the Availability Digest

www.availabilitydigest.com

More Never Agains V

July 2010

It seems that no matter how hard we try, ensured availability of our IT services continues to be evasive. In this article, we review a sampling of the many outages that have struck major companies in the past six months. Networking continues to be a top issue, causing over a third of all outages. Likewise, during this period, power and cooling failures accounted for another third. Interestingly, half of the power outages were caused by automatic transfer switches that did not cut the data center over to backup power. The rest of the outages were a mix of hardware faults, software bugs, and a denial-of-service attack.

A disturbing statistic is that half of these outages were experienced by large hosting, cloud, and SaaS providers such as Amazon, Hostway, Salesforce, Rackspace, and The Planet. This observation emphasizes the fact that shared services are not yet suitable for critical applications unless you have a good, tested failover plan.

Amazon's EC2 Services Down for Hours

Data Center Knowledge, December 10, 2009 – Amazon Web Services suffered an outage in its Virginia data center when a power distribution unit (PDU) failed at 4 AM. The problem started when a single component in the redundant PDU failed. Before that component could be repaired, its redundant partner failed; and the data center lost power for 45 minutes. It took hours to get customer instances back online, though most were up and running within five hours. The failure affected only one of Amazon's Availability Zones on the U.S. East Coast.

Australian Card Holders Get Hit with the Y2010 Bug

Topnews.net – January 3, 2010 – Shoppers at over 8,000 Bank of Queensland EFTPOS (electronic funds transfer/point of sale) terminals found that their credit and debit cards had "expired." It seems that at the stroke of midnight on New Year's Eve, the terminals rolled their dates over from 2009 to 2016; and any card with an expiration date prior to 2016 (which included all cards) was no longer valid. The bank quickly came up with a code that merchants could enter to make the terminals ignore the date.

Salesforce Taken Down By Dual SAN Failure

Tech Target, January 5, 2010 - Software-as-a-service provider Salesforce.com suffered a wide-spread outage on the first working day of the year. The outage took down most of Salesforce's 68,000 customers for more than an hour. Though unconfirmed by the company, it appears that the problem was a dual failure in a major redundant SAN that took out both the primary and backup systems. Salesforce.com system operators had to reboot systems to restore connectivity. The company runs its entire operation out of a data center in Silicon Valley and replicates its data

© 2010 Sombers Associates, Inc., and W. H. Highleyman

to another data center on the U.S. East Coast. It has announced plans to open a third data center in Singapore.

Y2K Bug Hits German Shoppers a Decade Late

Financial Times, January 6, 2010 – A Y2K-like bug triggered by the change of the decade left thirty-million German debit- and credit-card holders unable to make purchases or make ATM withdrawals. The bug was in the microchips embedded in the “chip and pin” cards that did not recognize the year 2010 as a valid year. The French card maker Gemalto admitted it issued the defective cards. Rather than replace millions of cards, banks are reconfiguring their ATMs and POS terminals to accept the faulty cards.

Don’t Put All Your Eggs in the Same Cloud

Search Cloud Computing, January 8, 2010 – Heroku, a web hosting service, uses Amazon’s EC2 cloud services to host its own cloud services. On January 2nd, all 22 instances of its hosted services running 44,000 applications suddenly vanished; and Heroku was down for an hour. Amazon blamed the fault on a router failure in its Virginia data center. All 22 Heroku instances were in a single Amazon Availability Zone. Though failover was built in, it was to the same Availability Zone. Configuring backups in other Availability Zones may well have prevented the problem.

VoIP Provider 8x8 Taken Down by ISP

TMCnet, January 15, 2010 – 8x8’s VoIP internet telephone service was taken down for four hours by an “unaffiliated ISP.” According to an 8x8 spokesperson, a Tier-1 ISP began misrouting 8x8’s IP messages, which use non-contiguous address blocks. 8x8 responded by broadcasting correct routing information from its backup data providers. During the outage, not only was Twitter alive with 8x8 customer postings of complaints and suggestions, but also with messages from 8x8 competitors advertising their services.

Rackspace Taken Down by a UPS Failure

Data Center Knowledge, January 18, 2010 – A UPS failure in a Rackspace London data center caused a power outage in that facility. It took hours to recover all of the servers. The outage occurred when a module failed in the UPS unit, and the unit failed to transfer the load properly. System personnel had to manually intervene to bring up 220 servers. In many cases, the staff had to replace power supplies, replace firewalls, reconfigure switches, and log on to the failed servers. Just a month earlier, Rackspace suffered a major outage by a routing error introduced during a test of the network linking its Dallas data center to its new Chicago data center.

Off-Track Betting in Australia and New Zealand Taken Down by Power Failure

Voxy, January 18, 2010 - The Totalizator Agency Board (TAB) of New Zealand provides off-track betting for member race tracks throughout New Zealand. About 2:25 in the afternoon of Sunday, January 17th, a power-supply failure in TAB’s Petone data center took down both of its servers, forcing TAB to cancel racing. The failure allowed some bets to be placed on the races in progress after the starting bell, but no bets could be placed on later races.TAB contacted customers who bet after the race started in order to pay for those bets on winners (horses and greyhounds) and to refund bets on losers. A new system is scheduled to be installed at the end of the year.

Twitter Goes Down Due to a Failover Fault

Cnn.com, January 20, 2010 – Twitter went down for about 90 minutes when it suffered a fault and was unable to fail over to its backup system. During the failure, Twitter users saw nothing but the

© 2010 Sombers Associates, Inc., and W. H. Highleyman

“fail whale,” the iconic symbol of a Twitter failure. Though there was no word on the cause of the failure, it occurred right after the Haitian earthquake, leading many to believe that Twitter’s system suddenly became overloaded with tweets. Twitter said that though it was down for an hour and a half, no tweets were lost.

South Africa Isolated for a Day by a Cable Fault

The Daily Maverick, January 21, 2010 – The SAT-3 undersea cable that carries most of the traffic between South Africa and Europe broke down for about 24 hours, effectively isolating South Africa from Europe and the rest of the world. The incident started when Telkom, the cable operator, began maintenance on the cable after informing customers that they might experience increased latency on the channel for four to six hours. However, an error by maintenance personnel working on the power units caused a massive failure of the cable.

Iowa Internet Routing Error Affects 22 States

Columbia Missourian, January 22, 2010 – Customers of Mediacom, a major ISP serving 22 states in the middle U.S., started having problems with Internet connectivity Tuesday evening. At first, only customers in Columbia, Iowa, were impacted. But by the next evening and through the following morning, the problem had spread to customers in 22 states. The problem was finally traced to a routing error at Mediacom’s Internet Network Operating Center in Iowa. Mediacom has installed additional monitoring facilities to address similar problems more efficiently in the future.

Power Surge Takes Out California Data Center

Data Center Knowledge, January 28, 2010 – During severe storms on January 19th, a local power outage caused a power surge that was not handled properly by the surge suppressors in the San Jose, California, data center of NaviSite, a managed hosting and cloud service provider. The surge blew out relay fuses and prevented the automatic transfer switch from starting the data center’s diesel generators. The battery UPS did not last long enough, and the entire data center was offline for almost an hour until the diesel generators could be started manually.

Minnesota’s North Shore Cut Off From World by a Steam Pipe

Minnesota Public Radio, February 4, 2010 – During the midmorning of Tuesday, January 26th, all counties in Minnesota’s North Shore along Lake Superior were cut off from the rest of the world for about twelve hours by a fiber cable break. The North Shore is connected to Duluth, MN, via a single cable – no redundancy. Conjecture is that the cable was laid alongside a steam pipe, and the heat destroyed the cable. Affected were 911 services (which are routed to Duluth), senior FirstCall emergency alert buttons, customs agents at the Canadian border, ATM and credit/debit card transactions, banks, and online businesses.

RAID Failure Takes Down Hosting Service for Five Days

The Register, February 10, 2010 – HostV is a hosting service with U.S. data centers in New Jersey and Chicago. HostV provides dedicated and virtual servers to its customers. In early February, it suffered a massive RAID failure that severely corrupted its database, taking down many of its servers. After trying to recover the data from the RAID disks, HostV finally realized that much of the data had to be restored from offline storage. However, the restoration of the encrypted data took much longer than anticipated. Five days later, at least one server and its data were still inaccessible.

© 2010 Sombers Associates, Inc., and W. H. Highleyman

Microsoft Windows Live Taken Down by a Single Server Failure

The Tech Herald, February 17, 2010 – The Microsoft Live sign-in service, which provides access to Microsoft services such as Hotmail and Messenger, refused to let customers in for about an hour when a server failed. The load normally handled by that server was rerouted to other servers, but they became overloaded; and the entire sign-in process became bogged down. As a result, many login attempts failed. System personnel moved in a new server and restored normal capacity in about an hour, but it took several hours for the backlog to be cleared and for service to return to normal.

WordPress Blogging Site Down for Two Hours Due to Routing Error

PC Magazine, February 19, 2010 - WordPress hosts over ten million blogs, including TechCrunch. On February 18th, 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.

Spotify's Music Silenced by Faulty Air Conditioner

The Next Web, February 21, 2010 – Spotify's music-streaming service stopped soothing the savage soul for several hours. It seems that there was a power failure at a large London data center where Spotify hosts some of its servers. The backup power kicked in as planned, but one of the large air conditioners in the data center did not start properly. Without full cooling, the data center's temperature began to rise; and servers started shutting down to protect themselves.

Chile Plunged into Darkness by a Transformer Failure

Seattle Times, March 14, 2010 – A power failure plunged almost all of Chile into darkness on a Sunday evening, leaving nearly the entire Chilean population of 17 million people without power. A transformer failure caused a ripple effect that ultimately caused a total collapse of the Chilean electrical power grid. It was hours before power was restored to most of the population. The outage occurred during aftershocks from the devastating 8.8 earthquake that rocked Chile on February 27th, but the outage did not seem related to earthquake activity.

Codero Downed by Faulty Automatic Transfer Switch

Data Center Knowledge, March 16, 2010 – Dedicated hosting provider Codero suffered a major power outage in its Phoenix, Arizona, data center that disrupted operations for several hours for most of its customers and caused even lengthier downtime for 10% of its customers. The backup generators started properly, but a faulty automatic transfer switch failed to switch the data center to the generator power. When the UPS batteries died, the servers died. The power outage also damaged a core router, further delaying recovery. Several servers suffered damaged hardware, causing extensive delays for customers hosted on those servers.

Wikipedia Succumbs to Heat, Failover Fault

PC Magazine, March 25, 2010 – Wikipedia had to shut down its European data center to protect its servers from overheating. No problem. Wikipedia has a rapid backup procedure and executed it to move traffic from the European center to servers in Florida. However, the failover procedure was broken, causing the DNS resolution of Wikipedia sites to stop working globally. It took about two hours to restore Wikipedia to service.

© 2010 Sombers Associates, Inc., and W. H. Highleyman

Sprint's New Customer Service Website a Fiasco

Sprint Connection, April 14, 2010 – Sprint Nextel stopped all customer support when it upgraded its web site, sprint.com. The new site was supposed to improve how customers managed their accounts and received help. Instead, as soon as Sprint flipped the switch, customers could no longer access the site. Customers could not activate new phones, change preferences, view their accounts, or pay their telephone bills. Problems continued for five days. What? No fallback plan?

Middle East Internet Access Still Fragile

TeleGeography, April 20, 2010 – Only three submarine cables link the Middle East with Europe, and 89% of all traffic is carried by just one of these cables. Last week, the main cable, SeaMeWe-4, was damaged; and traffic slowed to a crawl. It took about two weeks to restore service to normal. This outage followed major breaks in January and December of 2008. However, relief is on the way. Five new cables are scheduled to enter service between Europe and the Middle East later this year.

Failed Edge Router Isolates Colgate for a Day

Maroon News, April 22, 2010 – All of Colgate University's access to the Internet died for a day when its edge router crashed. 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 its six-figure cost. According to its service contract, Cisco delivered a replacement router within four hours; but it failed too. 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

Data Center Knowledge, 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.

Terremark Survives a Fire with No Downtime

Data Center Knowledge, May 4, 2010 – On April 30th, a fire broke out due to a transformer malfunction in one of the data-center electrical rooms at Terremark's Virginia data center. Terremark, a managed hosting provider in Virginia, had the good foresight to involve the local fire department in the design of its data center. By the time the fire trucks arrived, the emergency diesel generators had kicked in. However, the fire department, being familiar with the electrical distribution in the data center, was able to isolate power to the fire-affected room, thus avoiding having to push the dreaded Emergency Power Off switch.

Das Internet ist Kaput!

The New Internet, May 13, 2010 – Over thirteen million German web sites use the country's top-level domain, .de. Millions of these web sites became inaccessible for almost two hours when DENIC, the German Internet authority, uploaded new zone files that were empty. In effect, this meant that all web sites in those zones no longer existed. The web sites could not be reached, and email was rejected. Some reports indicated that all web sites beginning with "a" through "o" were down.

© 2010 Sombers Associates, Inc., and W. H. Highleyman

Car Crash Takes Down a Portion of Amazon's EC2 Cloud services

Data Center Knowledge, May 13, 2010 – Amazon's EC2 cloud-computing service suffered it fourth power outage in a week when a car hit a utility pole, and a transfer switch failed to manage the shift from utility power to the Virginia data center's diesel generators. The transfer switch was delivered by the manufacturer with an improper default setting and determined that the fault was inside the data center rather than outside. It therefore did not switch power to prevent danger to the building's occupants. A portion of EC2 users were down for a little more than an hour.

Bluehost Taken Down by Power-Induced Network Failure

Data Center Knowledge, May 25, 2010 – A maintenance error at a substation in Provo, Utah, forced a 138KV circuit breaker to open, resulting in an area-wide power blackout. The power outage caused the data center of web-hosting provider Bluehost to properly switch over to backup power. However, the power failure also took down telephone service for most of Provo, including the Internet connectivity for Bluehost. Bluehost was inaccessible to its customers for several hours.

Media Temple Hit by Denial-of-Service Attack

Data Center Knowledge, May 25, 2010 – Web-hosting provider Media Temple was hit by a denial-of-service attack on its DNS server, resulting in downtime for any hosted customers that used Media Temple's name servers. Due to the sophistication of the attack, Media Temple's firewall did not block the attack adequately, as the traffic appeared to be legitimate. Media Temple subsequently blocked all traffic from Asia, South America, and Mexico to reduce the impact of the attack. The blocks were later lifted.

ATT's VoIP Service Down for Hours

Associated Press, May 25, 2010 – ATT's U-Verse digital home service that provides telephone communications over the Internet went down for over four hours, silencing telephones in its 22-state U.S. local phone service area. The outage lasted from 10:30 AM until 2:25 PM. Calling a U-Verse number resulted in a message saying that the service had been disconnected. Support personnel told customers that a server failure had taken down ATT's U-Verse service in all 22 states.

Cisco Software Bug Takes Down a Piece of the Cloud

Data Center Dynamics, June 2, 2010 – Cloud-hosting infrastructure provider Hosting.com lost connectivity to its Newark, New Jersey, data center for almost two hours during a busy afternoon. 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.

© 2010 Sombers Associates, Inc., and W. H. Highleyman