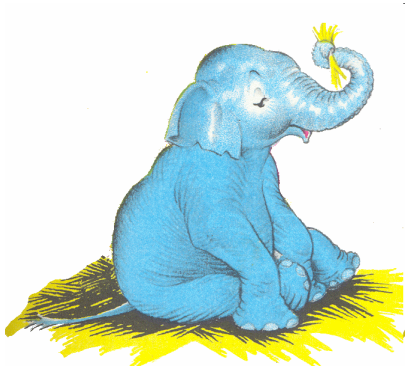
Networking Technologies for Big Data and Internet of Things



Raj Jain
Washington University in Saint Louis
Saint Louis, MO 63130
Jain@cse.wustl.edu

These slides and audio/video recordings of this lecture are at:

http://www.cse.wustl.edu/~jain/tutorials/gitma15.htm

Overview

1. What are Things?
2. What's Smart?
3. Why IoT Now?
4. Business/Research Opportunities in IoT
5. Why, What, and How of Big Data:
   It's all because of advances in networking
6. Recent Developments in Networking and their role in Big Data (Virtualization, SDN, NFV)

Cloud Computing

- August 25, 2006: Amazon announced EC2 ⇒ Birth of Cloud Computing in reality (Prior theoretical concepts of computing as a utility)
- *Web Services To Drive Future Growth For Amazon* ($2B in 2012, $7B in 2019) - Forbes, Aug 12, 2012
- Cloud computing was made possible by computing virtualization
- **Networking**: Plumbing of computing
  - IEEE: Virtual Bridging, …
  - IETF: Virtual Routers, …
  - ITU: Mobile Virtual Operators, …

What are Things?

- Thing = Not a computer
- Phone, watches, thermostats, cars, Electric Meters, sensors, clothing, band-aids, TV,…
- Anything, Anywhere, Anytime, Anyway, Anyhow (5 A's)

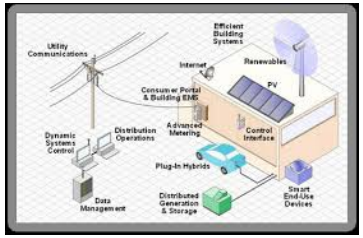
Ref: http://blog.smartthings.com/iot101/iot-adding-value-to-peoples-lives/
http://www.cse.wustl.edu/~jain/tutorials/gitma15.htm

Internet of Things

- Less than 1% of things around us is connected.
  Refrigerator, car, washing machine, heater, a/c, garage door, should all be connected but are not.
- From 10 Billion today to 50 Billion in 2020
  Should include processes, data, things, and people.
- $14 Trillion over 10 years
  ⇒ Third in the list of top 10 strategic technologies by Gartner
  (After Mobile devices, Mobile Apps, but before Clouds, …)
- a.k.a. **Internet of Everything** by Cisco
  Smarter Planet by IBM

Ref: "Gartner Identifies Top 10 Strategic Technologies,"
http://www.cioinsight.com/it-news-trends/gartner-identifies-top-10-strategic-technologies.html
Ref: J. Bradley, "The Internet of Everything: Creating Better Experiences in Unimaginable Ways," Nov 21, 2013,
http://blogs.cisco.com/ioe/the-internet-of-everything-creating-better-experiences-in-unimaginable-ways/#more-131793

Sample IoT Applications

 Smart Grid	 Smart Health	 Smart Home
 Smart Cities	 Smart Industries	 Smart TV
 Smart Watch	 Smart Car	 Smart Kegs

What's Smart?

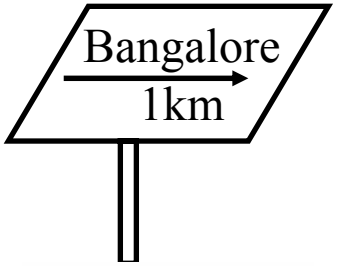
- IoT = Instrument, Interconnect, Intelligently process (3 I's)
- Old: Smart = Can think ⇒ Can compute
- Now: Smart = Can find quickly, Can Delegate
  ⇒ Communicate = Networking
- Smart Grid, Smart Meters, Smart Cars, Smart homes, Smart Cities, Smart Factories, Smart Smoke Detectors, …

Think

Communicate

4 Levels of Smartness

1. **Passive**: Communicate only when queried. Passive RFID, QR codes (*Nirjeeva*)
2. **Active**: Communicate when needed. Sensors. Home automation (*1-4 sense*)
3. **Aware**: Action based on simple computation. E.g., tele-health (*5-sense*)
4. **Autonomous**: Can make decisions based on rules. E.g., autonomous cars, smart grid (*Human*)



Ref: http://go.gigaom.com/rs/gigaom/images/gigaomresearch_the_internet_of_things_report.pdf

Internet of Brains

- Brain-to-Brain Interface
- A person's brain can send signals to other person's brain
- Useful for handicap people to communicate with others

Ref: http://homes.cs.washington.edu/~rao/brain2brain/experiment.html

Why IoT Now?

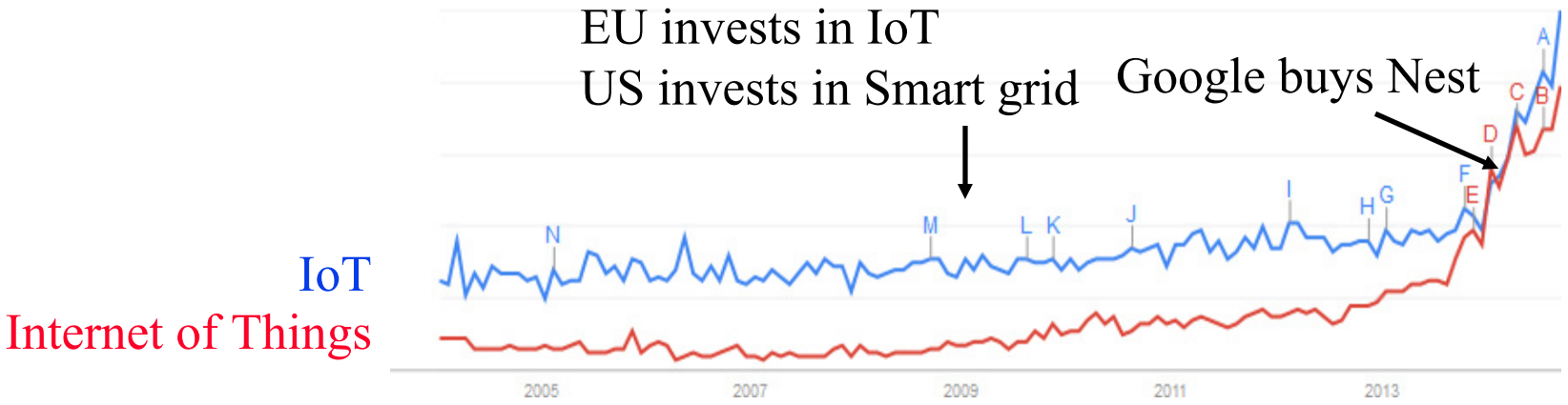
- IoT = Sensing + Communication + Computation

1. Micro-Sensors: Temperature, Moisture, Pressure, air quality, …
2. Tags: Radio Frequency Id (RFID), Quick Response (QR) Codes, …
3. Energy Efficient Communication: Small or no batteries, Personal area communication (PAN), Bluetooth, ZigBee, …
4. Micro-Computing: Micro multi-core chips, Raspberry Pi, Intel Galileo, Arduino, …
5. Cloud Computing: Little or no local computing
6. Open/Small operating systems: Linux

Ref: CTIA, "Mobile Cyber security and the Internet of Things,"
http://www.ctia.org/docs/default-source/default-document-library/ctia-iot-white-paper.pdf

Funding

Google Trends



- Around for 10 years
- IERC-European Research Cluster on the Internet of Things funded under 7th Framework in 2009
  ⇒ "Internet of European Things"
- US interest started in 2009 w $4B funding for **smart grid** in American Recovery and Reinvestment Act of 2009

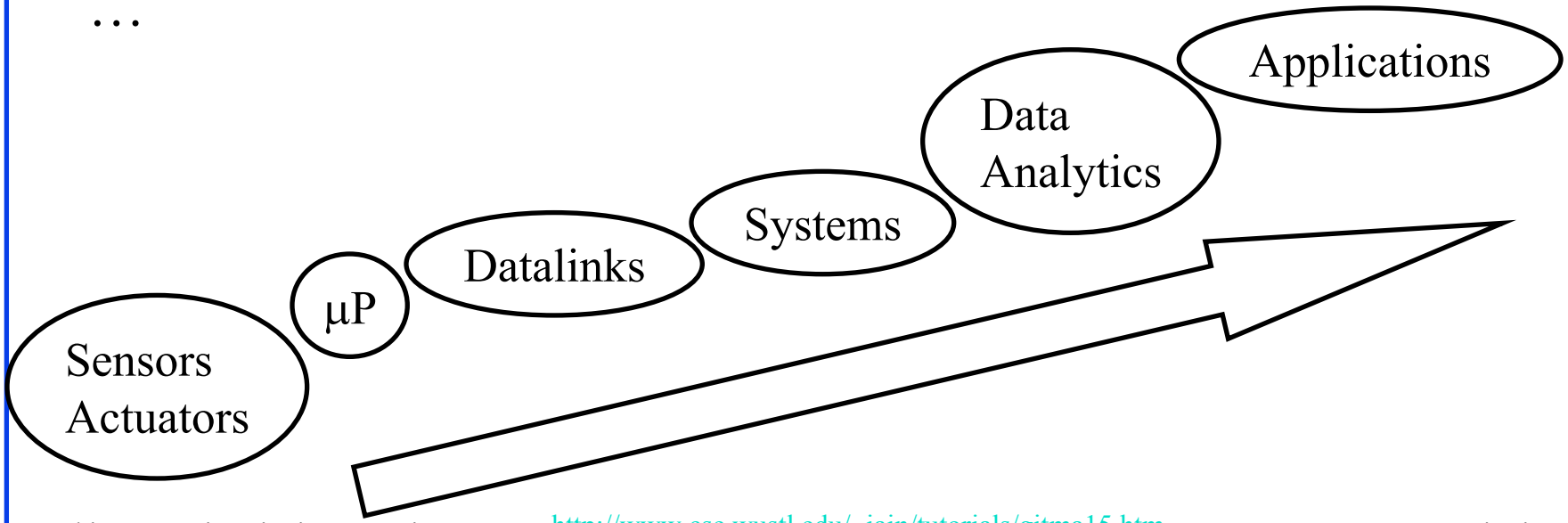
Research Funding for IoT

- 70 M € in European Research program FP7
  ⇒ Internet of European Things
- Networking and Information Technology Research and Development (NITRD)
  - Group of 15 Federal agencies: NSF, NIH, NASA, DOE, DARPA, ONR, …
  - Recommends supplement to the president's annual budget
  - CPS is one of the areas recommended by NITRD starting 2012 ⇒ Smart infrastructure
    - Smart Grid, Smart Bridges, Smart Cars, tele-operational surgical robots, Smart **Buildings**
- March 2014: £45M for IoT research in UK by David Cameron

Ref: NITRD, http://www.nitrd.gov/

Business Opportunities

- Components: Sensors, wireless radios, protocols,
- Smart Objects: Smart TV, Camera, Watch, …
- Systems: Buildings, Cars, Health, …
- Network service providers: ISP
- Application Service Providers: Monitoring, Analytics, Apps, …

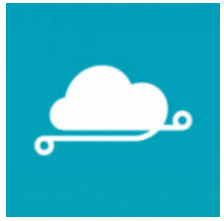
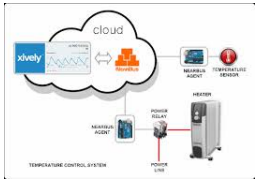
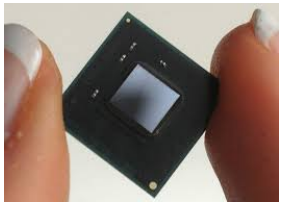
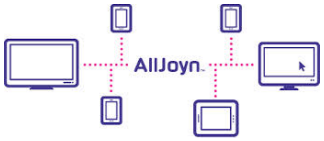


Venture Activities in IoT

- $1.1B invested in IoT startups by VCs in 153 deals in 2013
  - Quantified Self: Know your body and mind
  - Healthcare sensors: Wearable clock, sleep monitors
  - Energy management
  - Home Automation: Kitchenware, locks,
  - Environmental monitoring: Air Quality sensors, personal weather stations
- January 2014: Google buys NEST for 3.3B
- May 2014: $150M in VC investments in IoT by Cisco

Ref: http://www.cbinsights.com/blog/internet-of-things-investing-snapshot/
http://www.zdnet.com/cisco-invests-150m-in-internet-of-things-startups-7000028964/

Recent IoT Products

 NEST Thermostat	 Corventis: Wireless Cardiac Monitor	 WEMO Remote	 Tractive Pet Tracker
 Ninja Blocks	 Revolve Home Automation	 ThingWorx Application Platform	 Lings Cloud Platform
 Mbed Development Platform	 Xively Remote Access API	 Intel Quark Processor	 AllJoyn S/W Framework

IoT Research Challenges

1. **Naming and Addressing**: Advertising, Searching and Discovery
2. Service Orchestration
3. Power/Energy/Efficient resource management.
   Energy harvesting
4. **Things to Cloud**: Computation and Communication Gateways
5. **Miniaturization**: Sensors, CPU, network
6. **Big Data Analytics**: 35 ZB of data $2B in value by 2020
7. **Semantic technologies**: Information and data models for interoperability
8. **Virtualization**: Multiple sensors aggregated, or a sensor shared by multiple users
9. **Privacy/Security**/Trust/Identity/Anonymity
   Target Pregnancy Prediction
10. Heterogeneity/Dynamics/**Scale**

Internet of Harmful Things

Imagine, as researchers did recently at Black Hat, someone hacking your connected toilet, making it flush incessantly and closing the lid repeatedly and unexpectedly.

News

Worm may create an Internet of Harmful Things, says Symantec (Take note, Amazon)

Security firm Symantec says it has found a Linux worm aimed at Internet of Things devices

By Patrick Thibodeau
December 3, 2013 01:22 PM ET Add a comment

Share 6 g+1 Like 7 More

Computerworld - Security researchers are gradually raising warnings that the Internet of Things will increase, by multitudes, the number of things that can be hacked and attacked.

The Hitchcockian plotlines are endless. Replace *The Birds* with flying Amazon delivery drones. Or imagine, as researchers did recently at Black Hat, someone hacking your connected toilet, making it flush incessantly and closing the lid repeatedly and unexpectedly.

Beacons

- Advertizing based on proximity
- Peripherals (your phone) broadcasts its presence if Bluetooth is turned on
- Primary aim of these broadcasts is to allow device discovery
- Advertising packets consist of a header and max 27B of payload with multiple TLV-encoded data items
  - May include signal strength ⇒ Distance
- iOS7 iPhones can send/received iBeacons
- Can be used for customized advertising, indoor location, geofencing
- PayPal uses this to identify you. You can pay using a PIN and your phone.

Power per MB

Type	Bit rate	TX Power	mJoules/MB
802.11b	11Mb	50mW	36.4
802.11g	54Mb	50mW	7.4
802.11a	54Mb	200mW	29.6
802.15.1 Bluetooth	1Mb	1mW	8.0
802.15.3	55Mb	200uW	0.03

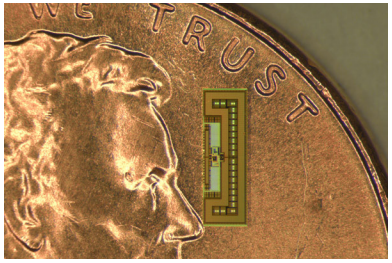
- Once connected, Bluetooth classic maintains connections even when there is no data. Low power but not low enough.

Datalink Issues

- Energy efficiency
  - ⇒ Need to decrease energy/bit by a factor of 1000
  - Energy/bit has gone down by a factor of 2 per year
  - Either wait ten years or design better protocols
- Small messages ⇒ Need low overhead
- Limited computing ⇒ Light weight protocols
  ⇒ lightweight Encryption, authentication, security
- Quality of Information (QoI)

Ant-Sized IoT Passive Radios

- Computer + Sensor + Radio in 3.7x1.2 mm from Stanford
- Can be added to dollar bills, band-aids, tools, …
- Monitor temperature, location
- 3 m range
- Extremely low power ⇒ No battery required (Similar to passive RFID
- Continuously monitor every part of the body of every patient



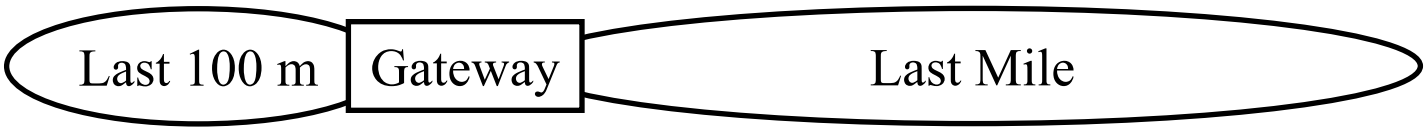
Ref: http://www.computerworld.com/article/2682854/stanfords-ant-sized-radios-could-connect-the-world.html?source=CTWNLE_nlt_pm_2014-09-12#tk.rss_all
http://web.stanford.edu/~arbabian/Home/Welcome.html

Networking Issues

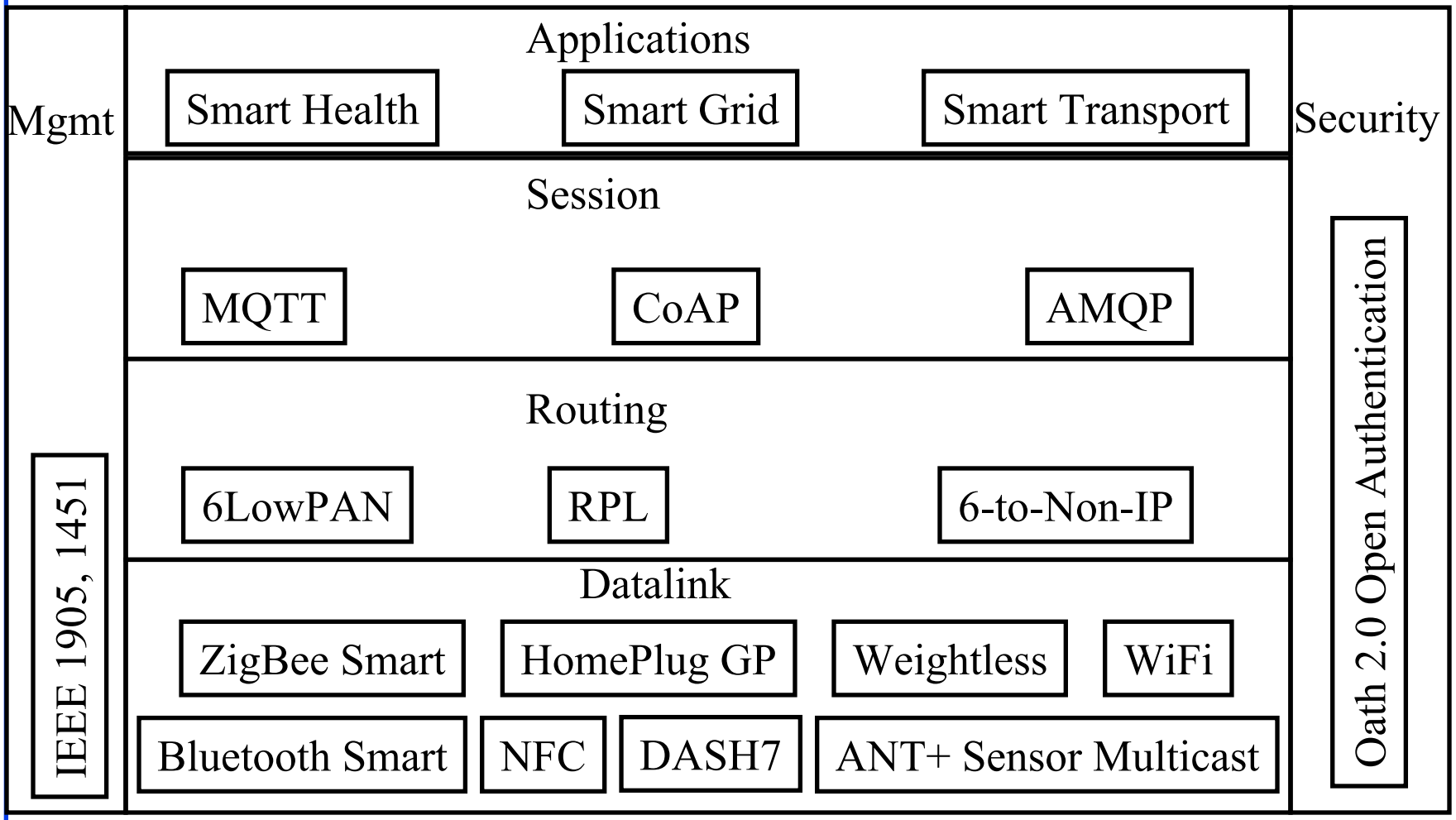
- Large number ⇒ 32-bit or 48-bit addressing not sufficient
- 32-bit IPv4 addresses too small
- 48-bit IEEE 802 too small
- 128-bit IPv6 addresses too large. Tiny things do not have energy to transmit such large addresses.
- 16-bit local addresses and 64-bit global addresses
- 6LowPAN, 6-to-NonIP

Last 100m Protocols

- The Last Mile: Mobile and Broadband Access revolution
  Smart Grid, Smart Cities, Smart Industries
- The last 100m: Smart home
- The last meter: Smart Healthcare, Smart Wearable's



Recent Protocols for IoT



Ref: http://tools.ietf.org/html/draft-rizzo-6lo-6legacy-00, http://en.wikipedia.org/wiki/OAuth, http://en.wikipedia.org/wiki/ANT%2B
http://en.wikipedia.org/wiki/Near_field_communication, http://en.wikipedia.org/wiki/Weightless_%28wireless_communications%29

Legacy IoT Protocols

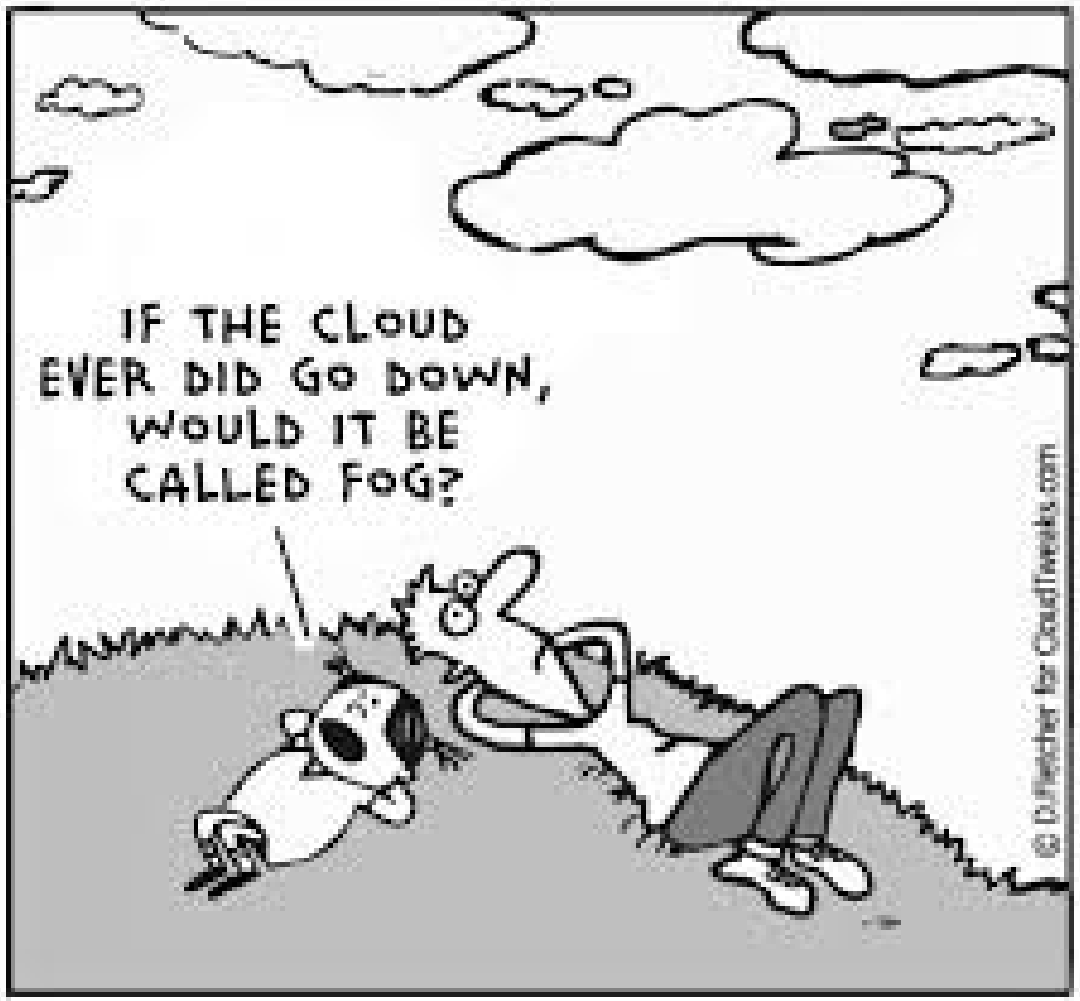
- **BACnet**: Building Automation and Control Network
- **LonWorks**: Local Operating Network (like BACnet)
- **ModBus**: Modicon (Schneider Electric)'s Serial Bus
- **KNX**: Home and Building Automation Standard
- **Z-Wave**: Wireless Communication for Home Automation
- **M-Bus**: Bus for remote reading of gas and electric meters
- **ANSI CI12.20**: Electric Meter Accuracy and Performance
- **DLMS**: Device Language Message Specification
- **COSEM**: Company Specification for Energy Metering

Standardization

- Almost every standards body is working on IoT: IEEE, IETF, ITU, ETSI, IPSO, …
- Seven organizations joined together to avoid duplication: ARIB, ATIS, CCSA, ETSI, TIA, TTA, TTC ⇒ oneM2M

Ref: http://www.onem2m.org

Fog Computing



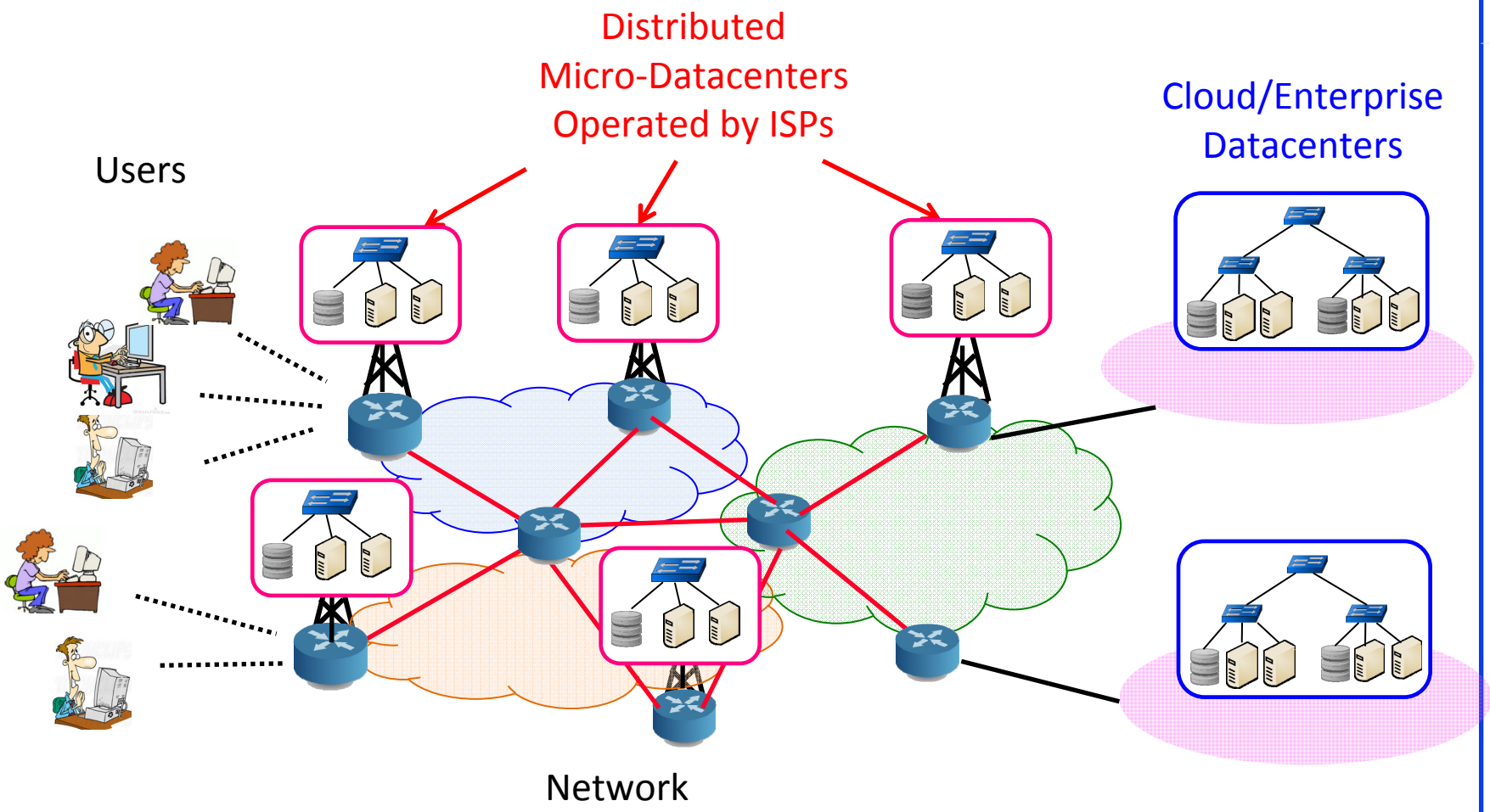
Ref: http://community.spiceworks.com/topic/254392-fog-computing-replaces-cloud-as-new-tech-buzzword
http://www.cse.wustl.edu/~jain/tutorials/gitma15.htm

Fog Computing (Cont)

- Location Aware and Location Sensitive
  ⇒ Low latency ⇒ Computing in micro clouds
  ⇒ Computing in the edge ⇒ Computing everywhere
  ⇒ Fog
- Geographically distributed => Everywhere/Anywhere
- Large Scale
- Mobility
- Real-Time

Ref: F. Bonomi, et al., "Fog Computing and Its Role in the Internet of Things," ACM MCC'12, August 17, 2012, Helsinki, Finland

Micro-Clouds on Cell-Towers



New Business Opportunities: Domain 2.0,
Datacenters on Towers, IoT, NFV, FV, Elastic Networks

The Problem Statement

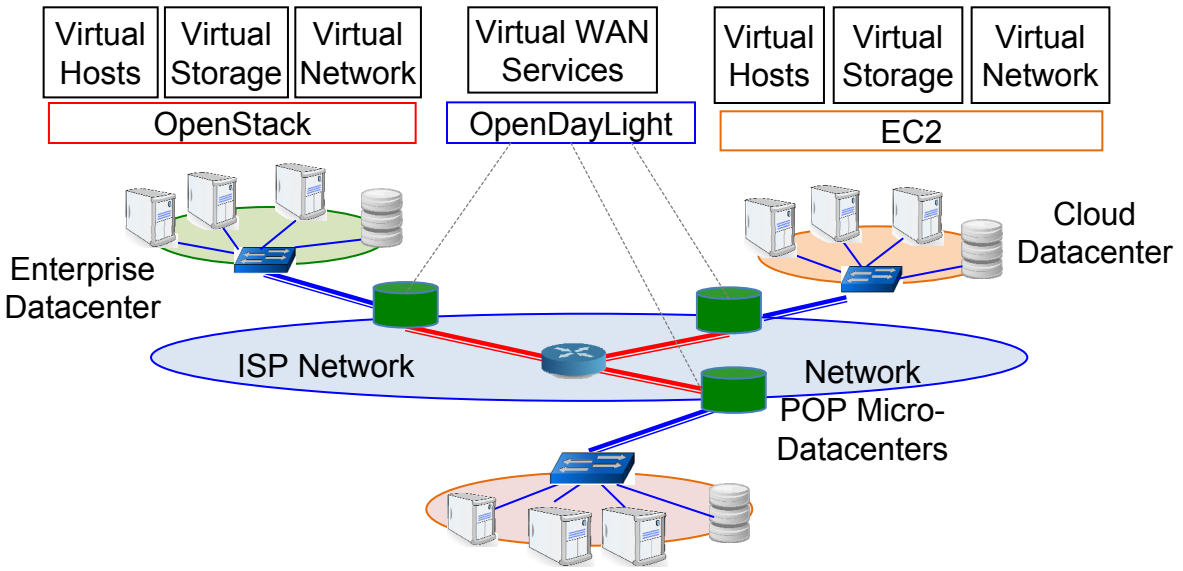
Massively Distributed Application Use Cases

Application Service Providers (ASPs)

Automatic Application Deployment and Delivery Platform

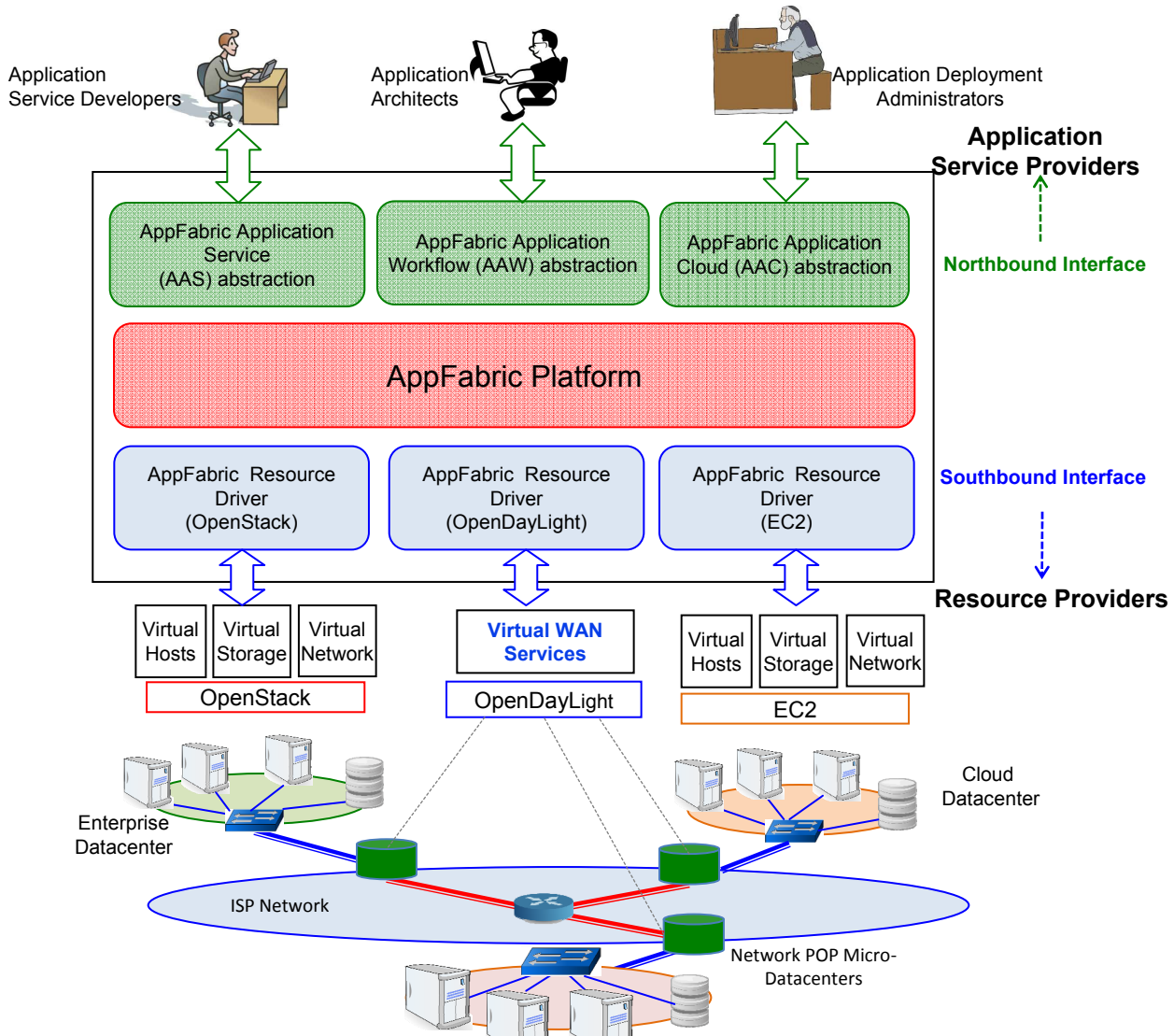
AppFabric

Distributed Virtual Infrastructure



Resource Providers (ISPs, CSPs)

Services in a Cloud of Clouds



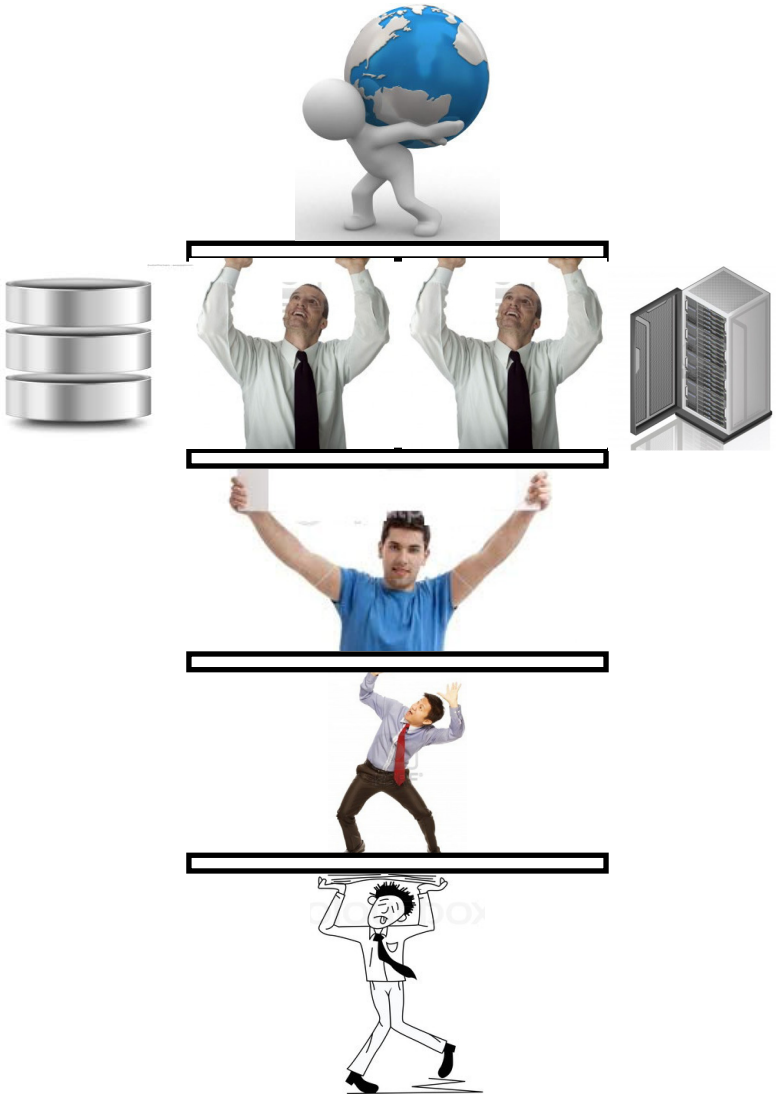
Why Big Data Now?

1. Low cost storage to store data that was discarded earlier
2. Powerful multi-core processors
3. Low latency possible by distributed computing: Compute clusters and grids connected via high-speed **networks**
4. Virtualization ⇒ Partition, Aggregate, isolate resources in any size and dynamically change it ⇒ Minimize latency for any scale
5. Affordable storage and computing with minimal man power via clouds
   ⇒ Possible because of advances in **Networking**

Why Big Data Now? (Cont)

6. Better understanding of task distribution (MapReduce), computing architecture (Hadoop),
7. Advanced analytical techniques (Machine learning)
8. Managed Big Data Platforms: Cloud service providers, such as Amazon Web Services provide Elastic MapReduce, Simple Storage Service (S3) and HBase – column oriented database. Google' BigQuery and Prediction API.
9. Open-source software: OpenStack, PostGresSQL
10. March 12, 2012: Obama announced $200M for Big Data research. Distributed via NSF, NIH, DOE, DoD, DARPA, and USGS (Geological Survey)

Big Data Enabled by Networking



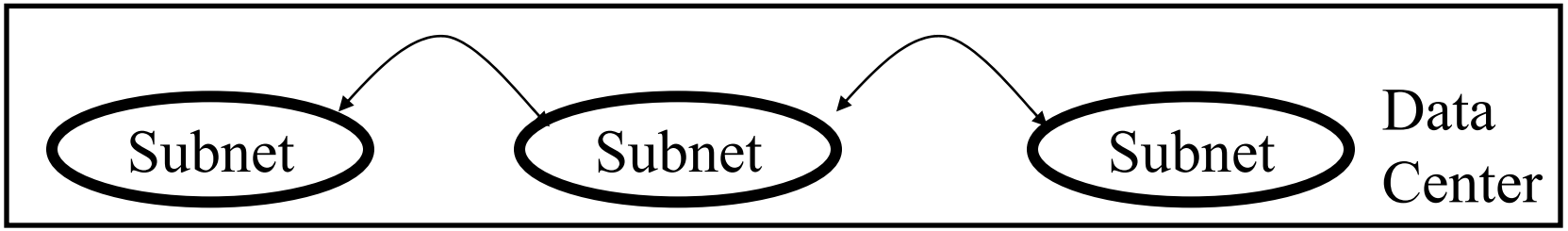
Recent Developments in Networking

1. High-Speed: 100 Gbps Ethernet
   $\Rightarrow$ 400 Gbps $\Rightarrow$ 1000 Gbps
   $\Rightarrow$ Cheap storage access. Easy to move big data.
2. Virtualization
3. Software Defined Networking
4. Network Function Virtualization

Virtualization (Cont)

- Recent networking technologies and standards allow:
  1. Virtualizing Computation
  2. Virtualizing Storage
  3. Virtualizing Rack Storage Connectivity
  4. Virtualizing Data Center Storage
  5. Virtualizing Metro and Global Storage

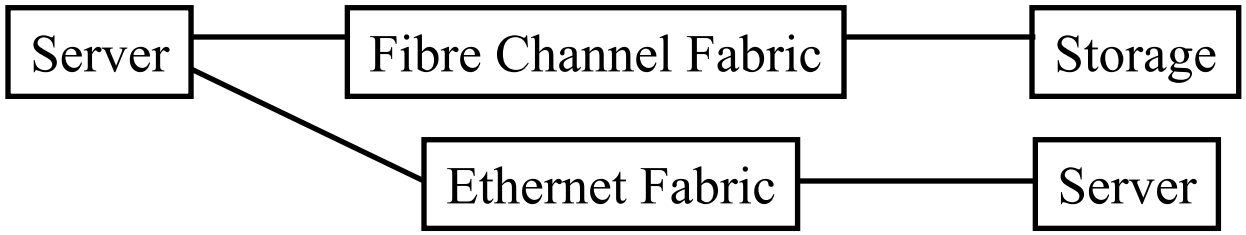
1. Virtualizing Computation



- Initially data centers consisted of multiple IP subnets
  - Each subnet = One Ethernet Network
  - Ethernet addresses are globally unique and do not change
  - IP addresses are locators and change every time you move
  - If a VM moves inside a subnet ⇒ No change to IP address ⇒ Fast
  - If a VM moves from one subnet to another ⇒ Its IP address changes ⇒ All connections break ⇒ Slow ⇒ Limited VM mobility
- IEEE 802.1ad-2005 Ethernet Provider Bridging (PB), IEEE 802.1ah-2008 Provider Backbone Bridging (PBB) allow Ethernets to span long distances ⇒ Global VM mobility

2. Virtualizing Storage

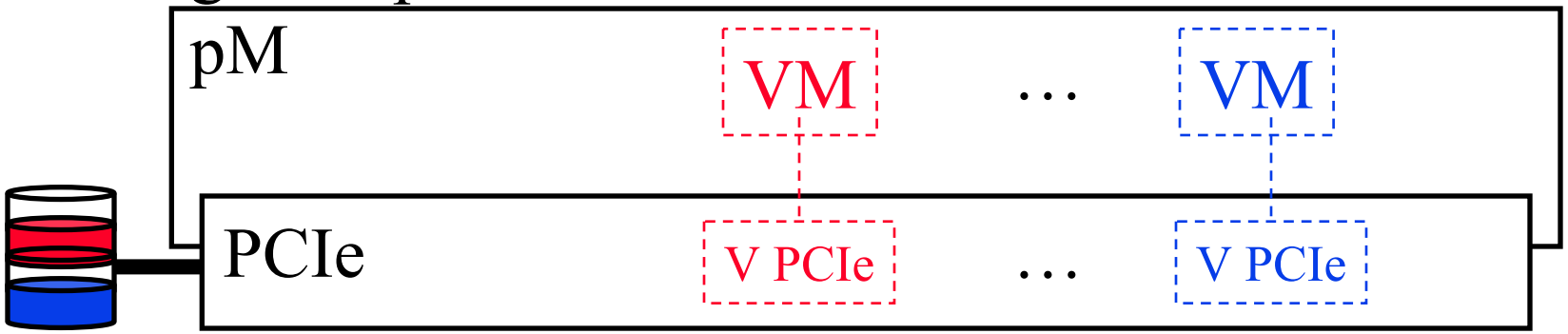
- Initially data centers used Storage Area Networks (Fibre Channel) for server-to-storage communications and Ethernet for server-to-server communication



- IEEE added 4 new standards to make Ethernet offer low loss, low latency service like Fibre Channel:
  - Priority-based Flow Control (IEEE 802.1Qbb-2011)
  - Enhanced Transmission Selection (IEEE 802.1Qaz-2011)
  - Congestion Control (IEEE 802.1Qau-2010)
  - Data Center Bridging Exchange (IEEE 802.1Qaz-2011)
- Result: Unified networking ⇒ Significant CapEx/OpEx saving

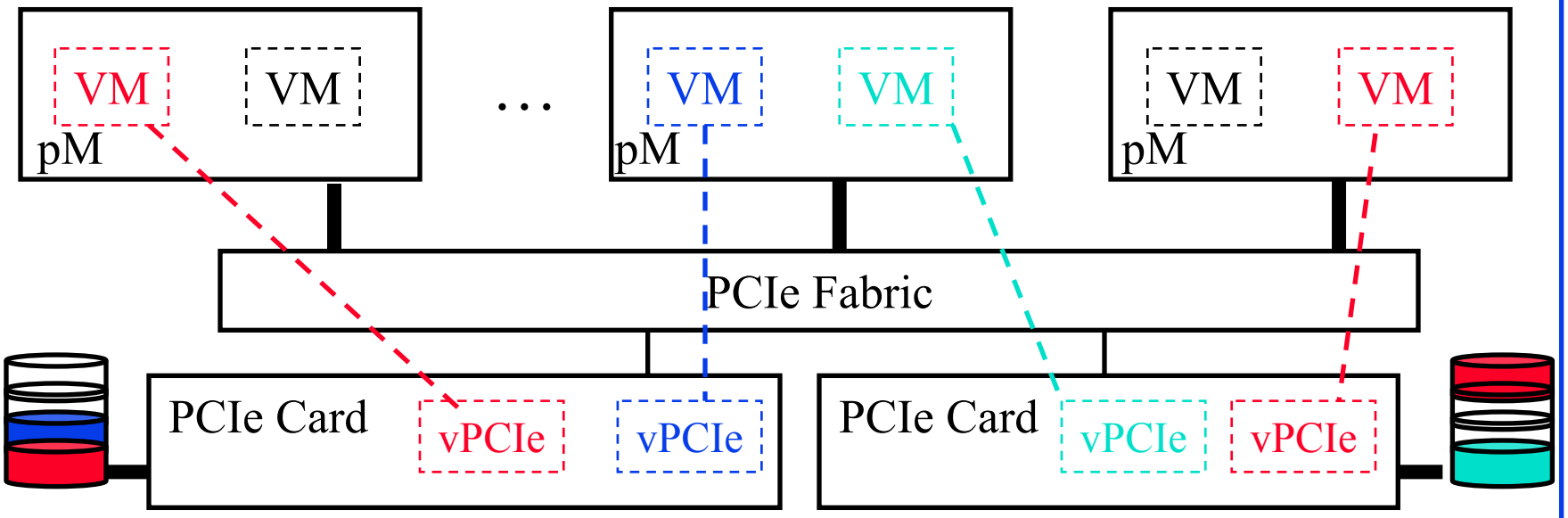
3. Virtualizing Rack Storage Connectivity

- MapReduce jobs are assigned to the nodes that have the data
- Job tracker assigns jobs to task trackers in the rack where the data is.
- High-speed Ethernet can get the data in the same rack.
- Peripheral Connect Interface (PCI) Special Interest Group (SIG)'s Single Root I/O virtualization (**SR-IOV**) allows a storage to be virtualized and shared among multiple VMs.



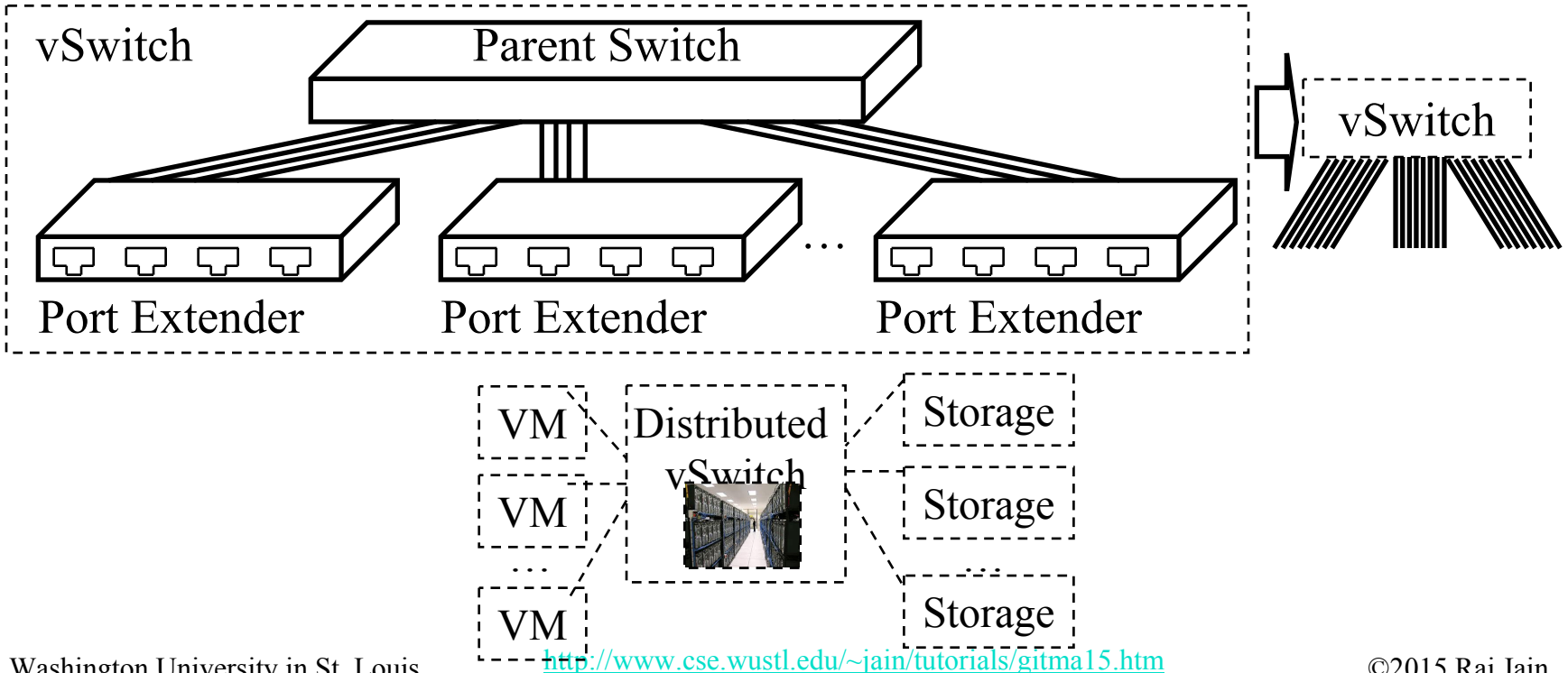
Multi-Root IOV

- PCI-SIG Multi-Root I/O Virtualization (**MR-IOV**) standard allows one or more PCIe cards to serve multiple servers and VMs in the same rack
- Fewer adapters ⇒ Less cooling. No adapters ⇒ Thinner servers



4. Virtualizing Data Center Storage

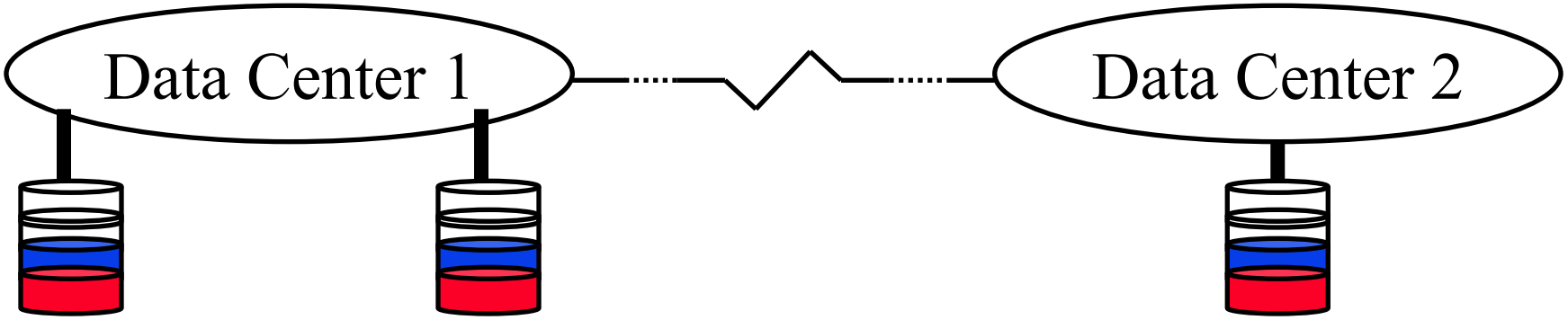
- IEEE 802.1BR-2012 Virtual Bridgeport Extension (VBE) allows multiple switches to combine in to a very large switch
- Storage and computers located anywhere in the data center appear as if connected to the same switch



5. Virtualizing Metro Storage

- Data center Interconnection standards:
  - Virtual Extensible LAN (VXLAN),
  - Network Virtualization using GRE (NVGRE), and
  - Transparent Interconnection of Lots of Link (TRILL)

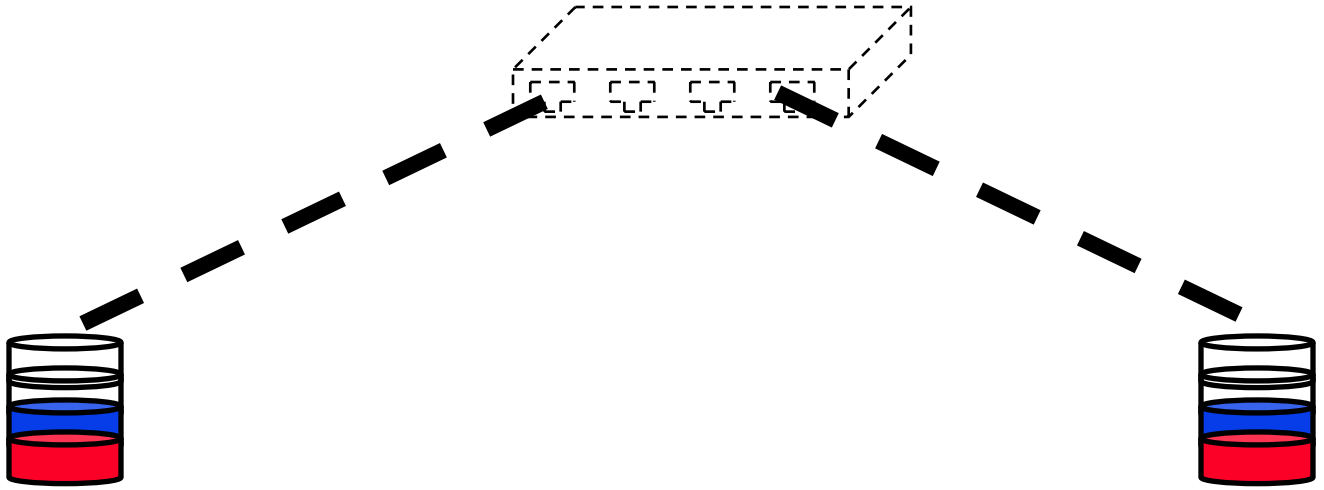
  ⇒ data centers located far away to appear to be on the same Ethernet



Ref: http://tools.ietf.org/html/draft-mahalingam-dutt-dcops-vxlan-04, http://tools.ietf.org/html/draft-sridharan-virtualization-nvgre-03, RFC 5556

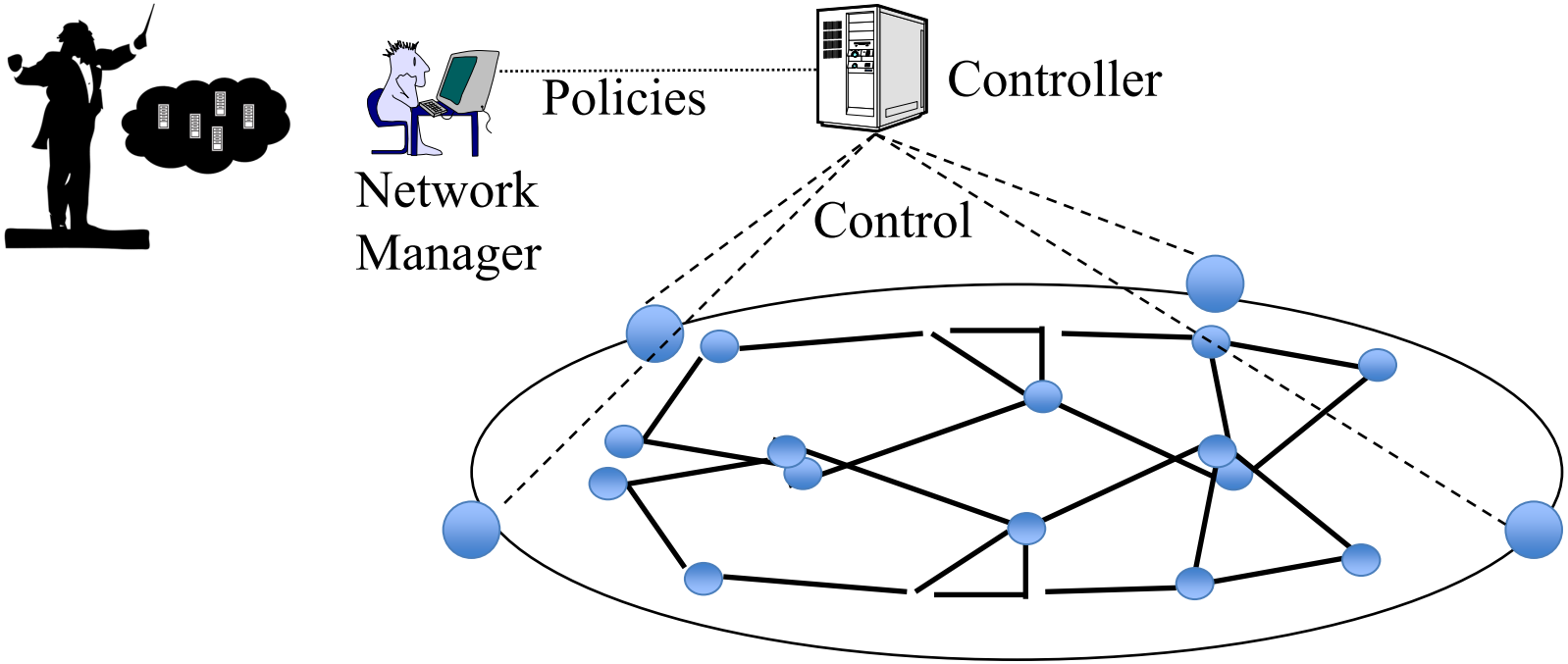
Virtualizing the Global Storage

- Energy Science Network (ESNet) uses virtual switch to connect members located all over the world
- Virtualization ⇒ Fluid networks ⇒ The world is flat ⇒ You draw your network ⇒ Every thing is virtually local



Ref: I. Monga, "Software Defined Networking for Big-data Science," http://www.es.net/assets/pubs_presos/Monga-WAN-Switch-SC12SRS.pdf

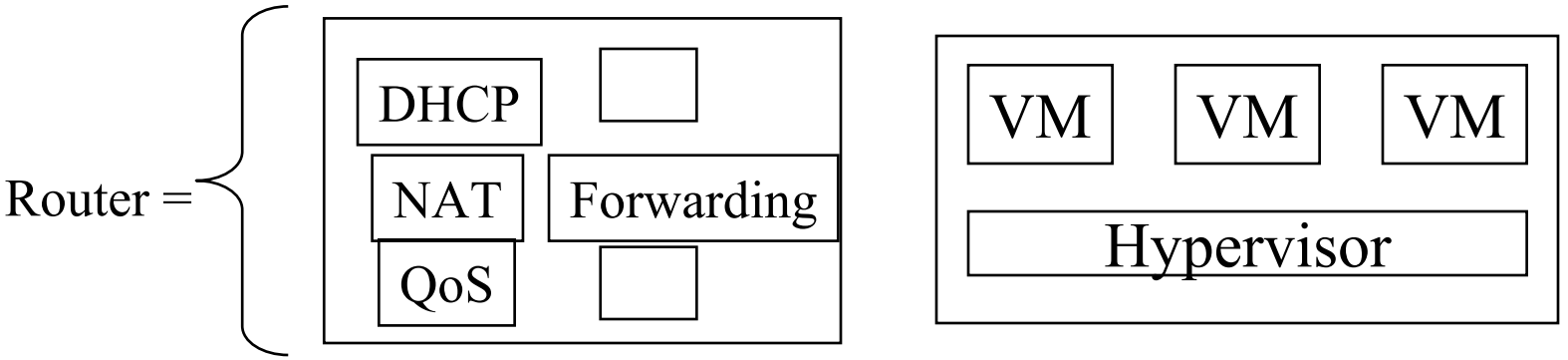
Software Defined Networking



- Centralized Programmable Control Plane
- Allows automated orchestration (provisioning) of a large number of virtual resources (machines, networks, storage)
- Large Hadoop topologies can be created on demand

Network Function Virtualization (NFV)

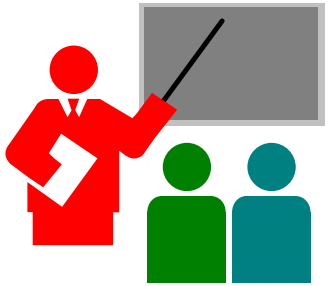
- Fast standard hardware ⇒ Software based Devices
  Virtual networking modules (DHCP, Firewall, DNS, …)
  running on standard processors
- Modules can be combined to create any combination of function for data privacy, access control, …
- Virtual Machine implementation ⇒ Quick provisioning
- Standard Application Programming Interfaces (APIs)
  ⇒ Networking App Market
  ⇒ Privacy and Security for Big data in the multi-tenant clouds



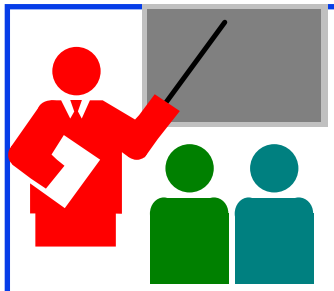
Big Data for Networking

- Today's data center:
  - Tens of tenants
  - Hundreds of switches and routers
  - Thousands of servers
  - Hundreds of administrators
- Tomorrow:
  - 1k of clients
  - 10k of pSwitches
    $\Rightarrow$ 100k of vSwitches
  - 1M of VMs
  - Tens of Administrators
- Need to monitor traffic patterns and rearrange virtual networks connecting millions of VMs in real-time
  $\Rightarrow$ Managing clouds is a real-time big data problem.
- Internet of things $\Rightarrow$ Big Data generation and analytics

Summary



1. Less than 1% of things are connected
   ⇒ IoT is a big opportunity for academics and industry
2. Smart Grid and Energy management is leading the change.
3. Smartness comes from communication capability since the computation can be delegated
4. Right at the knee: Academic and Startup Research opportunities in almost subfields of computing including hardware development, data analytics, security, and networking.
5. Cloud computing everywhere leads to fog computing and multi-cloud computing ⇒ AppFabric



Summary

1. I/O virtualization allows all storage in the rack to appear local to any VM in that rack ⇒ Solves the co-location problem of MapReduce
2. Network virtualization allows storage anywhere in the data center or even other data centers to appear local
3. Software defined networking allows orchestration of a large number of resources ⇒ Dynamic creation of Hadoop clusters
4. Network function virtualization will allow these clusters to have special functions and security in multi-tenant clouds.