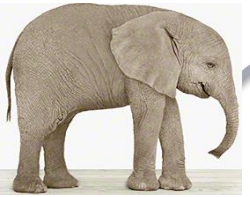
3 Issues

4 Solution

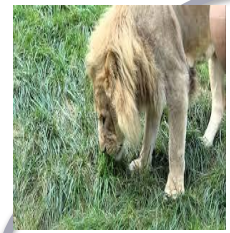
5 Case study

6 Conclusion

Features of the IoT



Large distributed system



Heterogeneous services

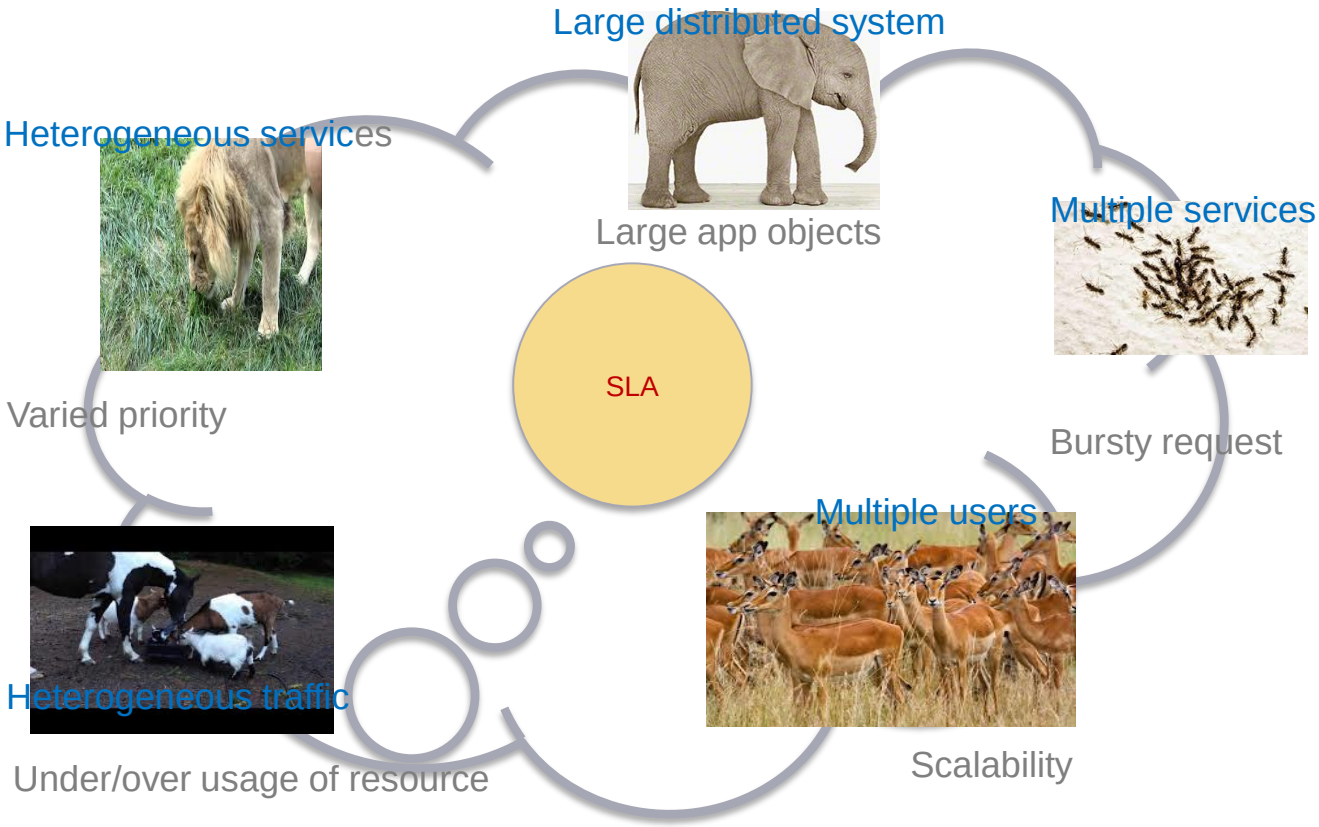
Multiple services

SLA

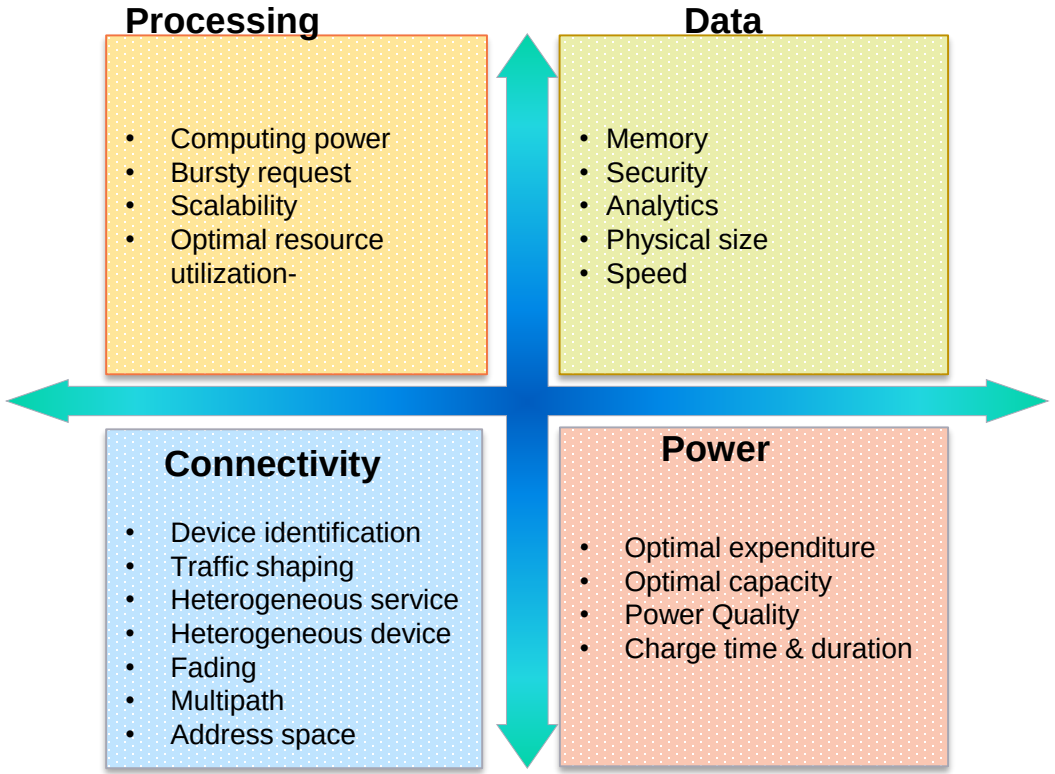
Heterogeneous traffic

Multiple users

Issues


Large distributed system
Heterogeneous services
Large app objects
Multiple services
SLA
Varied priority
Bursty request
Multiple users
Heterogeneous traffic
Under/over usage of resource
Scalability

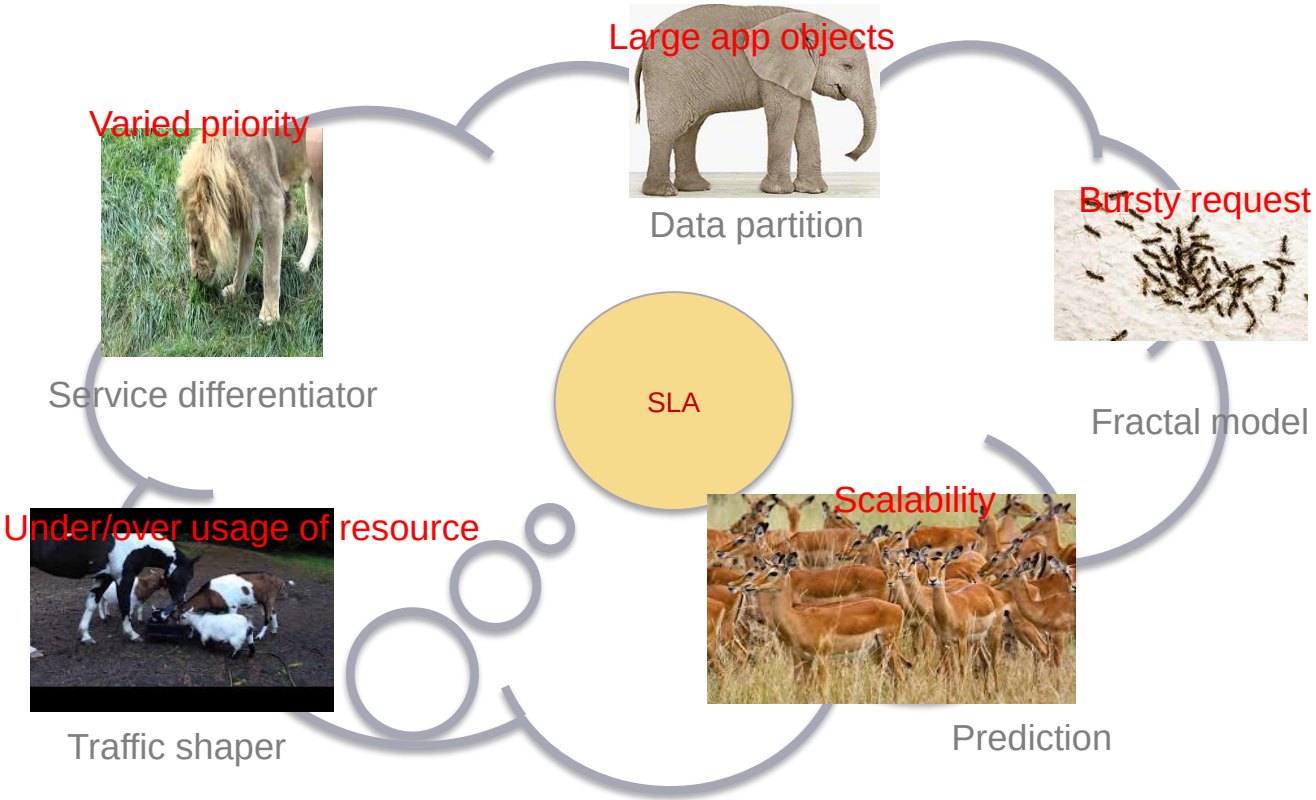
The issues


Processing
Computing power
Bursty request
Scalability
Optimal resource utilization-
Data
Memory
Security
Analytics
Physical size
Speed
Connectivity
Device identification
Traffic shaping
Heterogeneous service
Heterogeneous device
Fading
Multipath
Address space
Power
Optimal expenditure
Optimal capacity
Power Quality
Charge time & duration

Agenda

1. Objective
2. The IoT
3. Issues
4. Solution
5. Case study
6. Conclusion

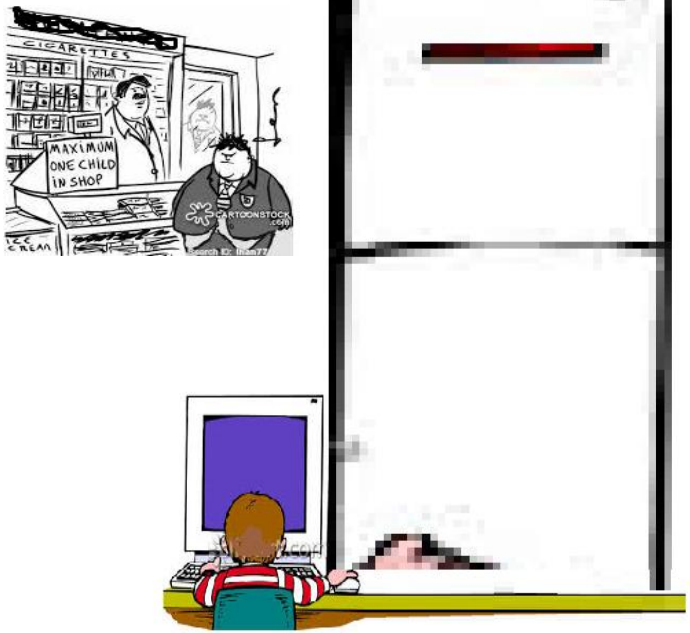
solution


Large app objects
Data partition
Varied priority
Service differentiator
Bursty request
Fractal model
SLA
Scalability
Prediction
Under/over usage of resource
Traffic shaper

Processing issues

Processing issue: Computing power

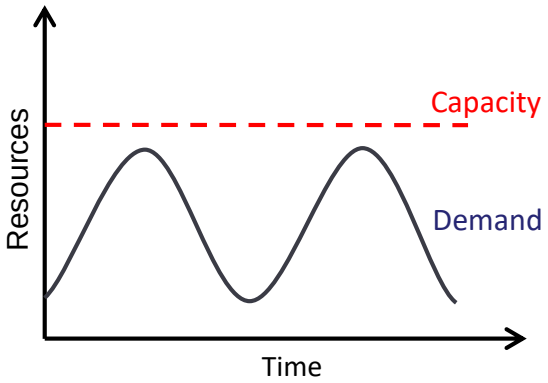
Cloud



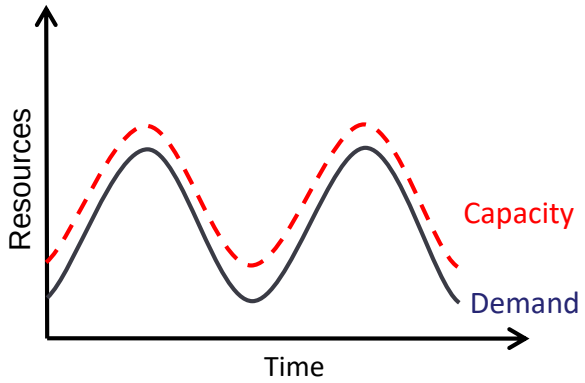
- Vast resources
- Heterogeneous platforms
- Multiple service APIs
- Multiple versions
- Multivendor tools
- Super computer!

- Buy what you want
- Buy when you want
- Buy *as much as* you want
- Jump the “Q”!

Resource utilization

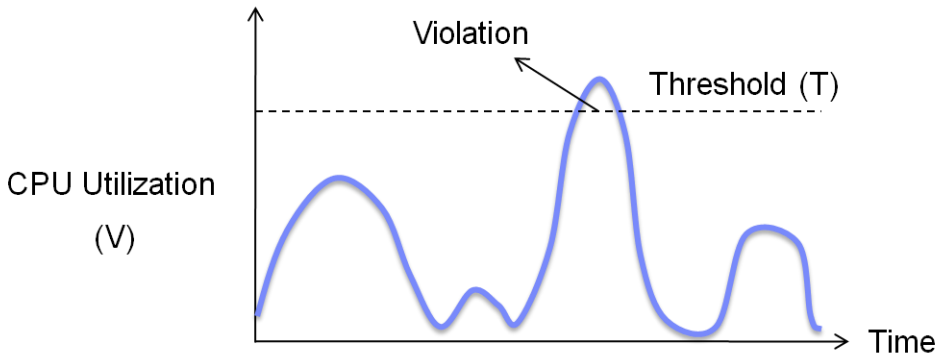


Static data center

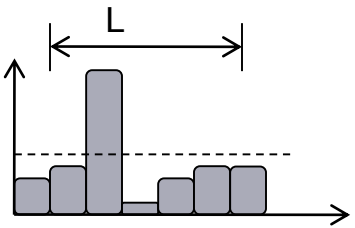


Data center in the cloud

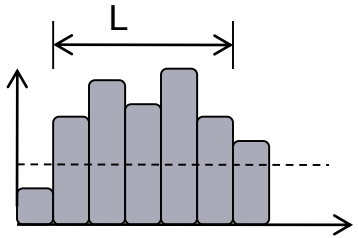
CPU usage



Demand monitoring



Short-term burst

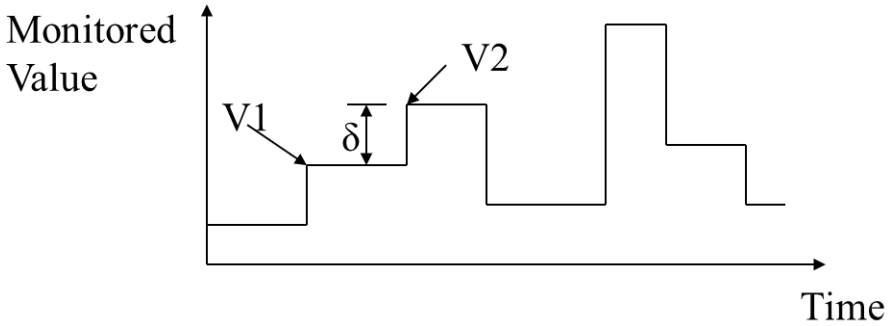


Repeated violation

Monitoring-as-a-service:Georgia tech

Sampling of the status

- Reduce sampling frequency if violation likelihood less than threshold

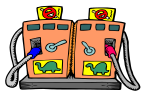


Processing Issue: Bursty request

Connected world heterogeneity





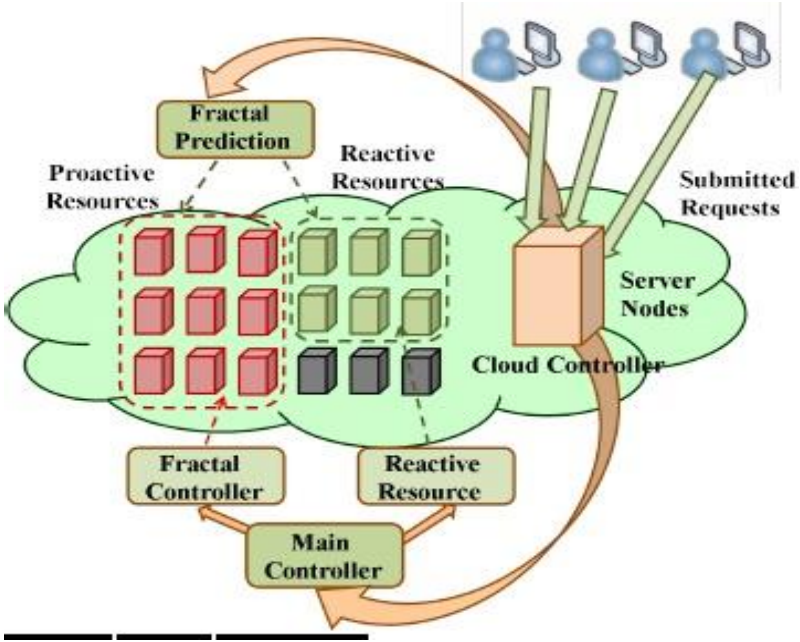

Load pattern

- “Bursty” request patterns especially when high levels of peak utilization.
- The workloads are varied and non-uniform in nature.
- under bursty workload conditions, **fractal model** can accurately predict bursty request processes
- the fractal model based optimization works better by an average of 30% for resource utilization , power reduction and job completion time.

Bursty traffic

- Fractal traffic in Hadoop:
  - In HDFS during input file loading and result file writing, high amount of data replication across cluster nodes make traffic bursty
  - In data shuffling phase of MapReduce multiple mapper nodes terminate and send their results to reducer nodes with Many to one traffic.

Handling of bursty traffic

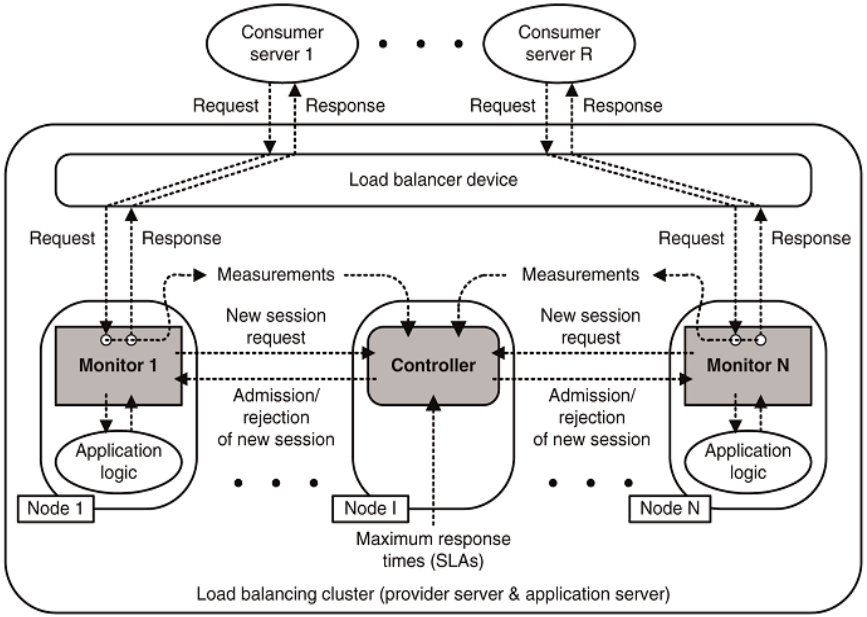


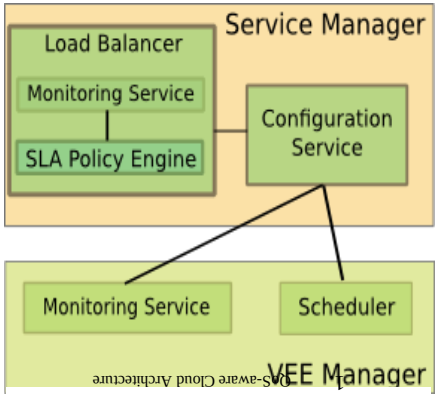
Handling the surges

- Flood control:
  - Static server provisioning (dam building) for large time scales and cross-IDC load shifting (water balancing) for small time scales
- load balancers
  - Each client has the visibility of each single resource allocated to it. Balance the load of the application among the resources of the platform. Prevent the occurrence of resource overload and handlre runtime failure of resources
- Add more web server instances, database replications
  - Return after flooding or put to power save mode
- set up web site health monitoring
  - Traffic sensing detect specific resource conditions, such as resource response times and pool membership configuration
- Fractal traffic predictor and controller

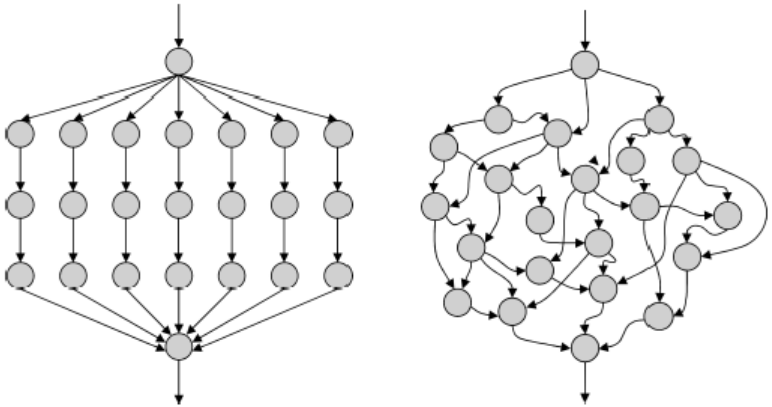
Processing Issue: scalability

Load balancing





Balanced and unbalanced workflow structure

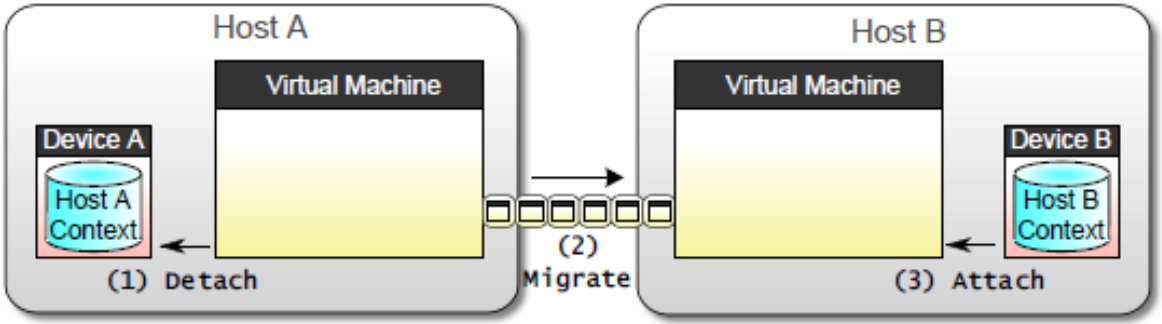


High Scalability

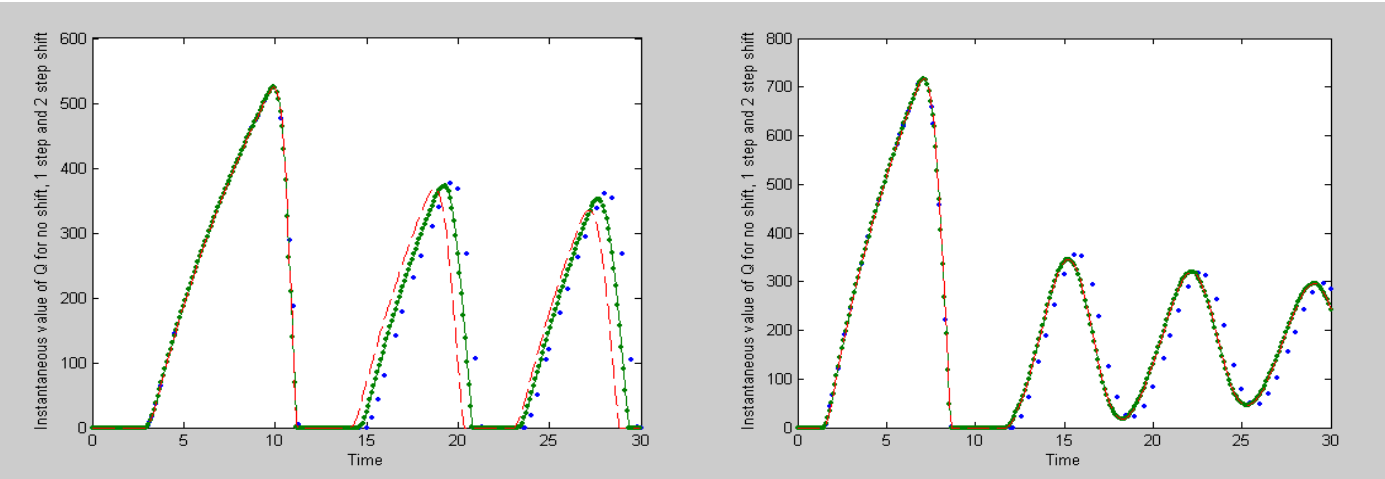
- **Service replication:**
  - cloning of services during run time to optimize the service load. Also used to support the nodes to achieve QoS especially during large service load.
  - enhances service scalability
- **Service migration:**
  - Calls for placing a running service on an alternative node when a particular node unable to meet agreed QoS .

VM migration: Resource re-allocation

- Re-contextualisation



Resource utilization with scaling of Requests



20 Simultaneous requests

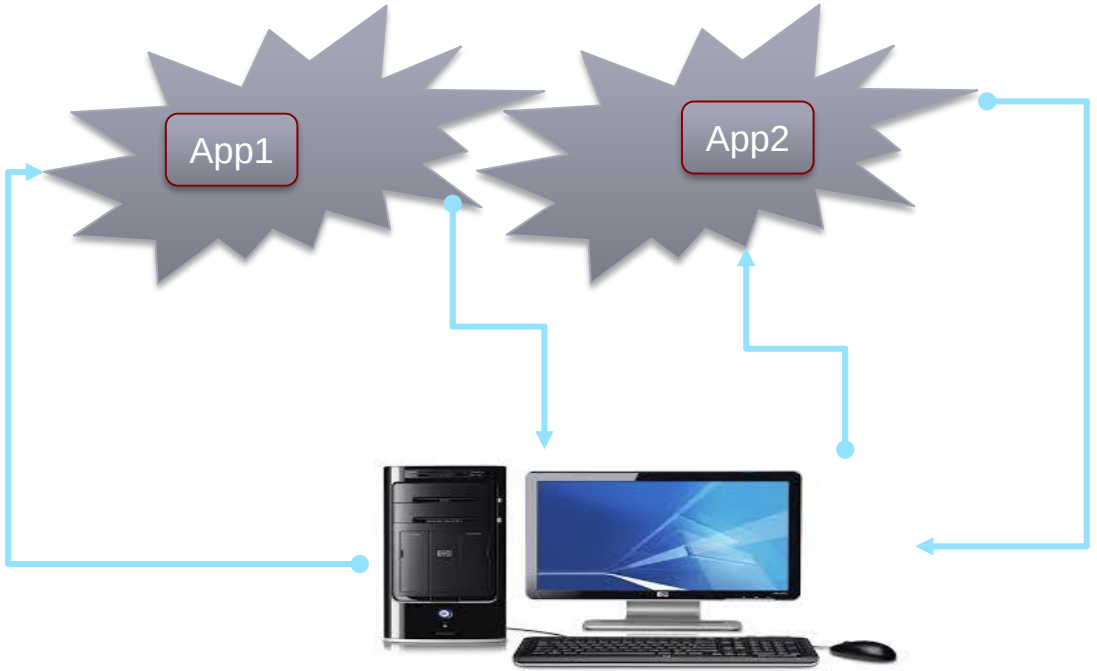
40 Simultaneous requests

Initial peak demand changes with scaling

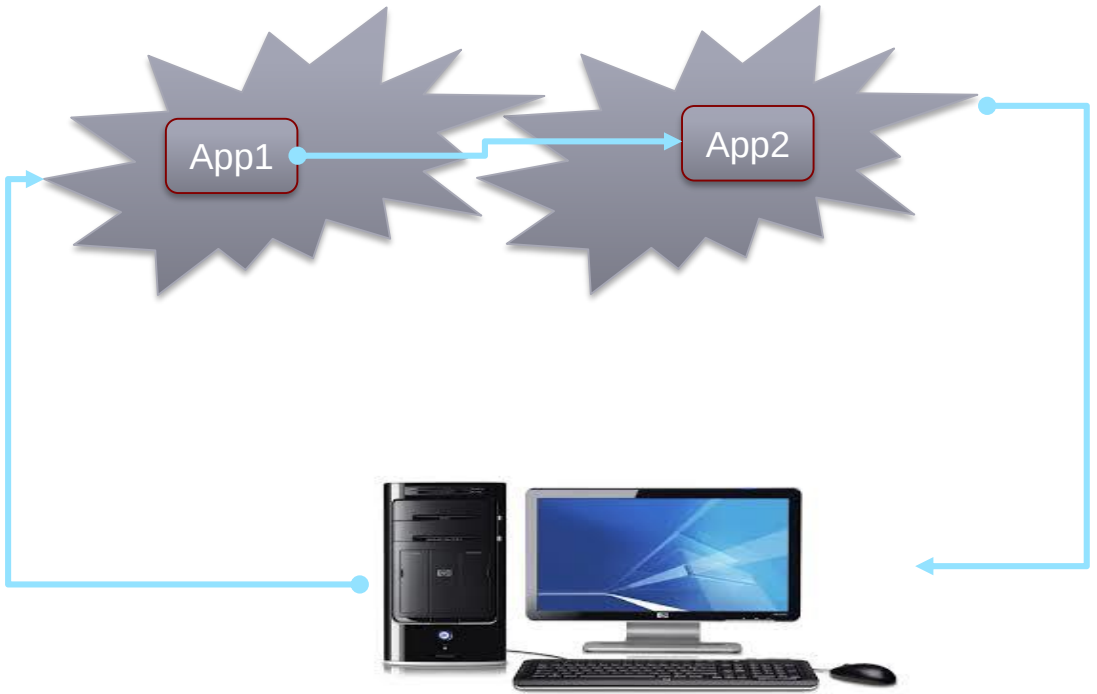
Legend: - . Instantaneous queue without shift – with shift 1 --with shift 2

Monolithic Application

Dynamic resource allocation

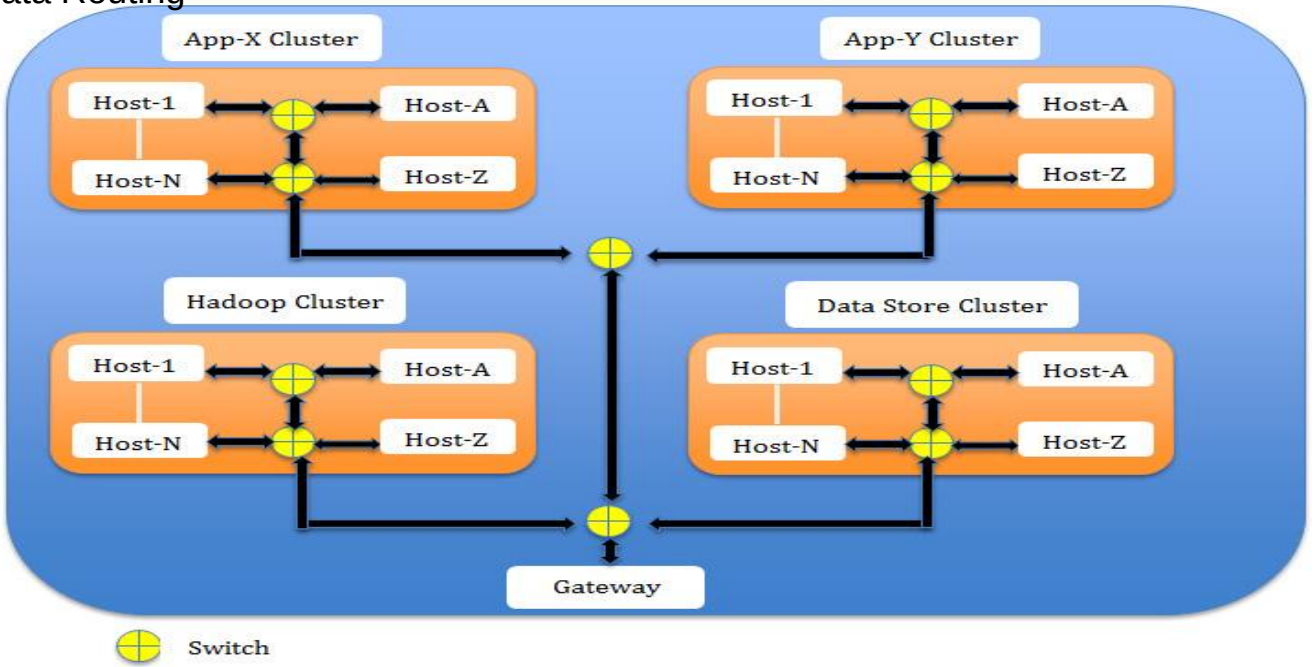


Optimal Cloud architecture



Processing issue: Optimal resource utilization

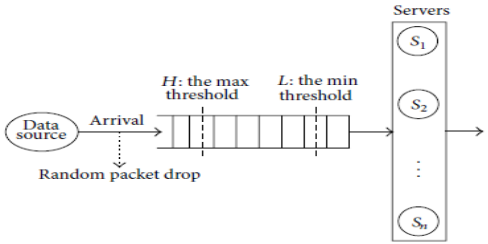
Data Routing

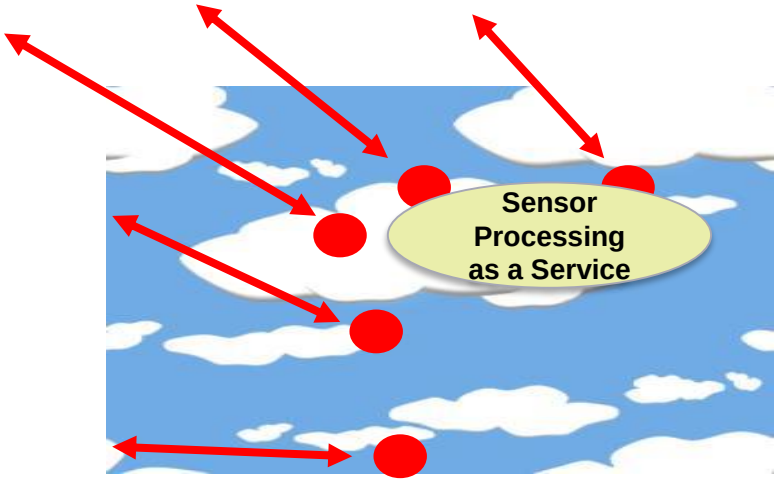


Data routing

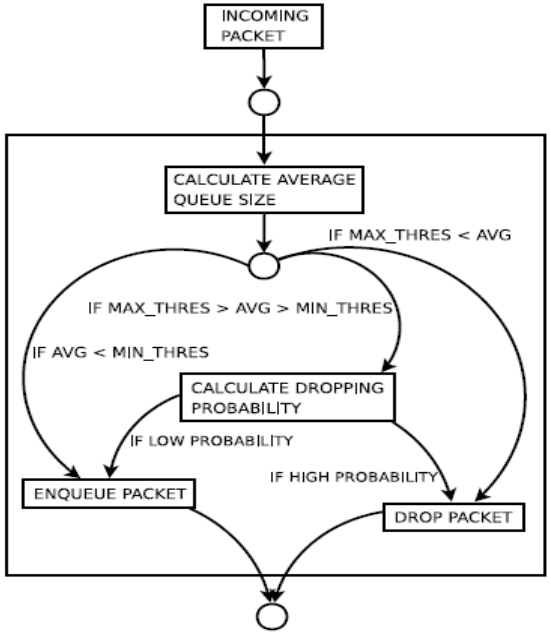
- Data routing happens over network switch
- During congestion packets are dropped
- Network buffers management policies can lead to poor latencies (if buffers become too large)
- also leading to a lot of packet droppings and low utilization of links

Sensor cloud example: Packet drop



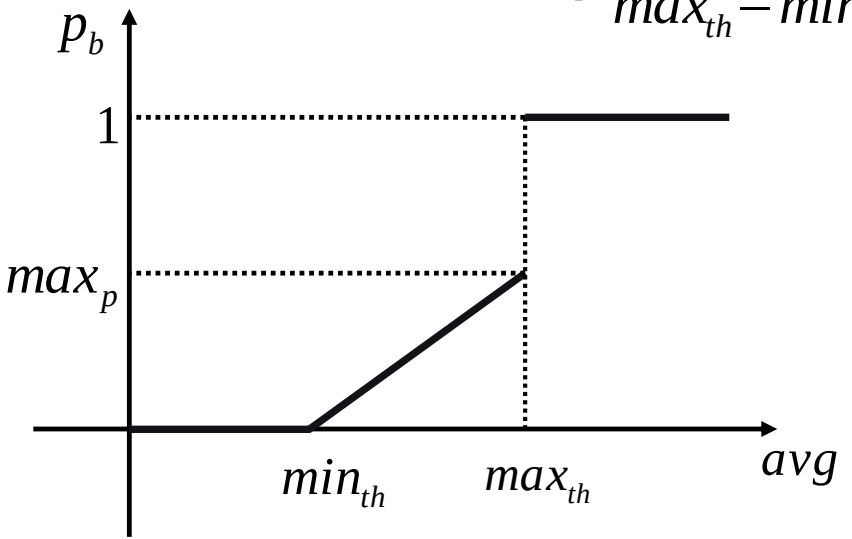


Working of RED



Packet Drop Probability

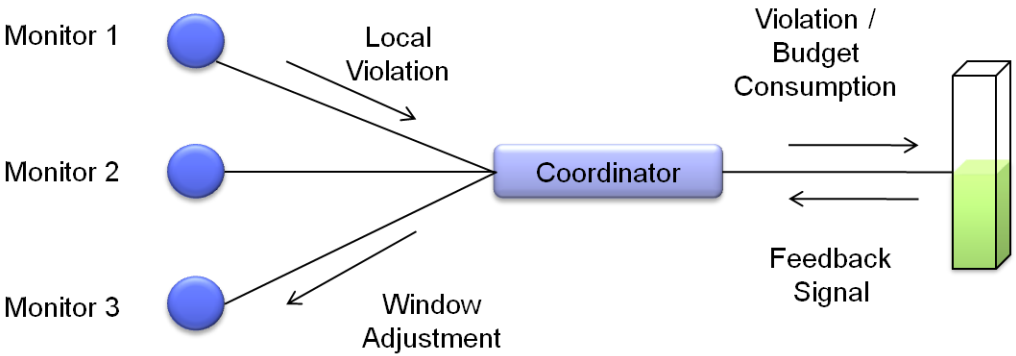
$$p_b = max_p \frac{avg - min_{th}}{max_{th} - min_{th}}$$



Working of predictor

- Prediction tools predicts the congestion based on data flow rate,packet loss etc
- The monitoring tools has option to setup alert for some threshold value, after which an alert is sent to sources
- The weighted random early detection (WRED) running inside switch alert the node to control the flow rate
- Better method is to collect the bandwidth usage from all nodes and switches and create the mapping for switches and nodes to help easily predict the data flow
- Random Early Detection (RED) make scheduling decisions at each switch.

WRED

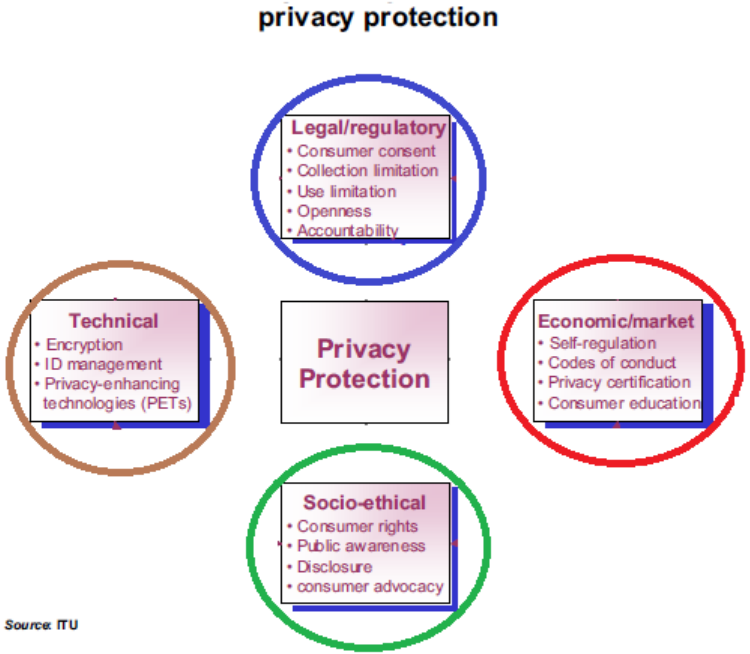

Monitor 1
Monitor 2
Monitor 3
Local
Violation
Coordinator
Violation /
Budget
Consumption
Feedback
Signal
Window
Adjustment

Performance with RED and 3 step prediction

sl	No.of sources	Variance with RED	Variance with prediction	Max queue With RED	Max Queue with prediction
1	20	125.2279	106.8508	404	404
2	30	134.0159	120.8611	475	475
3	40	140.5793	128.3137	539	539
4	60	142.8687	111.8134	654	654
5	80	177.0254	126.0417	738	735
6	100	194.5093	138.2350	822	822

Data management issues

Data management issue : security



Data Management issue: Memory

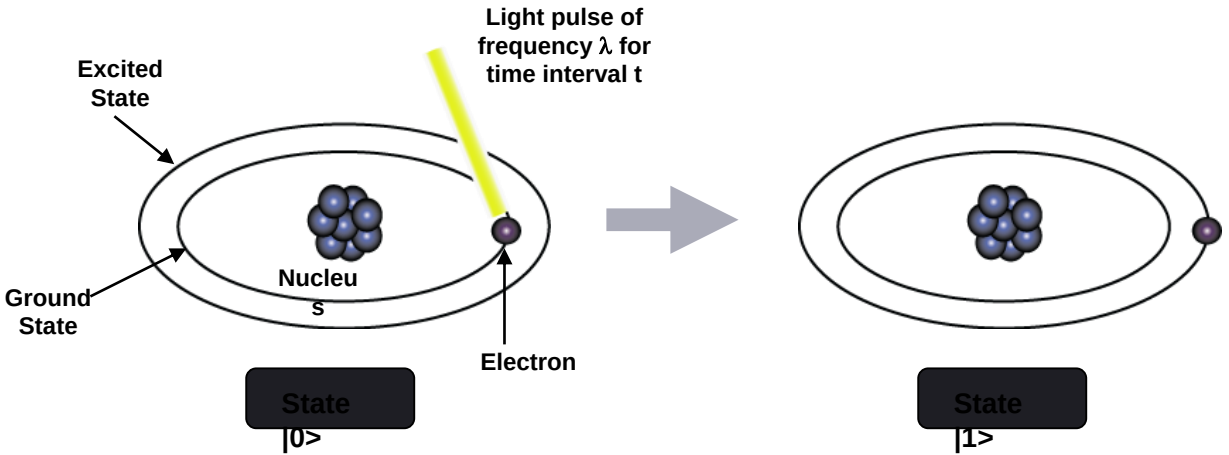
In-place computation

#	Sensor attributes	In place computation	Far off computation
1	Energy	more	less
2	storage	more	less
3	communication	less	more
4	Overall delay	less	more
5	Accuracy	?	?

Data management issue: Physical size

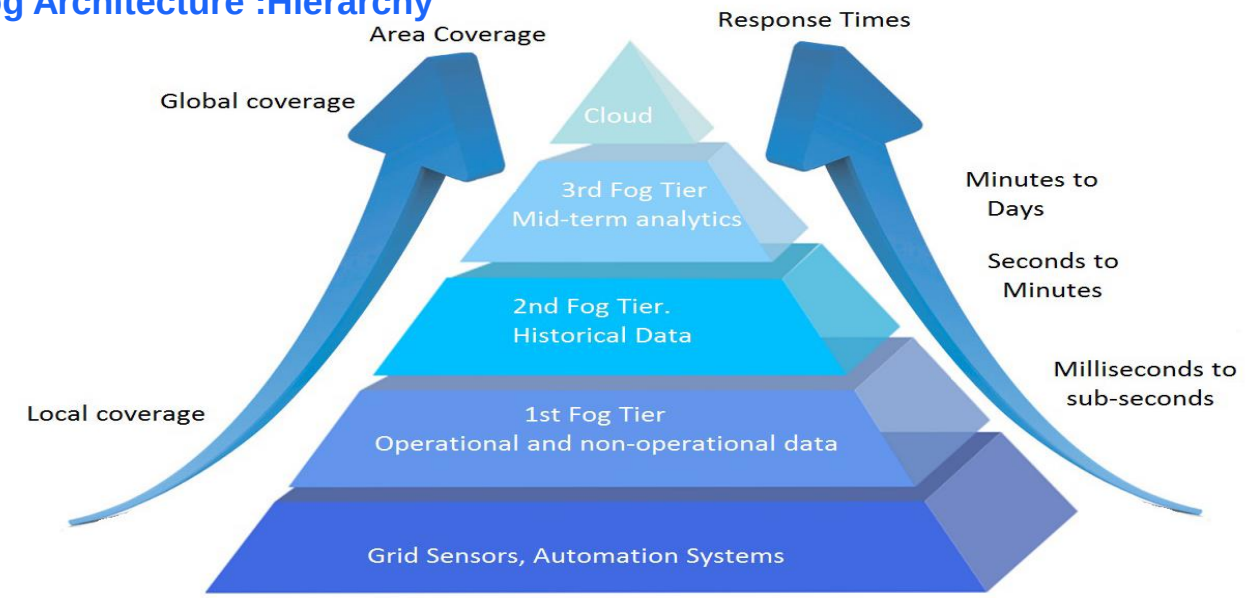
Quantum memory

- A bit of data in one of two states denoted by |0> and |1>.
- A single bit of this form is qubit
- $|\psi\rangle = \alpha\ |0\rangle + \alpha\ |1\rangle$



Data management issue: Access speed

Fog Architecture :Hierarchy

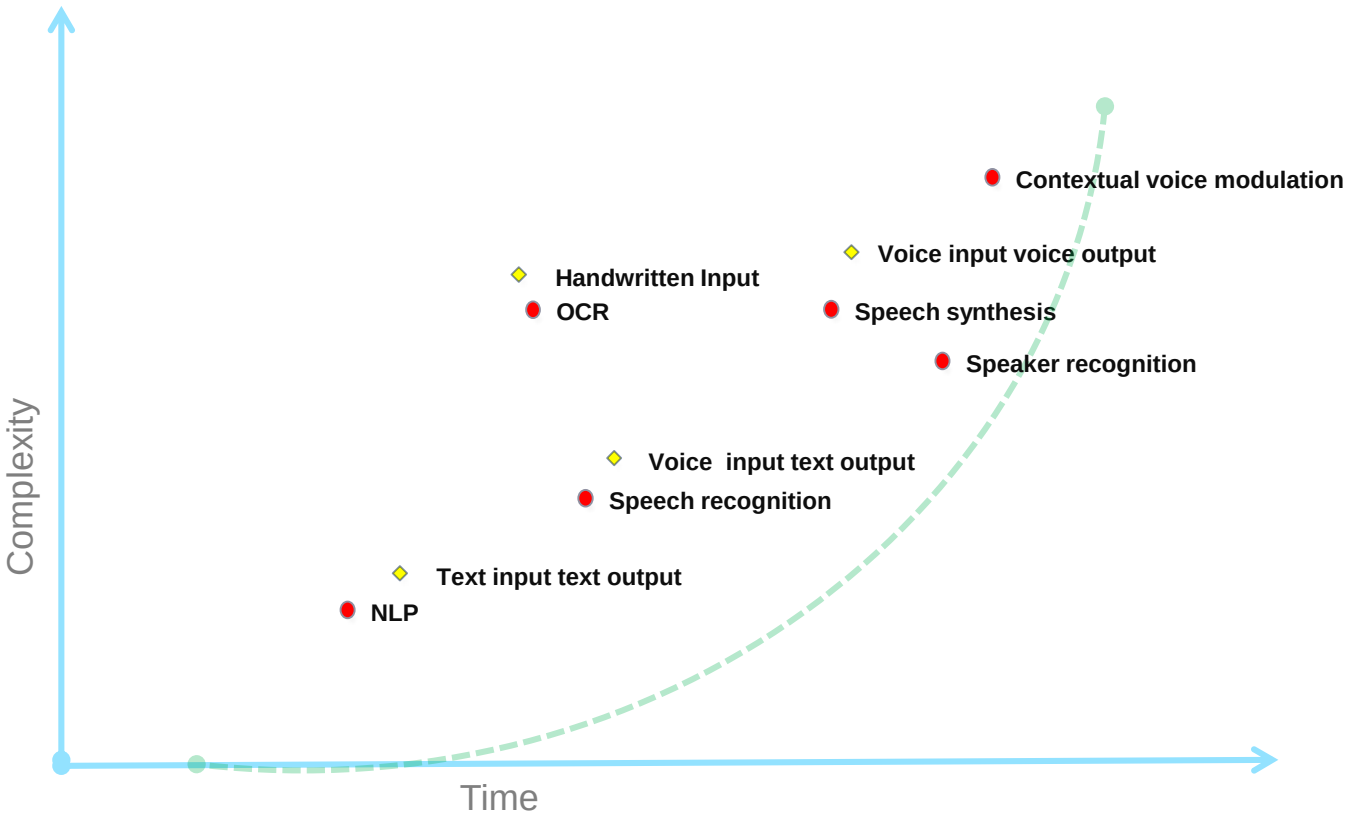


Source: Wikipedia

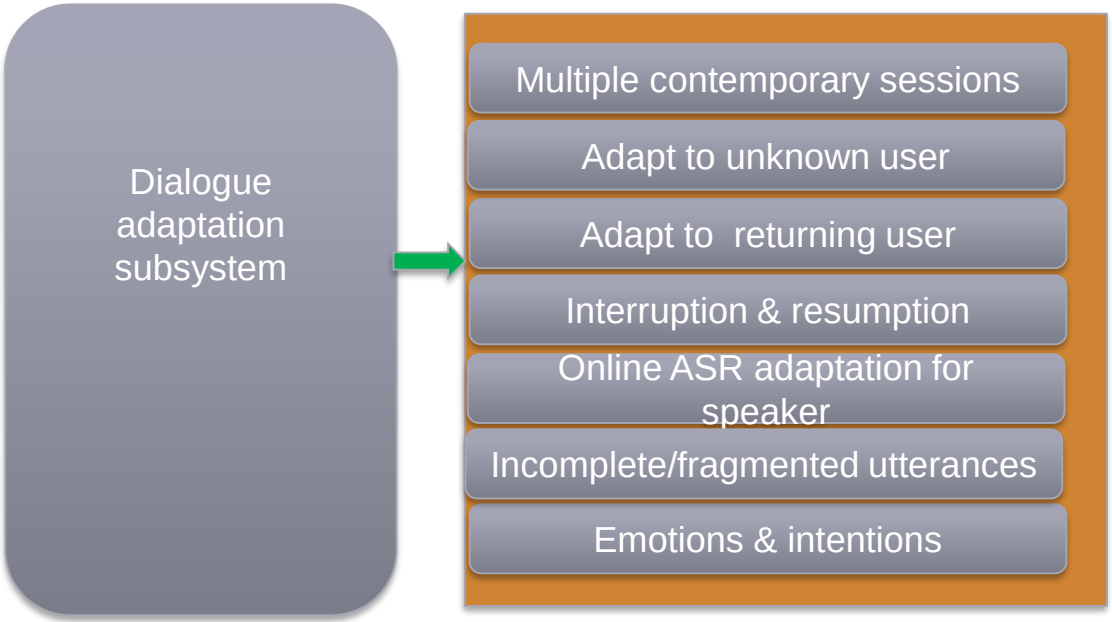
Fog computing

Requirement	Cloud Computing	Fog Computing
Latency	High	Low
Delay Jitter	High	Very low
Location of server nodes	Within the Internet	At the edge of the local network
Distance between the client and server	Multiple hops	One hop
Security	Undefined	Can be defined
Attack on data en route	High probability	Very low probability
Location awareness	No	Yes
Geographical distribution	Centralized	Distributed
Number of server nodes	Few	Very large
Support for Mobility	Limited	Supported
Real-time interactions	Supported	Supported
Type of last mile connectivity	Leased line	Wireless

Data management issue: Analysis



Cognitive components: eg. Dialog system



Power issues

Power issues: Optimal expenditure

Expenditure load balancing

- In Wireless Mesh Networks, by allowing nodes to **relay messages** for other nodes, the distance that needs to be bridged can be reduced, reducing the energy needed for a transmission.
  - The number of transmissions a node needs to perform increases costing more energy.
  - Decides whether or not to relay traffic. The usage of nodes having low capacity should be avoided for relaying traffic.
  - Routing algorithm provides cheapest route based on number of hops within a mesh network.

Power Issues: Optimal capacity

Energy harvesting

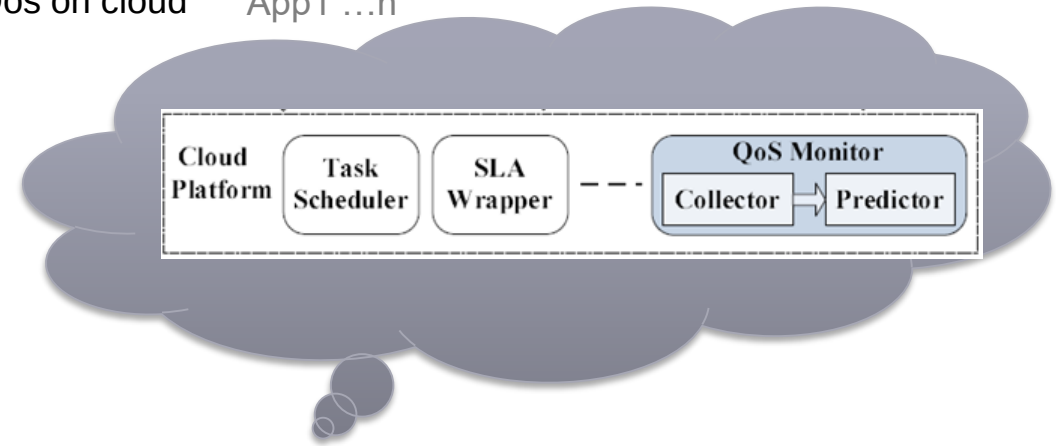
- Ambient energy **from light, vibration**, and **heat** may be used to run a low power WSN.
  - generate more than 150µW of power, enough to run a IPv6 routing node in a 802.15.4e network
- More **brightly lit areas**, such as workstation areas and reading surfaces, have 500 lux of lighting.
  - With 200-300 lux of light, small photovoltaic cells can supply sufficient power to operate an IPv6 router in a 802.15.4e network.
- **Thermal Electric Generators** (TEGs) produce power from the heat of hot surfaces, such as computer monitors or high-current motors.
- The energy produced from a **temperature differences** of even 10ºC becomes usable as an energy source.
  - The typical difference between internal body temperature and room temperature is about 15º C.

Connectivity issues

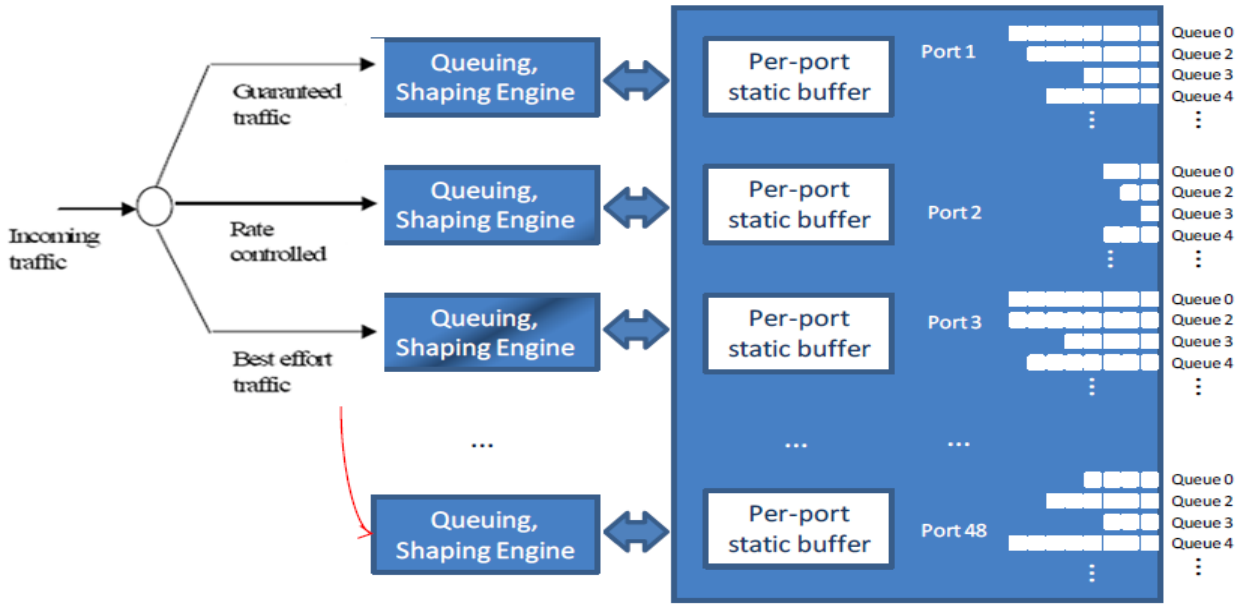
Connectivity issue: Varied priority

Qos on cloud

App1 …n



Traffic class



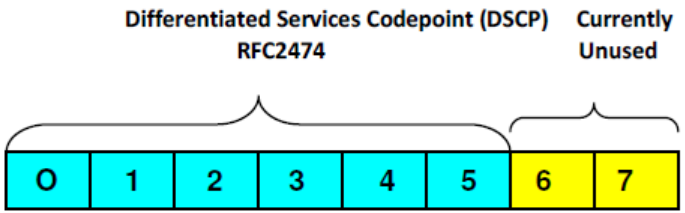
Resources for multiple classes

- 1. Immediate reservation:
  - Resources are provided right away or rejected.
- 2. In-advance reservation:
  - Resources must be available at specified time. Often an a fixed price charge is required to initiate a reservation and another rate is charged for the instances throughout the duration of the reservation.
- 3. Best effort reservation:
  - request are queued and serviced accordingly.
- 4. Auction based reservation: customers bid
  - for a particular configuration. As soon as dynamically adjusted resource price lowers the bid amount the resources are allocated.

Predictor

- Auto-regressive moving average (ARMA) model used for incoming workload
- The forecasting module predicts the work-load ahead of time
- The controller estimates the number of necessary resources (e.g., processing cores)
- The Predictor controller ensure a base level of guaranteed rate,
- Predictor can proportionally share available resources among applications with more demands than their guarantees

Packet marking



Relative QoS

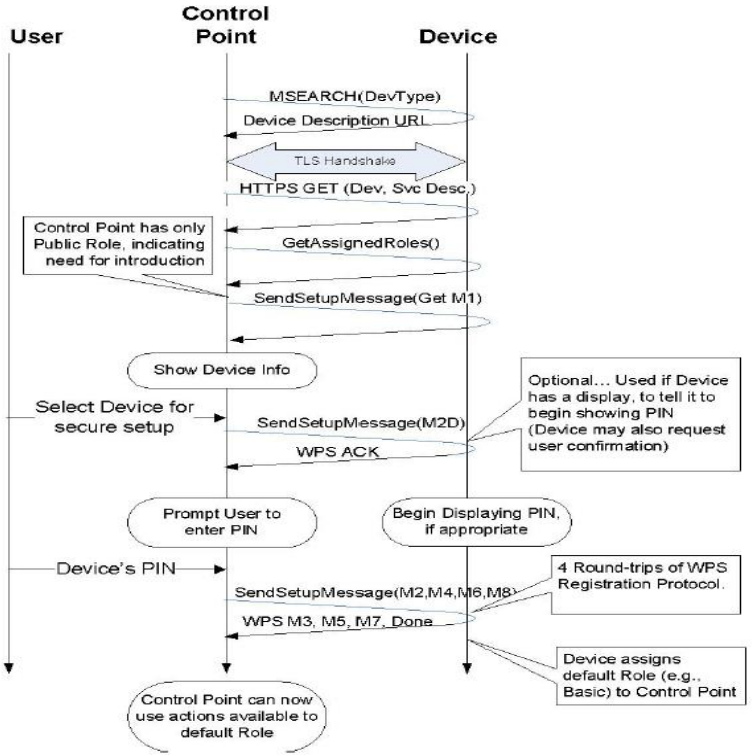
- Need: different goals are set for different entities.
  - Some users take the **shortest finishing time** as priority,
  - some take the **lowest cost** as priority
  - For some, the goal is to achieve **optimal matching** between simulation tasks and virtual machines
  - For others, to satisfy the applied multi-dimension **QoS** needs.
- Tiered model
  - different clients get different levels of service
  - Jobs of low-priority clients may be preempted (aborted or suspended) by jobs of high-priority Clients
  - on-demand instances available when there is no demand for reserved instances
  - **each client gets an absolute guarantee** (for receiving resources and for price paid)

Connectivity issue: device discovery

Ease of plugin

- It’s important to immediately know
  - when an IoT device drops off the network and goes offline
  - when that device comes back online
- Presence detection of IoT devices gives an exact, up to the second state of all devices on a network.
  - ability to monitor IoT devices and fix any problems that may arise with the network

Device discovery through upnp



Resolution

	#	Issue	Resolution
Processing	1	Processing power	Cloud
	2	Bursty request	Fractal model
	3	Scalability	Migration & Load balance
	4	Optimal resource utilization	Prediction based allocation
Data	5	Memory	Inplace computation
	6	Security	
	7	Analytics	Cognitive
	8	Physical size	Qbit
	9	Speed	Abstraction, fog computing

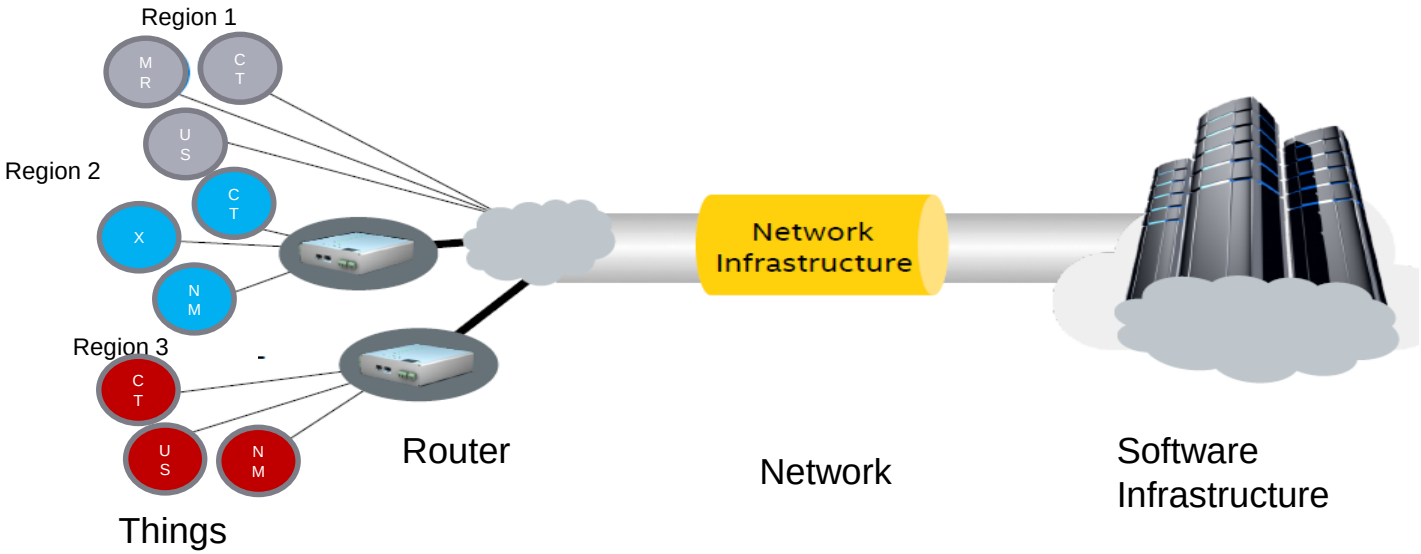
Resolution

	#	Issue	Resolution
Connectivity	10	Device discovery	uonp
	11	Resource utilization	Traffic shaping
	12	Varied priority/heterogeneous services	QoS
	13	Heterogeneous devices(renderers)	
	14	Fading	
	15	Multipath	
	16	Address space	Transition to IPv6 – Internet protocol v6
Power	17	Optimal expenditure	Computation vs acquisition, power load balance
	18	Optimal capacity	power loading, Energy Harvesting
	19	Power Quality	
	20	Charge time & duration	

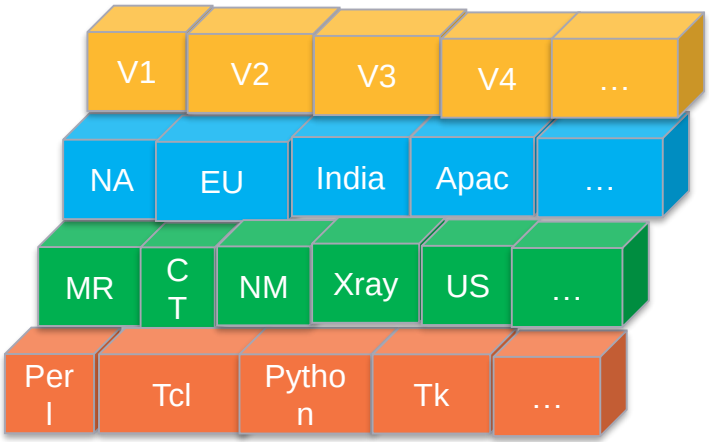
Agenda

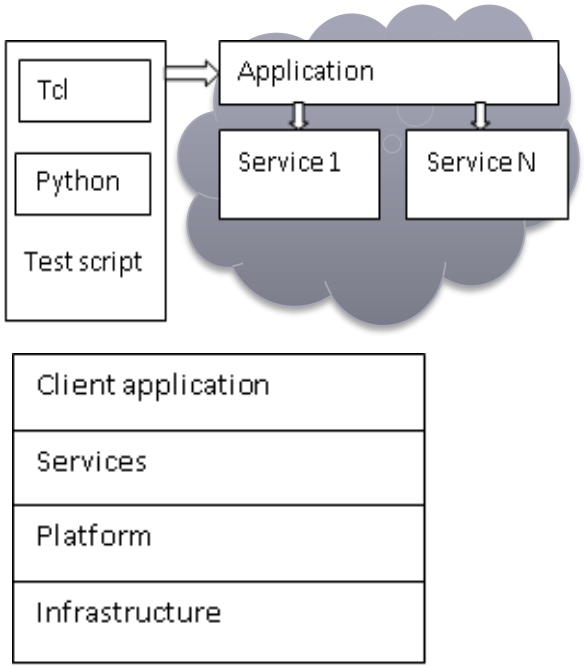
- Objective
- The IoT
3. Issues
4. Solution
5. Case study
6. Conclusion

Enterprise Dev environment

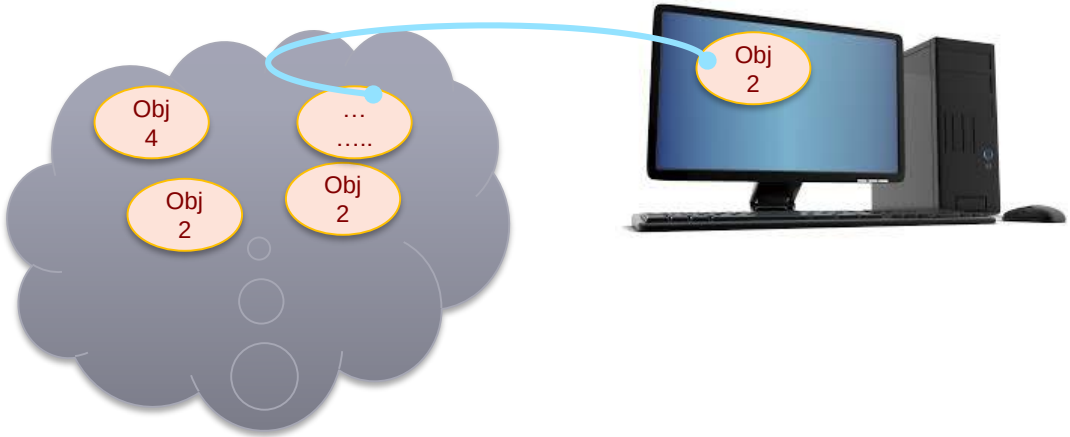


Enterprise IoT: Test automation





UI objects



Client Application
Object selector
Services
Platform
Infrastructure

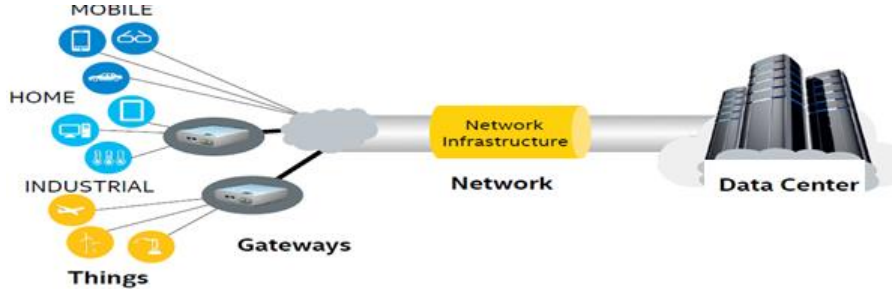
Design Examples

Design using IoT

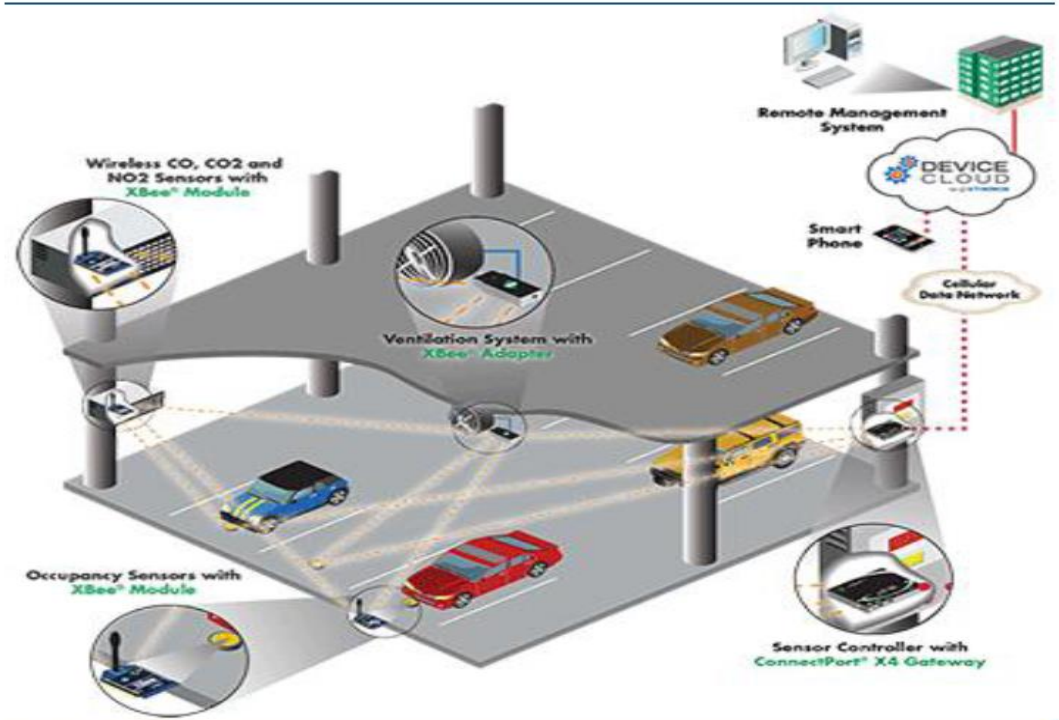
- Parameters
  - Right choice of scope
  - IoTs
  - Technologies
  - Connectivity & solution
  - Design
  - Algorithms
  - Objects/components
  - Stake holders
  - features
  - Issues

Design issue

- Possible use-cases and solution(technologies, input, output..) around
  - google glass



- Best solution for car parking in a lot: It should indicate location of empty slots



Source: Digi.com.

Machine Augmented Advisor

Hey! Take A diversion to left. Today U can't go right as there is an accident blocking the road 2 kms ahead

Maa Integrates GPS, Local news channel (FM) & daily routines

Hey! Take an umbrella . It is likely to rain in the football ground although Sunny here

Maa Integrates GPS, weather report and inbox (calander)

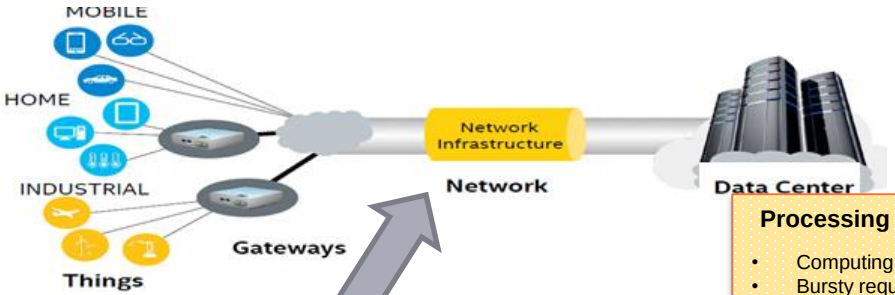
Hey! Your friend, whom u met yesterday , had a cardiac arrest in the morning. Would u like to visit him ?

Maa Integrates & tracks Personal activities, Web accounts & social network

Agenda

1. Objective
2. The IoT
3. Issues
4. Solution
5. Case study
6. Conclusion

Key Takeaway



Processing

- Computing power
- Bursty request
- Scalability
- Optimal resource utilization-

Data

- Memory
- Security
- Analytics
- Physical size
- Speed

Connectivity

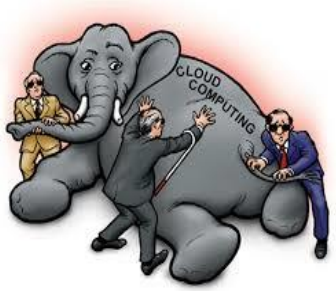
- Device identification
- Traffic shaping
- Heterogeneous service
- Heterogeneous device
- Fading
- Multipath
- Address space

Power

- Optimal expenditure
- Optimal capacity
- Power Quality
- Charge time & duration

Summary

- Optimal utilization of the resources plays a crucial role for the performance of the IoT devices
- Better prediction of the data traffic holds the key for the optimization
- Data traffic from devices exhibits a variety of statistical features, that may be exploited for prediction
- Better organization of the software in to multiple modules helps to achieve enhanced performance



References

- http://surrylearn.surrey.ac.uk
- ETSI, Machine to Machine Communications, http://www.etsi.org/technologies-clusters/technologies/m2m
- Machine-to-Machine Communications, OECD Library, http://www.oecd-ilibrary.org/science-and-technology/machine-to-machine-communications_5k9gsh2gp043-en
- W3C Semantic Sensor Networks,http://www.w3.org/2005/Incubator/ssn/XGR-ssn-20110628/
- Mahboobeh Ghorbani et al Prediction and control of bursty cloud workloads: a fractal framework codes'14