the Availability Digest

www.availabilitydigest.com

Help! My Data Center is Down!

Part 4: Intranet Outages

January 2012

Today's data centers are incredibly complex. No matter how much redundancy data-center designers build into their infrastructures, things fail. Data centers fail, often with disastrous consequences. Sometimes, a fault will take down a data center for hours. Sometimes, a fault will take down a data center for days.

In our previous articles, we related several major data-center disasters caused by power failures, storage subsystem faults, and in our last article, Internet outages.

A data center is no good to anyone if it cannot be accessed by its users. The Internet outages we described in our last article were faults external to the data centers and that seriously impacted their operations. But data centers also rely heavily on internal networks to interconnect their servers, to connect internal users, and to provide connectivity to the external Internet. These internal networks are called *Intranets*.

In this article, we explore some notable Intranet failures that rendered data centers useless even though they were otherwise fully operational. The stories are taken from the Never Again archives of the *Availability Digest*.

Intranets

A large data center can comprise hundreds or thousands of servers. These servers must not only be interconnected, but they must also connect to local users and to the Internet. Today's internal networks use Internet technology and are called Intranets. Large Intranets are built with a great deal of redundancy to ensure that they provide reliable service. Unfortunately, Intranets can fail. When they do, company operations can be severely impacted.

Network Problem Takes Down Johannesburg Stock Exchange

An undisclosed network problem shut down the Johannesburg Stock Exchange in South Africa for most of the day on Monday, July 14, 2009. The problem prevented the Exchange from disseminating trade data from its morning opening until late in the afternoon. Affected were the Stock Exchange News Service, equity trading, and equity derivatives. It was estimated that billions of rands in trades were lost. The Exchange reports that its network had an availability of 99.6% over the last six years. Is that an availability of which to be proud?

NIC Closes Dublin Airport

Aircraft position and identification information started to intermittently disappear from controllers' screens in early June, 2008, as aircraft departed and approached Dublin Airport, Ireland's busiest air-traffic hub. The problem continued through early July, when it finally became so bad that controllers shut down the air traffic control system on Wednesday afternoon, July 16th, and rerouted planes to other airports. The problem was ultimately traced to a faulty Network Interface Card (NIC) providing radar data to the control systems. There was no backup link. Tens of thousands of passengers were left stranded or delayed over a several day period.

Router Failure Delays Airing of ALCS Game

With U.S. sports bars crowded for baseball's American League Championship Series (ALCS) Game 6 between the Tampa Bay Rays and the Boston Red Sox on October 18, 2008, howls of anger were raised when TV sets showed instead the Steve Harvey show. It turned out that circuit breakers for Turner Broadcasting System's master router and its backup in Turner's Network Operations Center in Atlanta both independently tripped, causing routing of the live feed to fail. By the time the game went live, the Rays were ahead 1 to 0. The Tampa Bay Rays went on to win Game 7 and the American League slot in the World Series, but lost to the Philadelphia Phillies in the Series.

Network Problem Costs UK Punter £1 Million

A network fault took down the U.K.'s National Lottery network just two days after Christmas, 2008, and prevented ticket sales for that day's lottery drawing. In addition, the network fault crashed the online games' web site. Sales in excess of £1 million pounds were lost, the margin on which would have been applied to good causes throughout the U.K. A single ticket won the Lotto jackpot, netting the lucky winner £3.4 million pounds. However, he would have been £800,000 pounds richer if the other tickets had sold, which would have resulted in a larger pool.

US Airways Suffers Fiber Cut

About 100 US Airways flights were delayed nationwide on January 29, 2009, when a fiber optic cable was cut near one of its data centers in Phoenix, Arizona. US Airways operates about 3,100 flights per day. The cable cut affected the airline's flight dispatch systems and some of its airport computer systems. Data-processing services were restored in about two hours.

Tokyo Commodity Exchange Taken Down by Router

The Tokyo Commodity Exchange (TOCOM), Japan's largest commodity market, had to suspend trading for over three hours on May 16, 2009, when connectivity between its member firms and the floor was lost. The problem occurred just days after the exchange upgraded to a new technology platform from Nasdaq OMX Group. The fault was traced to a router that was showing a 99% load during a time that a 5% load was expected. The router was replaced; and connectivity was restored, allowing the Exchange to resume trading a half-hour before its daily close. The night session beginning two hours later was uneventful.

NYSE Trading Halted Due to Routing Problems

Trading on the floor of the New York Stock Exchange for 242 stocks, including American Express, Merck, General Electric, and Exxon, was halted at 10:43 AM on June 12, 2009, when orders could not be routed to brokers on the floor. Eight of the 27 NYSE-traded stocks that make up the Dow Jones Industrial Average were affected. Though floor trading in these stocks was halted for several hours, electronic trading continued. The Dow was calculated from share prices that weren't being updated but that were corrected once floor trading resumed.

Router Failure Takes Down WestJet Airlines

On August 7, 2009, a router in WestJet's Calgary network center took down the airline's computer system that it uses to check in passengers. The router was redundant, and its backup should have taken over immediately. The problem was that the router didn't quite die. Though it failed to forward traffic, it was not sick enough for its backup to take over. About 1,000 passengers were affected for over an hour, and several flights out of Toronto and Montreal were seriously delayed.

Maintenance Disrupts California Airspace

A maintenance subcontractor's mistake shut down the FAA Telecommunications Infrastructure (FTI) in the Oakland Air Traffic Control center on August 28, 2009, affecting flights over Northern California, Nevada, and the Pacific Ocean. Controllers had to rely on cell phones to coordinate flights with neighboring FAA facilities. It seems that a maintenance error created problems with the communication system. The backup system was put into service but failed the next day, causing the outage. The controllers were never notified that they were running with a single point of failure. Normally, such a notification would have put them on alert status.

BBC's Web Site Off the Air for Half a Day

The BBC web site is the site that everyone in the U.K. turns to when they think the Internet is down because its web site never fails. But the U.K. 'go to' web site did fail on November 5, 2009. In a statement, BBC acknowledged that the failure was due to a network problem but elaborated no further. BBC apologized for the outage and indicated that BBC engineers were monitoring the network to ensure that no further problems occurred.

WordPress Blogging Site Down for Two Hours Due to Routing Error

WordPress hosts over ten million blogs, including TechCrunch. On February 18, 2010, WordPress suddenly went offline for almost two hours. It is estimated that over five million page accesses were lost, but no data was compromised. It turned out that a latent cabling error in one of its data-center providers caused an alternate route to be improperly configured. The erroneous route could handle only 10% of the normal WordPress traffic. The routing error also broke the failover mechanisms between WordPress' San Antonio, Texas, and Chicago data centers.

Failed Edge Router Isolates Colgate University for a Day

All of Colgate University's access to the Internet died for a day when its edge router crashed on April 22 , 2010. During this time, the Colgate network was inaccessible to all off-campus users. All Internet traffic flows through this one router, but Colgate has only one because of the router's six-figure cost. According to its service contract, Cisco delivered a replacement router within four hours; but this router also failed. By the time a good router was received and installed, the University had suffered almost a day of isolation.

Network Outage Isolates Dallas Data Center of The Planet

On May 3, 2010, one of four border routers failed at hosting provider The Planet and affected connectivity with the company's core network in its Houston data center. The outage cut off access between some hosted servers and the Internet for almost two hours. The failure also dropped connections to several Internet transit providers directly connected to the router. Shortly after the network was restored, The Planet suffered a link failure between its Dallas and Houston data centers. This network outage isolated some customers from their servers.

Cisco Software Bug Takes Down a Piece of the Cloud

Cloud-hosting infrastructure provider Hosting.com lost connectivity to its Newark, New Jersey, data center for almost two hours during a busy afternoon on June 2, 2010. The company reported that a software bug in a Cisco Catalyst 6509 switch not only caused the problem but also disabled both the primary and the backup switches. Many major cloud providers were affected, including Hostway, Rackspace, and Amazon Web Services.

Bell Canada Takes Down Air Canada

Air Canada, Canada's major airline, saw its computer operations come to a halt for several hours on the morning of October 20, 2008. The culprit was a nationwide glitch caused by a routing problem in Bell Canada's trans-Canadian backbone communications network. The outage affected all of Air Canada's online IT operations, from self-service kiosks to sign-in desks, gate operations, ticketing, and reservations. There were massive delays in boarding passengers. Bell Canada finally corrected the problem around noon by routing around the fault. What? No automatic rerouting? Backbone networks do fail!

Summary

Complex Intranets are the glues that hold together data centers. When an internal network fails, data centers can lose much if not all of their usefulness. Even worse, a data center may depend upon other data centers to properly function. If those data centers suffer an internal network failure, the data centers to which they are supplying services may also come to grief.

Most internal corporate networks today are based on Internet technology. Typically, however, a company does not invest in the redundant paths and the dedicated administration of its internal Intranet that a public Internet service provider does. When a network path fails, there is often no fast way to recover from the fault. Identifying the source of the fault and correcting it can take hours, during which the data center may effectively be down.

So far in this series, we have focused on technical failures. But a disturbingly large number of failures are caused by human actions, whether accidental or malicious. In our next article, we will examine upgrades that have gone wrong – all due to poor planning by people.