Availability and Policy

Clement Chen and Craig Lewis

Agenda

- Availability Definition
- The causes of disruption
- Aspects of Availability& How to achieve availability
  - Data
  - Network
  - Communication
  - IT system
  - Power
  - Humans
- Measurement
- BCDR
- Policy Analysis

Things to think about

- What does “availability” mean?
- What does “availability” mean in the case study?
- Developing Availability Policy

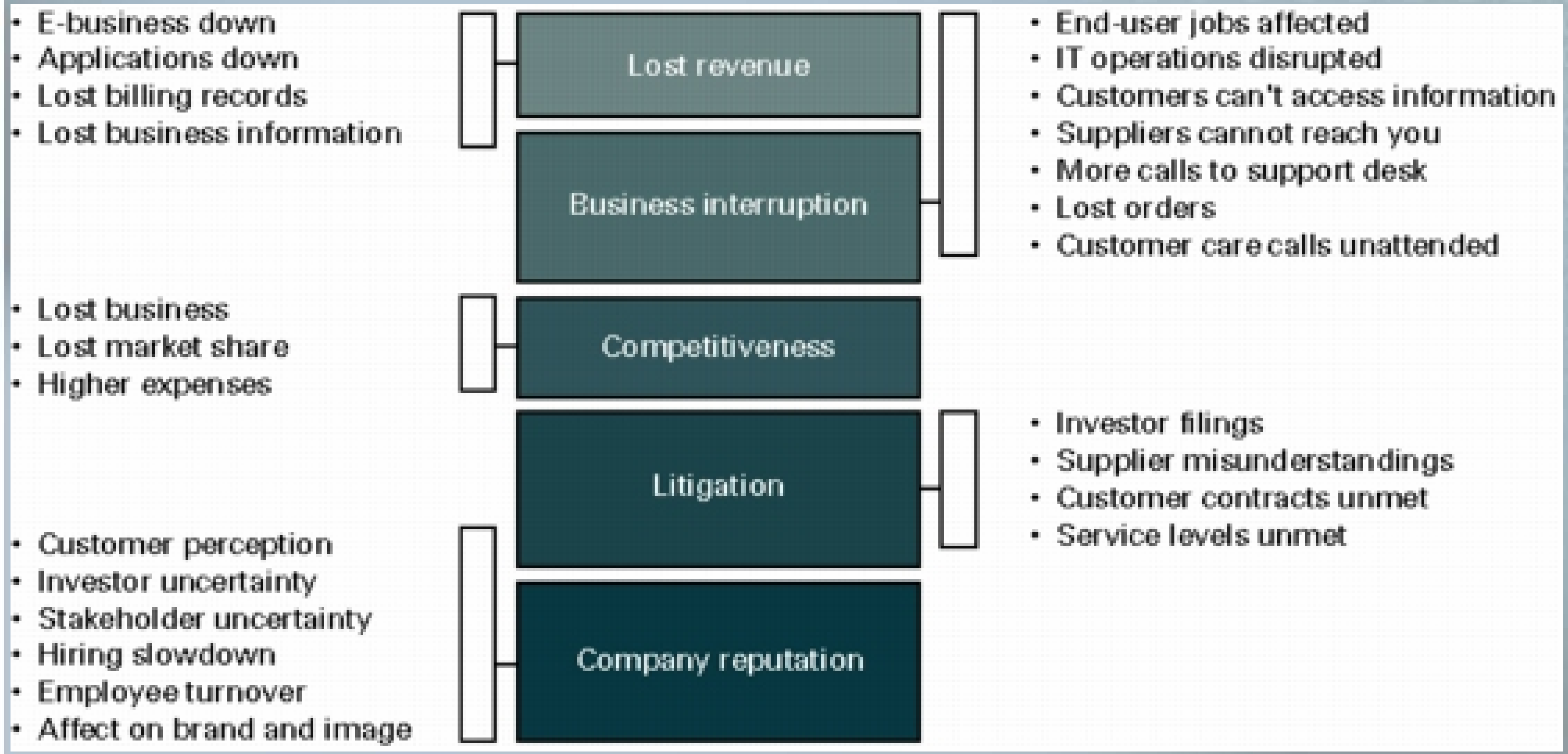
Availability: Definition

- The degree to which data or systems are accessible and in functioning condition.
- Looking at it another way, the degree to which the system is fulfilling the intended function.

Availability vs. Reliability

- Availability and Reliability are not the same thing.
- Availability means that the system is ready for use.
- Reliability means that a device or system can perform its job when called upon to do so.
- There is overlap but they are not the same thing.

Impact of Disruption to Availability



- E-business down
- Applications down
- Lost billing records
- Lost business information

Lost revenue

Business interruption

- End-user jobs affected
- IT operations disrupted
- Customers can't access information
- Suppliers cannot reach you
- More calls to support desk
- Lost orders
- Customer care calls unattended

- Lost business
- Lost market share
- Higher expenses

Competitiveness

Litigation

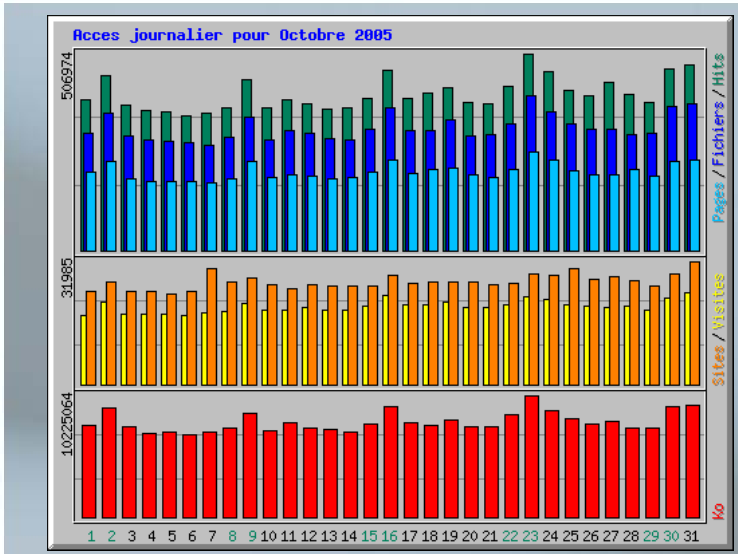
- Investor filings
- Supplier misunderstandings
- Customer contracts unmet
- Service levels unmet

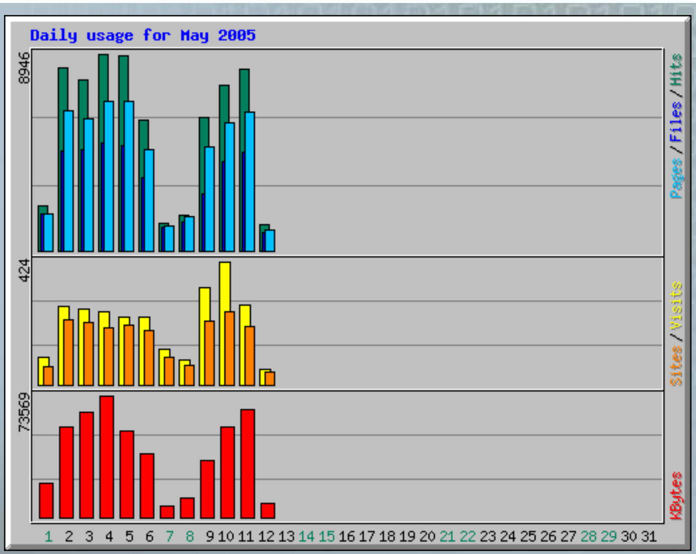
- Customer perception
- Investor uncertainty
- Stakeholder uncertainty
- Hiring slowdown
- Employee turnover
- Affect on brand and image

Company reputation

- Source: http://www.cisco.com/en/US/netsol/ns206/networking_solutions_white_paper09186a008015829c.shtml

Interruption to Availability

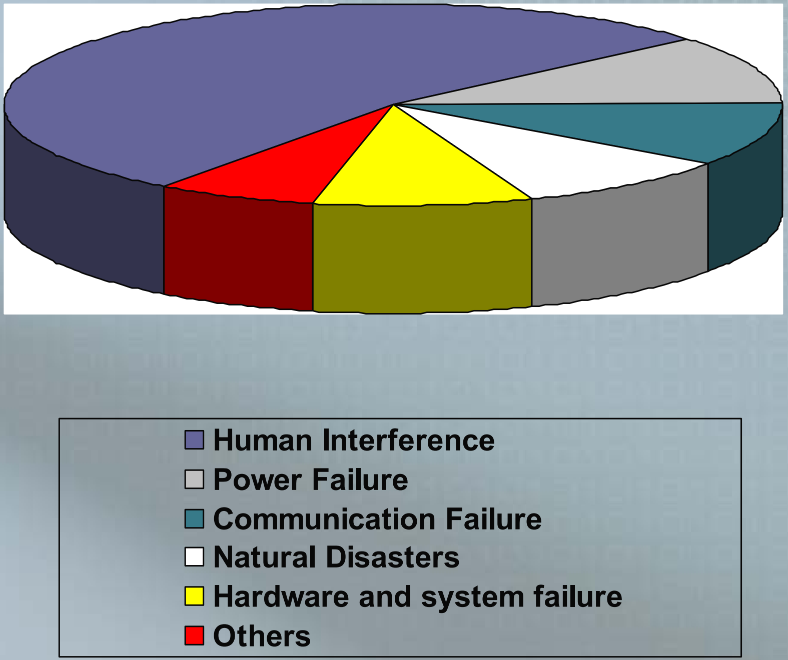




- What do you notice about the two graphs?
- Can you tell which server has higher availability?
- Can you tell *why* one server has higher availability?

Major Causes of Disruption

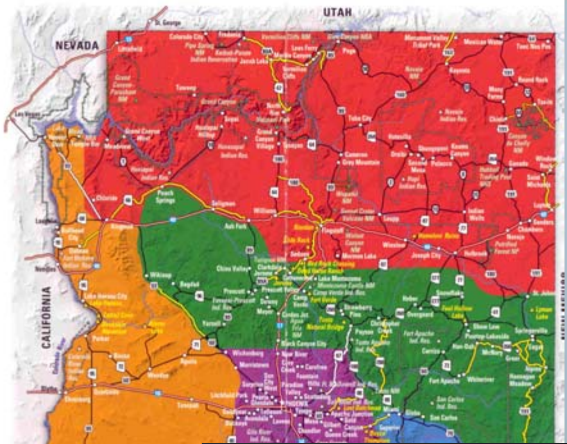
- Human Interference
  - Operator error;
  - Virus and hacker attack;
  - Theft or sabotage;
  - Terrorism (post 9/11);
  - ………
- Communication Failure
- Hardware or system failure
- Natural Disasters
- Power Failure
- Water Damage
- Fire
- ……

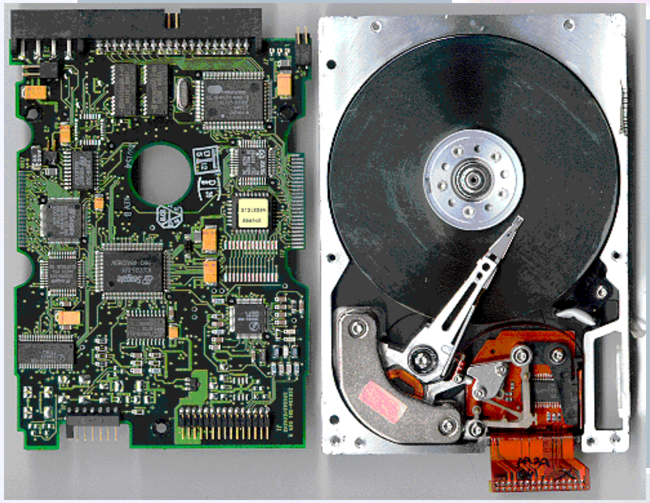


Source: Accenture and Gartner

Magnitude of Disruption

- Regional
- Metropolitan
- Building
- System
- Component



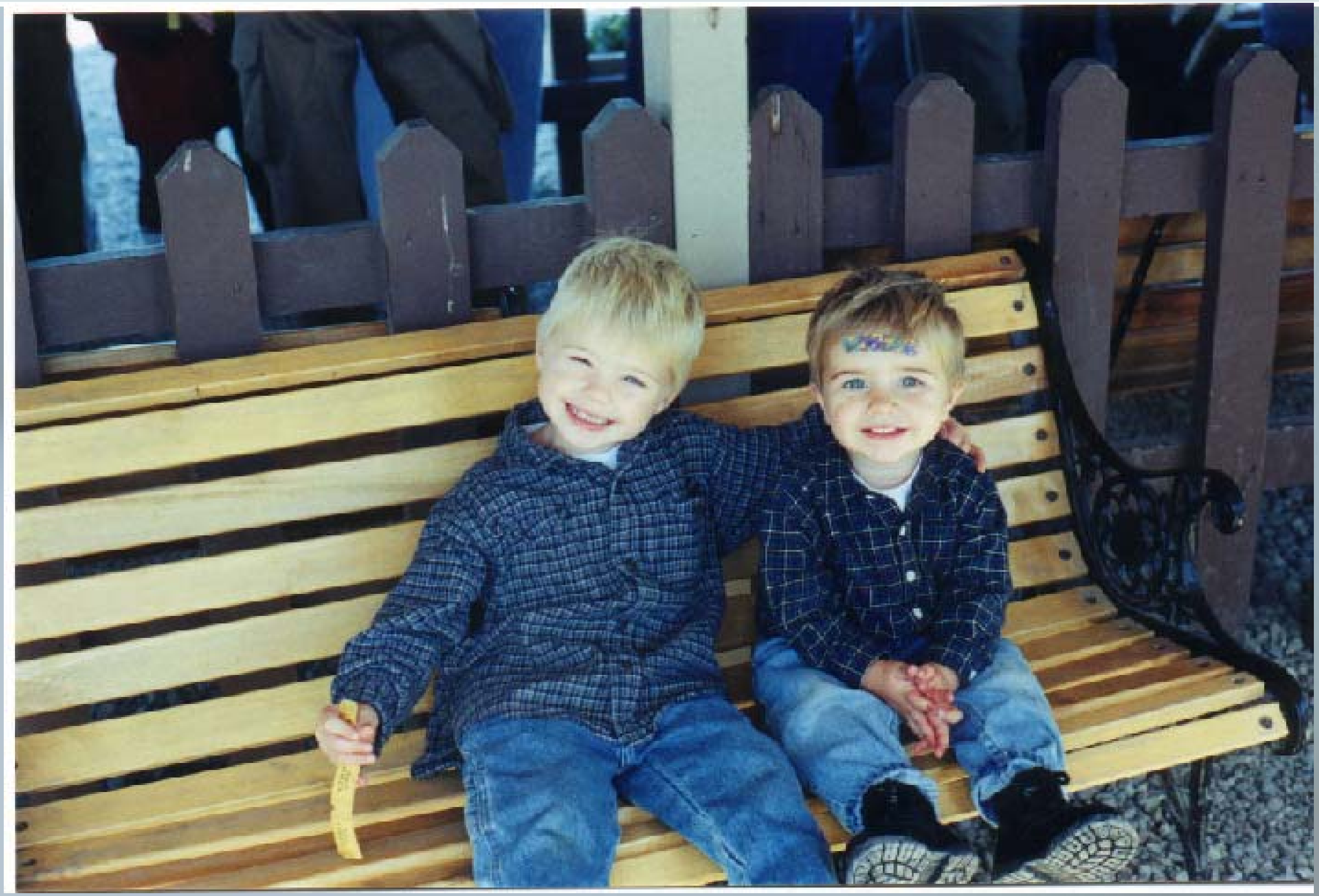


Aspects of Availability& How to achieve availability

- Data Availability
- Network Availability
- Communication Availability
- System Availability
- Power Availability
- People Availability
- Other Resources Availability

Data Availability

How important is data?



“You should protect your data like you would your children”

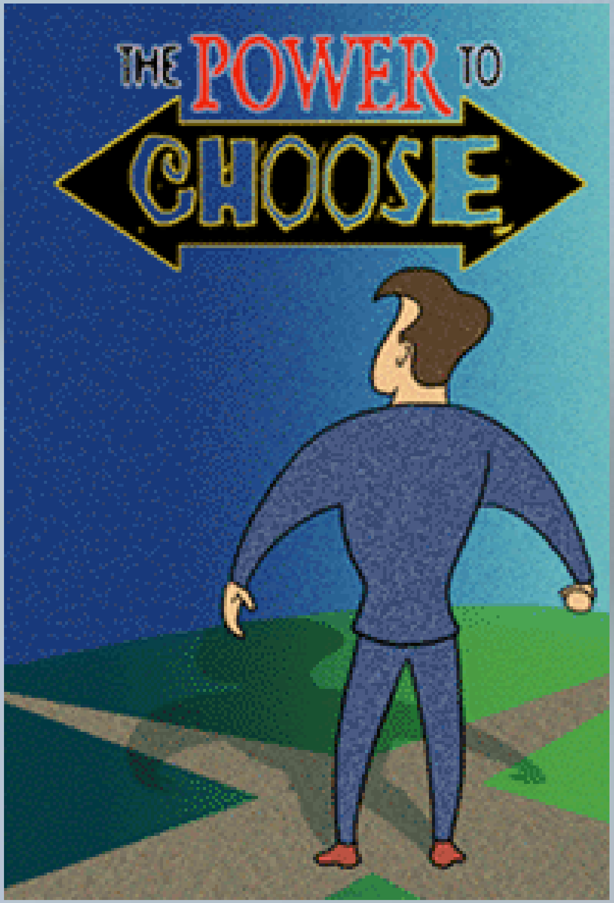
-San Jose Mercury News, January 2002

Data Retention - Legal Implications….

- **Sarbanes Oxley**
  - **All electronic company information must be retained for at least five years.**
  - **Accounting firms that audit publicly traded companies must retain all related documents for 7 years after audit.**
- **HIPPA**
  - **Members of health care industry must retain patient information for 6 years**
- **SEC 17a-3 and 17a-4**
  - **Brokers/dealers must retain records for 3-6 years and more**
- ……………………………….

Quick Survey

- If you have to choose between losing your laptop and your data, which would you choose?



How to Achieve Data Availability

- Rule #1: Backup !
- Rule #2: Backup !!
- Rule #3: Backup !!!

Common backup methods

- Full Backup
  - Backup every file
  - Takes a lot of storage space
- Incremental Backup
  - backs up files that have been created or modified only since the last backup;
  - backup operator needing several tapes to do a complete restoration
- Differential Backup
  - backs up files that have been created or modified only since the last full backup
  - backup operator need only the full backup and the one differential backup to restore the system.

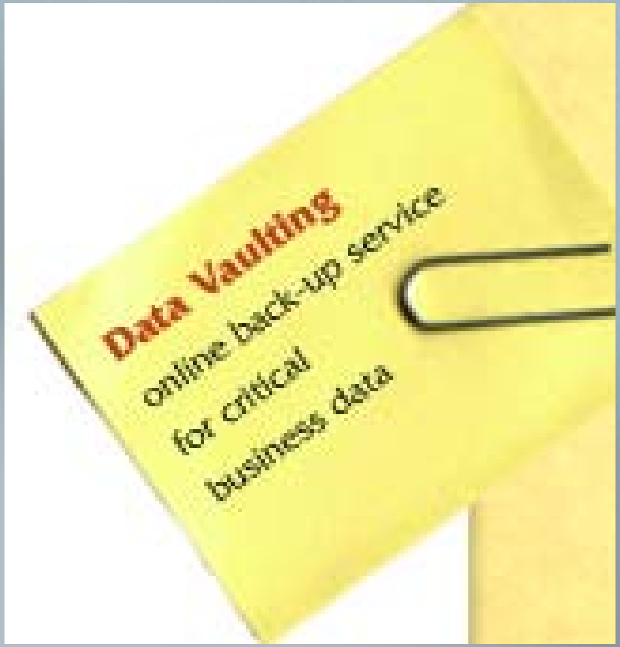
Let’s talk a little about data backup technologies ……………..

- Tape (offline-storage)
  - Pros: Typically the least expensive medium
  - Cons: Longest data recovery times
- ATA-based storage systems (near-line storage)
  - Pros: Relatively quick data recovery times, cheaper than fiber-channel and SCSI systems
  - Cons: May lack performance and reliability characteristics of Fiber Channel- and SCSI-based systems
- Fiber Channel- and SCSI-based systems (online storage)
  - Pros: Very Fast data recovery
  - Cons: Very expensive medium

All these technologies can only sustain disruptions that occur at a small range …….

Data Vaulting

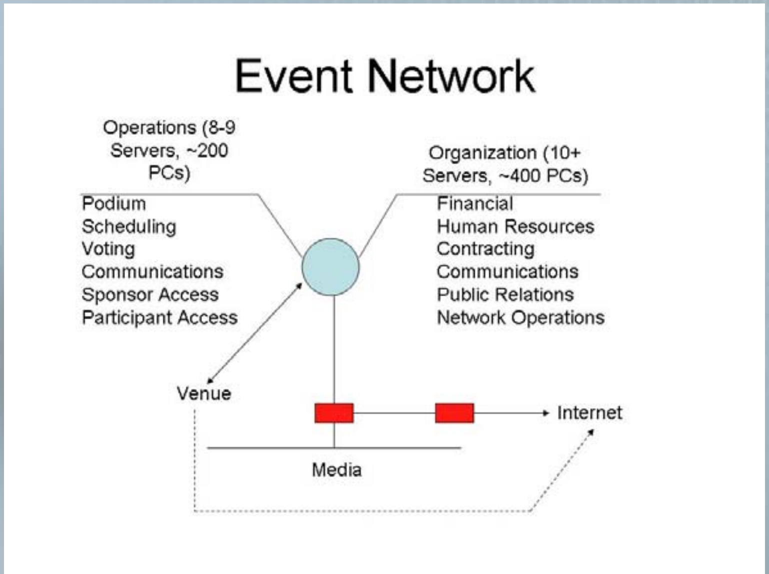
- Copy of data is saved at a remote site
  - Periodically or continuously, via network
  - Remote site may be own site or at a vendor location
- Minimal or no data may be lost in a disaster
- There is typically some delay before data can actually be used



Network Availability

Availability Decisions

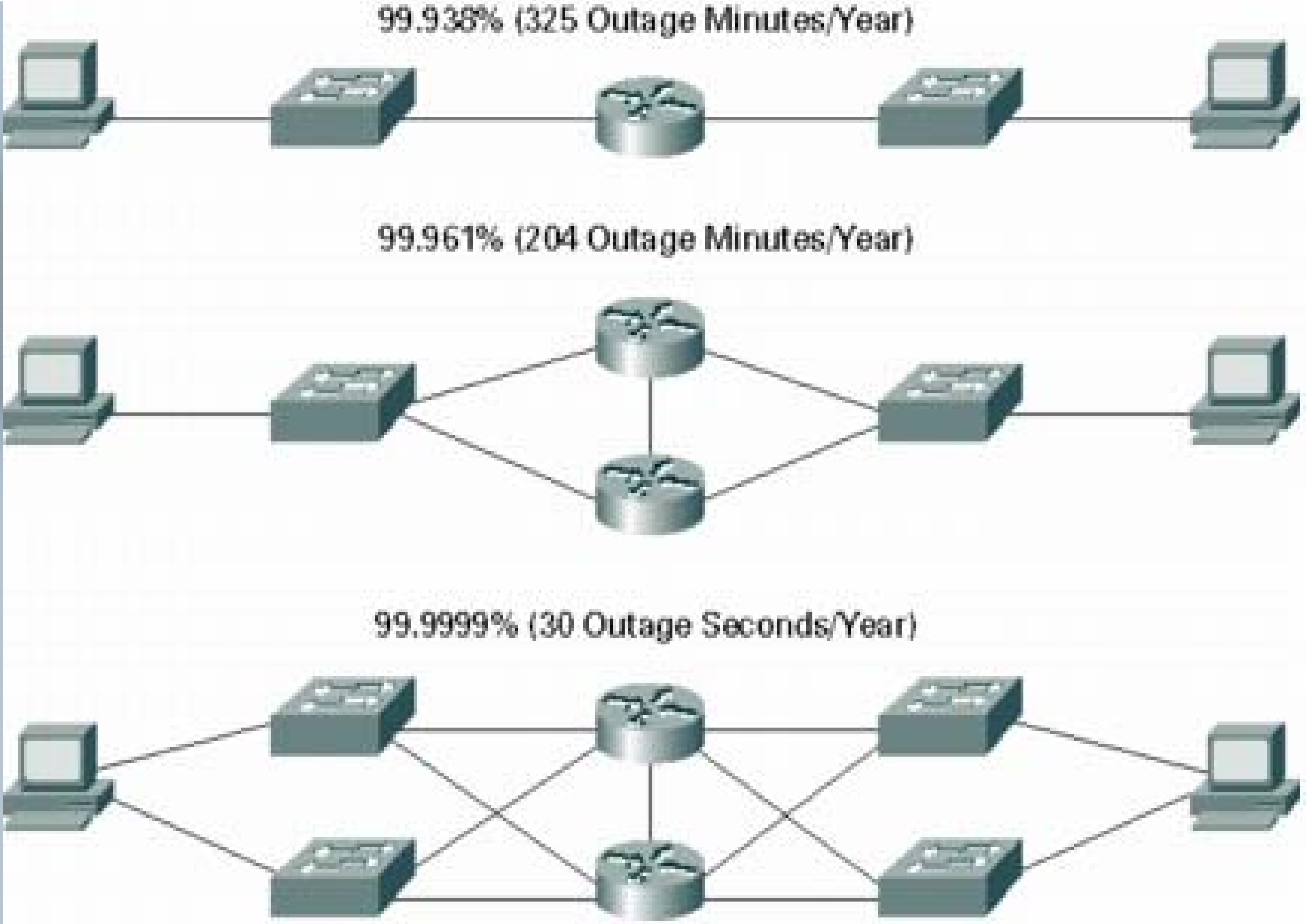
- Availability based on component
  - Teleprompter
  - Voting System
  - Sponsor Access
  - Participant Access
  - Others
- Who approves the SLA for availability?



Network Availability

- Prioritize the systems needing network access
- Measure the amount of bandwidth needed to fulfill purpose of each component
- Calculate overhead of protective measures.
- Decide what (if anything) can drop

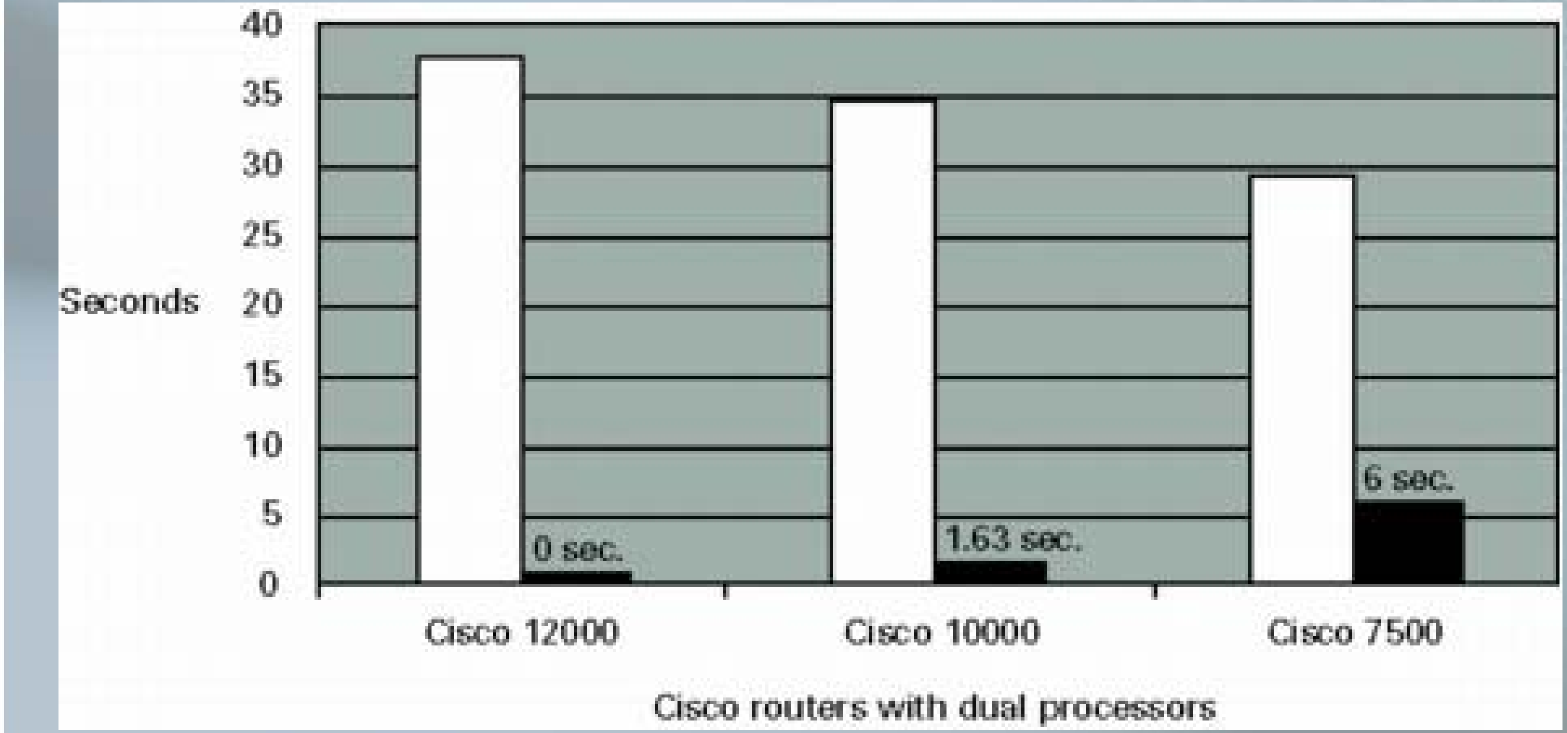
Architecting High Availability


99.938% (325 Outage Minutes/Year)
99.961% (204 Outage Minutes/Year)
99.9999% (30 Outage Seconds/Year)

- Source: http://www.cisco.com/en/US/netsol/ns206/networking_solutions_white_paper09186a008015829c.shtml

Meeting your SLA

- Can your equipment handle it?

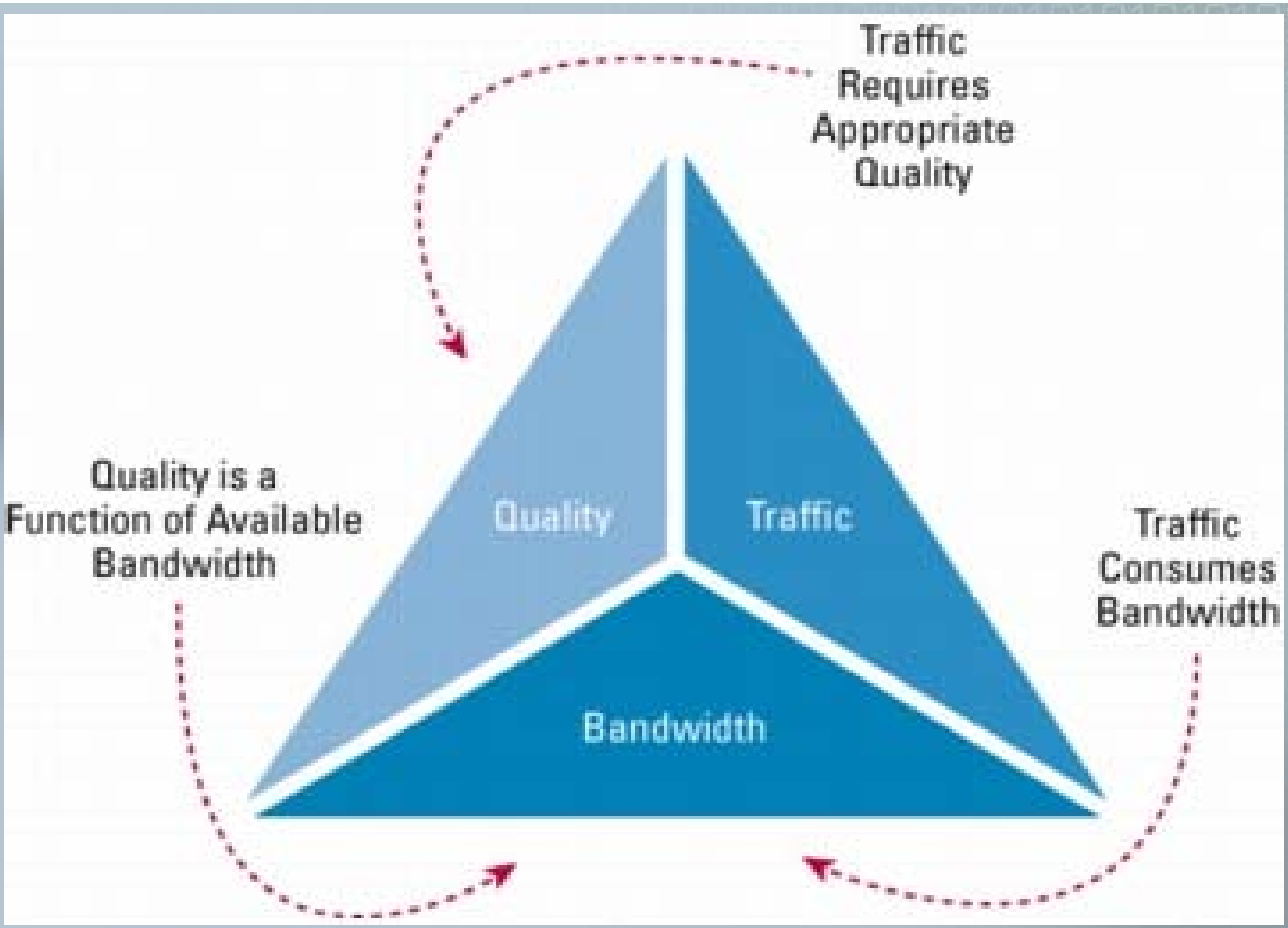


- Source: http://www.cisco.com/en/US/netsol/ns206/networking_solutions_white_paper09186a008015829c.shtml

Threats to Network Availability

- DDoS
- Flash traffic
- Component failure
- Misconfiguration / Inefficient configuration
- Legitimate use which network can’t handle. Capacity underestimated by network designers.

Determining Bandwidth


Traffic Requires Appropriate Quality
Quality is a Function of Available Bandwidth
Quality
Traffic
Traffic Consumes Bandwidth
Bandwidth

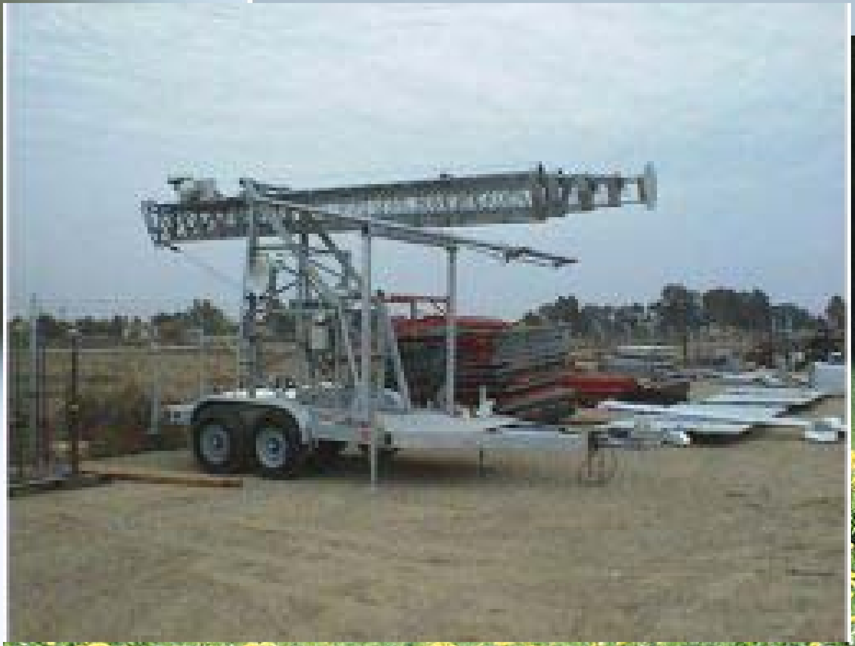
- Source: http://www.cisco.com/en/US/products/ps6558/products_white_paper0900aecd8024d42d.shtml

Solutions

- Overprovision
- Reduce bottlenecks
- Less rich content
- Content provider (e.g. Akamai)
- QoS Prioritization of data on network

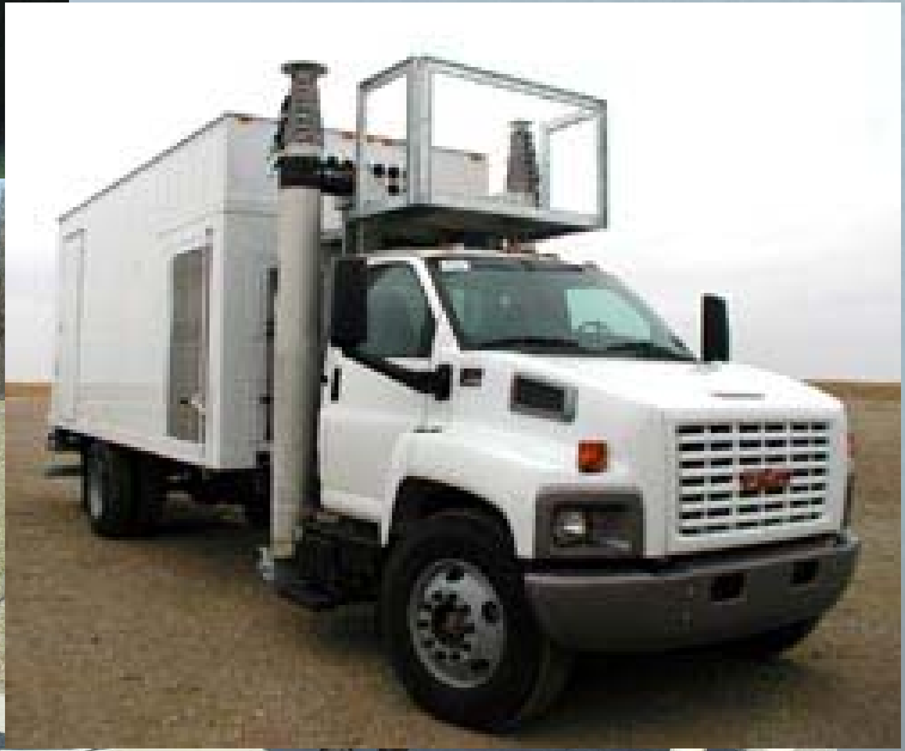
Technologies for Communication Availability

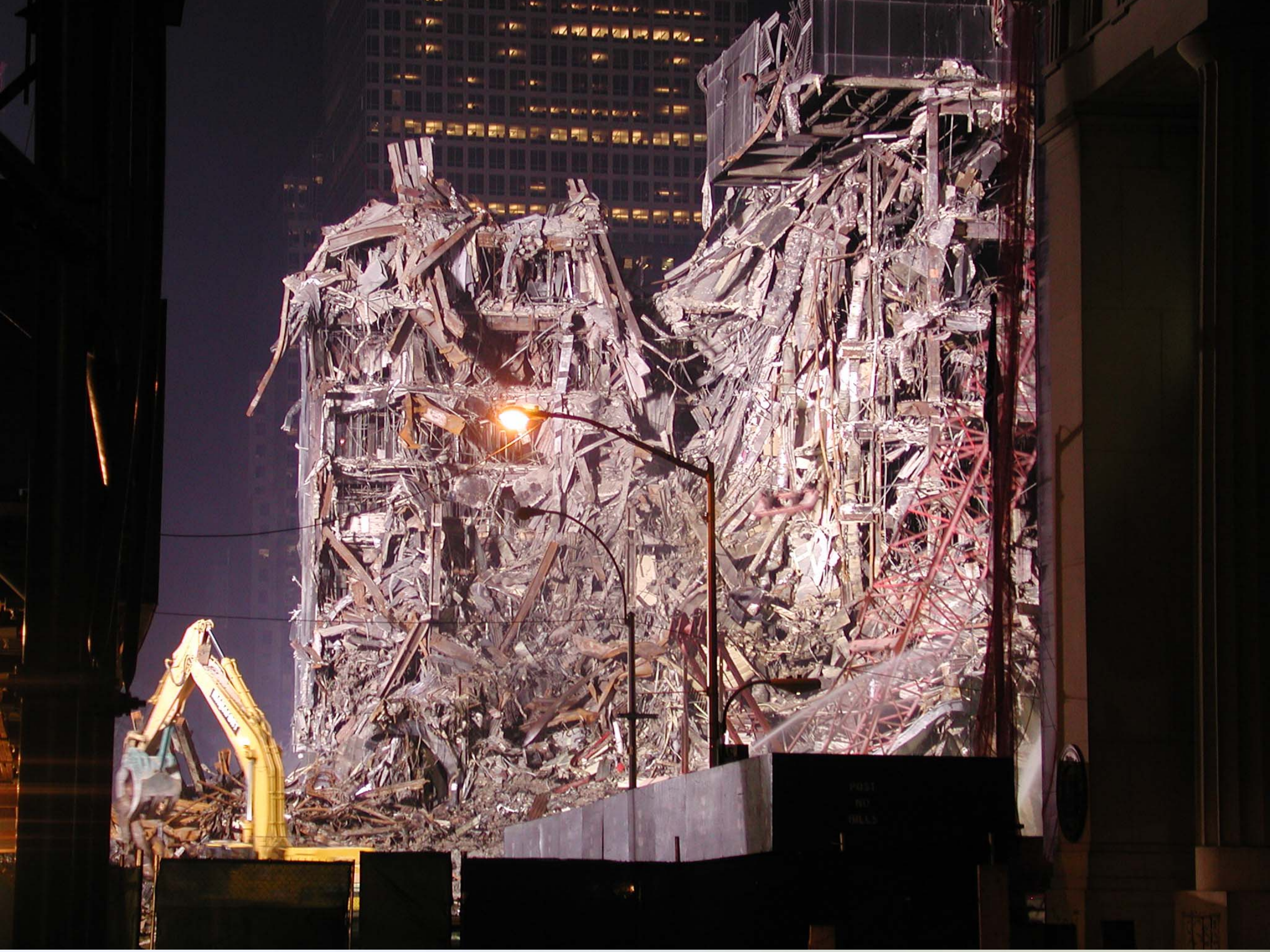
COW and COLT



COW (Cell on Wheel)

COLT (Cell on Light Truck)




POST
NO
BILLS

Communication Recovery - 9/11

21 COWS

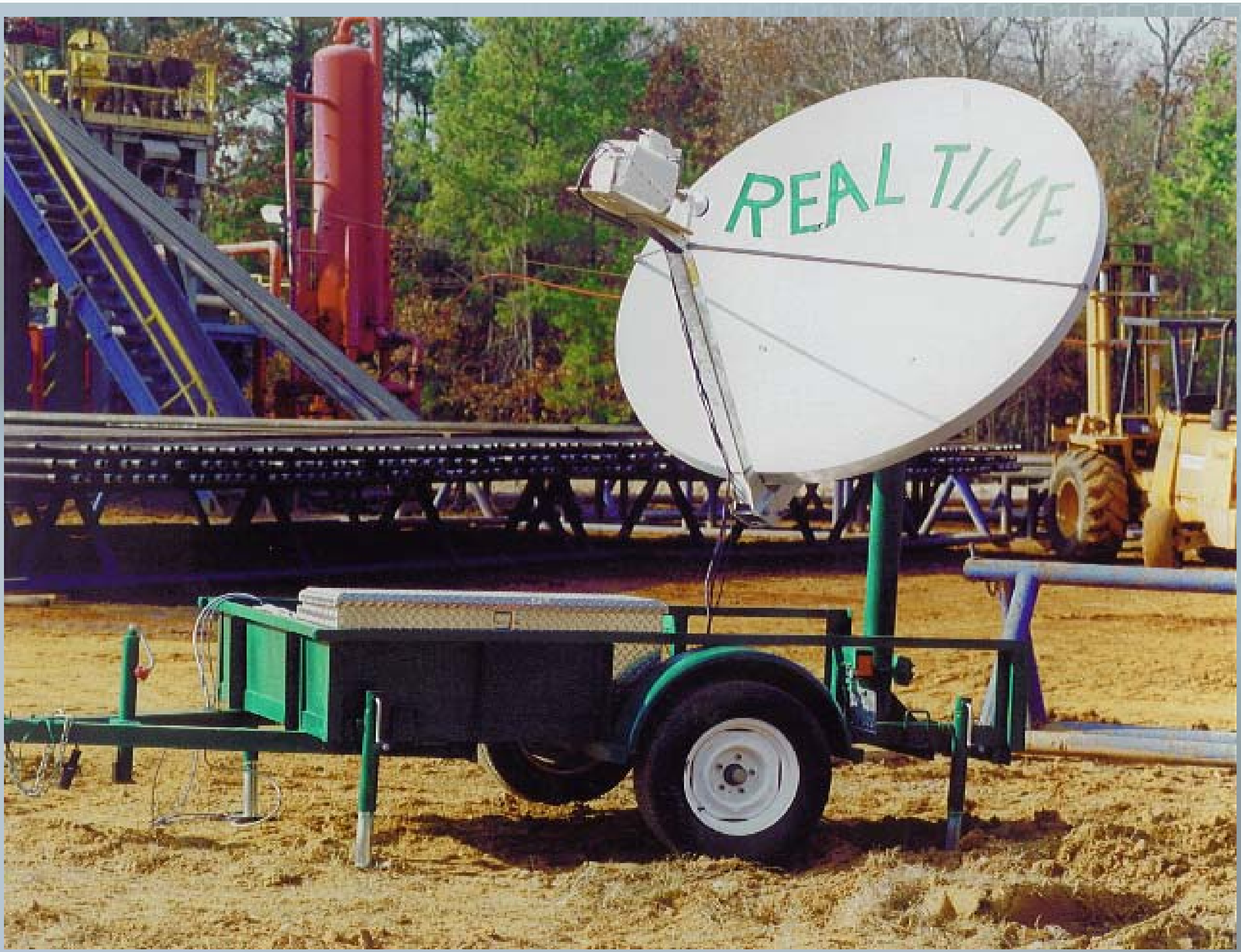
15 COLTS

Source: Lucent Technologies

Is Satellite an Option?

Real Time Communication, www.rtc-vsat.com

Real Time puts telephone voice service, fax, email, and data capability on remote land-based sites and offshore locations.



Is Satellite an Option? (Cont)

Service: Mobile Satellite Communications

Bandwidth: 1-2 Voice connections, 1 64k data connection

Connectivity: Satellite link is established from customer site to Real Time Communications(RTC) Hub. Voice connections are charged from McLain VA to the end-point. Charges can be billed back to the customer through their selected long distance service provider. The data connection can be established directly to the internet.

Equipment Connections: Voice and data connections are through an access box at the back of the Control unit. Voice is connected through a standard RJ-11 connection to a telephone. Data is connected through a standard Ethernet connection.

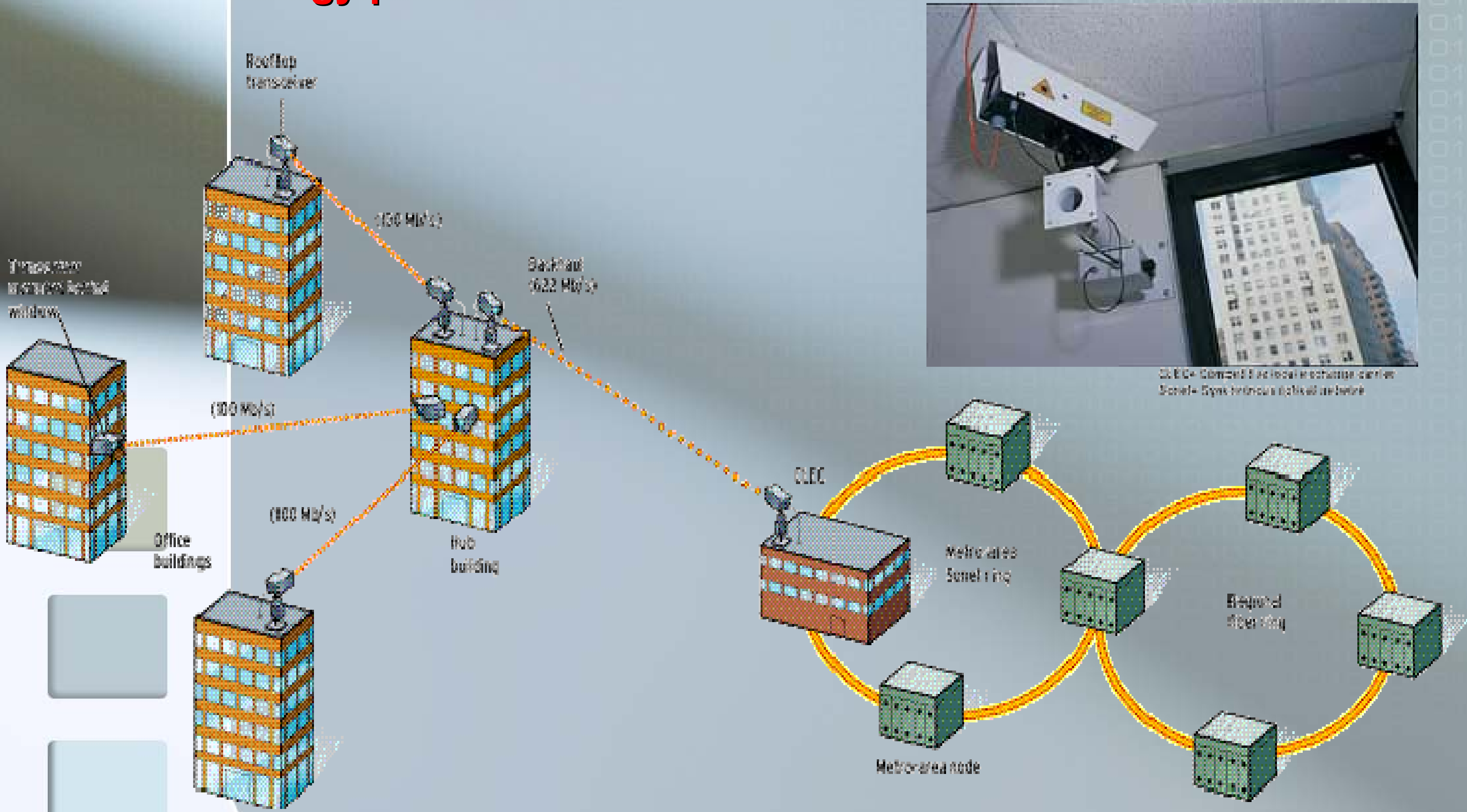
Power requirements: 110VAC at approx. 15Amps. Mobile units contain a UPS.

Deployment: Upon activation by the client, RTC will deploy qualified technicians with either a mobile satellite package on a trailer or a mobile satellite package which uses a 4 legged non-penetrating mount. At the client site the technicians will coordinate with site personnel an optimum set-up location and set-up and test connectivity. They require 110VAC power and phone communications to contact their Hub to verify the set-up. They have cell phones which may be used if they are within coverage area of their provider.

Once set-up and testing is complete the unit may not be moved without re-initiating the set-up and testing procedure by the technicians.

Lasers for Communication Availability

Technology proven in the WTC disaster



Lasers for Communications Availability(Cont)

- Technology using laser beams over short distances that can be set up within 24 hours and currently provide OC12 data rate.
- Very reliable
- Fast, easy setup
- One time capital charge or monthly charge
- Distance limits of 2.5 miles. Closer is better
- Environmentally and personally safe
- Reference:

LightPointe www.lightpointe.com

Terabeam www.terabeam.com

System Availability

- Software Reliability
  - How to write a reliable software?
  - A huge topic ………
- Hardware Reliability
  - Redundant equipments, standby, …
  - Cluster

How Google Achieves High System Availability

- The Google Infrastructure Challenge
  - Indexed >8 billion web pages;
  - Appx. 40 million searches/day;
  - Storage capacity >5 petabytes;
  - Gmail has millions of users, each user's storage > 2GB;
  - 60s*60m*24h*7d*365 availability;
  - cannot lose people's email;
  - Traffic growth 20-30%/month.
  - Capital and operating costs at fraction of large scale commercial servers;

The Google™ Infrastructure

- GFS (Google File System)
  - GFS replicates user email in three places; if a disk or a server dies, GFS can automatically make a new copy from one of the remaining two. Compress the email for a 3:1 storage win, then store user's email in three locations, and their raw storage need is approximately equivalent to the user's mail size.
- Google Clusters
  - 100,000+ commodity-class PCs
  - Running custom fault-tolerant software
    - Replication of services across many different machines
    - Automatic fault detection & handling

Dimensions of a Google Cluster

- 359 racks
- 31,654 machines
- 63,184 CPUs
- 126,368 Ghz of processing power
- 63,184 Gb of RAM
- 2,527 Tb of Hard Drive space

More Technical Details:
L.A. Barroso, J. Dean, H. Holze: “Web Search for a Planet: The Google Cluster Architecture”, IEEE Micro, 2003
S. Ghemawat, H. Gobioff, S.T Leung: “The Google File System”, Proc. ACM SOSP, 2003

Other Important Availability Concerns

- Power, HVAC, other system components
- People and resources

Availability of and to the Infrastructure

- Availability of the infrastructure can have a direct impact on availability of information

- Voice communications
- Power
- HVAC
- Physical access

Solutions

- Voice
  - Cellular Phones (remember COWS and COLTS)
  - WiFi Phones
  - Walkie-talkies
- Power
  - Uninterruptible Power Supply (UPS)
  - Generators
- HVAC
  - Portable coolers
- Physical Access
  - Security guards
  - Transportation shuttles
  - Backup/alternative to electronic access controls

Resource Availability

- Availability doesn’t just refer to the computers:

- Schedules (print and electronic)
- People being where they need to be
- People knowing what they need to know
- Elevators, wheelchairs, access for disabled, escorts and concierge
- International event – signs, handouts, translation, etc.

People and Availability

- People are a source of information.
- Staff with knowledge of how to fix a problem not being there to fix it negatively impacts availability.
  - Positional redundancy – *"Worker X can do that, but she's not here until tomorrow."*
  - Shared knowledge – *"What if I get hit by a bus?"*
  - Limitations on physical access – *"It's a 30 second fix, but it will take me 10 minutes to get there."*
  - Limitations placed by policy – *"I know how to fix it, but I'm not allowed to go in the server room."*

People and Availability

- Will your workers ever…?
  - Become sick
  - Know enough
  - Get injured
  - Have a family emergency
  - Slack

Source: http://imageserver1.textamerica.com/user.images.x/37/IMG_425837/_0930/TZ200930004832874.jpg

People and Availability

- Problems affecting availability:

- The event from the case study has a limited IT budget.
- The organizers want most of the money to be spent on the “visible.”
- Limited time and staff controls.
- Too much work to do and not enough time to do it.

People and Availability

- Solutions:

- Shift some work to sponsors. E.g. “The Summer Games powered by IBM”
  - Vendor will have some interest in getting things right and bring expertise.
  - Can bring extra labor and equipment.
- Convince* the organizers that dedicating resources to availability is for the good of the event. There is a value and return on focusing on availability.

* may be easier said than done. ☺

Measuring Availability

Measuring availability

- What does it mean to be available and how can it be measured?
- Availability means that systems or data are accessible but does not guarantee:
  - Performance
  - Typical ways of doing things can still be used
  - Full system capacity

Measuring Availability: Uptime

- Definition:
  - Mean Time Between Failure (MTBF) is the amount of time between failures, where failure is defined as a departure from acceptable service for a system. This is a measure of reliability.
  - Mean Time to Recover (MTTR) measures the amount of time required to repair or recovery for a failed system.
  - Availability is the ratio of the time a system is actually available to the time it should have been available.
  - Availability = MTBF / (MTBF + MTTR)

Measuring Availability: Uptime

- For the case study, let’s say that a system needs to be available for one week

Availability %	Downtime per week
98%	3.4 Hours
99%	1.7 Hours
99.9%	10.1 Minutes
99.99%	1.0 Minute
99.999%	6.0 Seconds
99.9999%	.6 Seconds

Measuring Availability: Uptime

- Measuring availability for 1 week from 8:00 am – 11:00 pm

Availability %	Downtime per week
98%	2.1 Hours
99%	1.1 Hours
99.9%	6.3 Minutes
99.99%	37.8 Seconds
99.999%	3.8 Seconds
99.9999%	.38 Seconds

Business Continuity/Disaster Recovery

Business Continuity/Disaster Recovery

- **Business Continuity: Ability to maintain the constant availability of processes and information across the business enterprise**

- **Disaster Recovery: Immediate and temporary restoration of computing and network operations after a natural or manmade disaster *within defined timeframes***

Source:

Lucent Technologies

Need Continuity of Business, not just Availability of IT System

Defining a BCP

- Every Business Continuity strategy includes three fundamental components:
  - Business Impact Analysis
  - Recovery Strategy
  - Design and Develop the disaster recovery process
- BCP should consider every type of interruption from a brief power outage up to the worst possible natural disaster or terrorist attack

Why BCDR is Important?

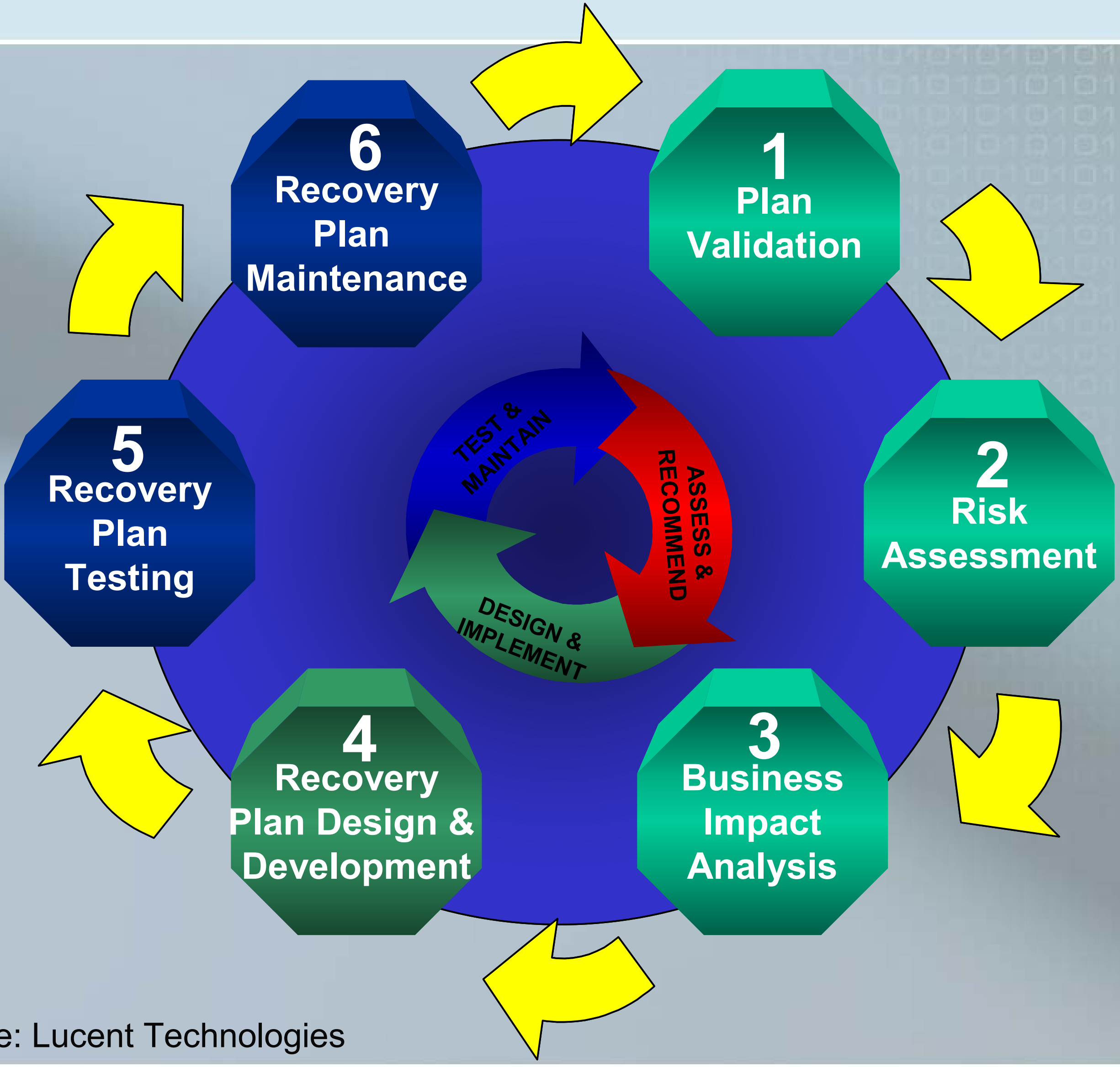
- 80% of companies having an extended disaster are out of business within five years.[1]
- 50% of companies having a disaster without a plan go out of business within two years.[2]
- 29% of companies with a major disaster will close within two years; 43 percent never reopen.[3]

1 University of Minnesota
2 IBM Business Recovery Service
3 DATAPRO

Requirements of a BCP

1. Provide procedures and listing of resources to assist in the recovery process.
2. Provide an immediate, accurate and measured response to emergency situations.
3. Identify vendors that may be needed in the recovery process and put agreements in place with selected vendors.
4. Avoid confusion experienced during a crisis by documenting, testing an training plan procedures.
5. Clear guidance for declaring a disaster
6. Provide the necessary directions to ensure the timely resumption of critical services
7. Document recovery processes so they can be executed by knowledgeable people

Stages of Business Continuity Planning



Source: Lucent Technologies

Business Impact Analysis (BIA)

The BIA is a functional analysis that identifies the impacts should a outage occur. Impact is measured by the following:

- Allowable Business Interruption – the Maximum tolerable downtime
- Functional and Operational Considerations
- Regulatory Requirements
- Organizational Requirements

The BIA sets the stage for determining a business-oriented judgment concerning the appropriation of resources for recovery planning efforts.

Source: The CISSP Prep Guide

BCDR Resources

- Survive: The Business Continuity Group

  http://www.survive.com/

- Emergency Information Infrastructure Partnership

  http://www.eiip.org/

- Disaster Recovery Journal

  http://www.drj.com/

Policy Analysis

- Review first draft of policy relating to systems in the case study
- Discuss shortcomings of draft policy
- Make recommendations for improvement of the policy
- Get feedback from the event directors (a.k.a. the professors)

Voting System Availability Policy

- The Voting systems will be available with 98% uptime for the conference.
- The organization IT group shall provide a number of voting systems commensurate with the number of attending voters. There will be 1 voting station for every 25 delegates.
- Use of a voting system requires the use of a smart card. The voting system will automatically run a voting application upon smart card insertion. The system will then tally the vote and lock the system when the smart card is removed.
- There shall be 4 clusters of voting stations placed at various locations throughout the event venue. These clusters must be located in areas which are accessible to voting delegates.

Network Availability Policy

- Switches and routers will use Inter-Switch Link protocol to maintain VLAN information as traffic flows between devices. This will allow for better network security and flow monitoring.
- All external network traffic will flow through the main Internet connection. The venue network connection is only to be used for failover in the case of an outage.
- All traffic over the internal network will be encrypted.

Questions