Vodacom's One-Year Recovery

Brett Dismore >> **Principal Engineer** >> Business Connexion

Paul J. Holenstein >> **Executive Vice President** >> Gravic, Inc.

Vodacom (Pty) Ltd and Prepaid Front End (PPFE)

Vodacom (Pty) Ltd is one of the largest cellular telephone service providers in Africa. From its roots in South Africa, Vodacom has expanded to include cellular telephone networks in Tanzania, Lesotho, and the Democratic Republic of the Congo. It currently offers voice and messaging services to over 55 million customers. Vodacom is owned by Vodafone, the world's second largest mobile phone company (behind China Mobile).

Prepaid calling cards, the fastest growing cellular option in Africa, are a major Vodacom service. If this service is unavailable, much of Africa's cellular capability comes to a halt as subscribers exhaust their cellular minutes. Therefore, Vodacom uses HP NonStop server pairs to provide prepaid calling card services via its Prepaid Front End (PPFE). The PPFE, primarily a NonStop OSS application, is used by customers to recharge their accounts. If the PPFE fails, subscribers cannot add money to their cellular accounts; if subscribers exhaust their accounts, they no longer have cell phone service. To minimize PPFE outages, the HP NonStop server pairs implement the Shadowbase bi-directional Sizzling-Hot-Takeover (SZT) data replication architecture to ensure multi-second recovery times.

As within its other geographical areas of service, Vodacom provides a PPFE HP NonStop SZT pair in Tanzania. However, in August, 2013, a battery explosion downed Tanzania's production PPFE NonStop server. Vodacom was able to switch PPFE operations to its backup system in just a few minutes; however, problems with the Online Charging System/Intelligent Networking (OCS/IN) platform restricted full operational capability of the remote (backup system) until the next day. Even worse, a year later, the backup capabilities for this installation were still being brought back into service. This article investigates the causes of the lengthy failure/recovery cycle as a sober lesson for other mission-critical businesses that are implementing business continuity architectures.

The Tanzanian PPFE Configuration

Call Handling

A simplified overview of how cell calls are handled is shown in Figure 1. In this figure, cell phone subscribers connect to the Vodacom cellular network via Intelligent Network (IN) systems. Each subscriber is assigned an IN based on the first digits of his telephone number. Each cell phone tower connects to an IN. Various information is collected about a cell phone user (such as the number of minutes he has left in his account, and the geographical location of the subscriber).

When a subscriber turns on his phone, his mobile signal is picked up by the closest cell phone tower. The IN to which the cell phone tower is connected determines the subscriber's assigned IN from his mobile- phone number and notifies the assigned IN of the subscriber's location. The subscriber is connected to an IN, which also obtains the subscriber's account data from his assigned IN.

When a subscriber places a call, it is handled by the connected IN, which contacts the assigned IN of the called party to determine the location of that party and then routes the call to the connected IN of the called party. The call has now been established.

When the call has been completed, both the IN handling the calling party and the IN handling the called party update their respective subscribers' account data (for instance, deducting minutes from the subscribers' accounts of prepaid minutes). The updated data is then sent to the subscribers' assigned INs to update their databases.[3]

Adding Minutes

In order to add airtime to a prepaid cellphone account, a subscriber has a number of choices:

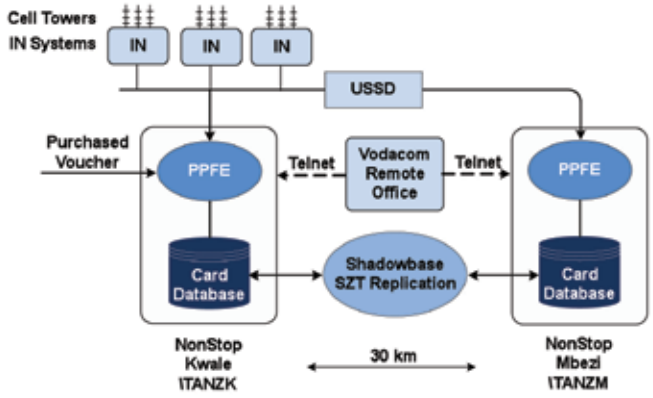

Figure 1 – PPFE Architecture

- *Purchase a voucher at a store*
  To activate this airtime, the subscriber has to redeem the voucher by sending the unique number printed on the voucher to the Vodacom PPFE system using Unstructured Supplementary Service Data (USSD).[4] The PPFE system then performs a lookup in its database to get the amount of airtime associated with this unique number from the voucher and applies it to the subscriber's IN.
- *Purchase airtime from a Bank ATM*
  This transaction is sent to the Vodacom PPFE system via

[3] The actual details of how calls are handled are much more complex than described here, which is just a simple overview for the purposes of this case study.
[4] USSD is a protocol used by cell phones to communicate with the service provider's computers.

a standard financial format message (called an ISO8583 message). The PPFE system verifies the request by using data in its database and applies the correct airtime to the subscriber's IN account.

- *Purchase airtime from a street vendor*
  Street vendors have Vodacom accounts (similar to bank accounts) that are stored on the Vodacom PPFE system. The vendor uses USSD to interact with the PPFE system and the PPFE applies the airtime to the subscribers IN account when the vendor completes the transaction.

Continuous Availability for the PPFE

If the PPFE experiences an outage, subscribers will not be able to add minutes to their accounts. Cell phone service will become unavailable to subscribers once their balance is exhausted. Therefore, the PPFE is implemented as a dual HP NonStop server pair using a bi-directional Shadowbase SZT data replication configuration, where one system is the production system and handles all transactions (see Figure 1). However, the backup system is actively running the PPFE application and is ready to take over in seconds if the production system fails. Its database is kept synchronized with the production system by the Shadowbase data replication engine from Gravic, Inc. (www.gravic.com/shadowbase).[5]

Shadowbase data replication is accomplished in sub-second times so that the backup database is synchronized with the production database. If the production system fails, then rerouting transactions to the backup system is all that needs to be done. By configuring bi-directional replication, any changes made on the promoted system queue are delivered to the original system to resynchronize its database once it is recovered. The backup system can be assessed continually with test transactions against test accounts to ensure that it is functional and working end-to-end. This assessment is quite useful since validating the backup system's application processing can be done at any time without requiring a production application outage. This testing helps to ensure the backup system will be able to take over processing without any problems if the production system fails.

On behalf of the customer, Vodacom installed one PPFE NonStop system in Oyster Bay, Kwale Street, Tanzania, and the other in Mbezi, Tanzania (Figure 2). The two systems are about 30 kilometers apart and are situated on high ground to address the primary cause of disasters in Tanzania – tsunamis. The systems are managed for the customer by Vodacom personnel located in a remote office connected to the systems via Telnet links.

The node name for the NonStop system in Kwale is \TANZK, and the node name for the NonStop system in Mbezi is \TANZM. Node \TANZK was the production node and the Shadowbase architecture replicated its database changes to its SZT backup node, \TANZM.

The IN systems at the backup site are typically down, and require a restart sequence to bring them up in the event of a failover. This means that during normal processing at Kwale, all IN systems are connected to the production node in Kwale. IN access to the backup site in Mbezi is provided by an Unstructured Supplementary Service Data (USSD) gateway.

The Production PPFE System is Downed by an Explosion

Disaster Strikes

On Friday afternoon, August 16, 2013, disaster struck. A UPS battery exploded and caught fire in the Kwale datacenter. The explosion, along with the fire-suppression system, damaged the \TANZK PPFE production node, taking it out of service. The sudden outage of \TANZK caused significant database corruption, as updates in progress could not be completed.

Even worse, the IN systems were in the same data center as the \TANZK PPFE node, and their local database was corrupted by a hard down when the failure occurred. Their communication links with the backup PPFE server in Mbezi were also damaged.

Failover to the Mbezi IN systems failed, with the Mbezi IN systems being unable to complete the DB startup sequence and allow access to the IN customer database application. Later, it was learned that this was caused by a massive number of simultaneous

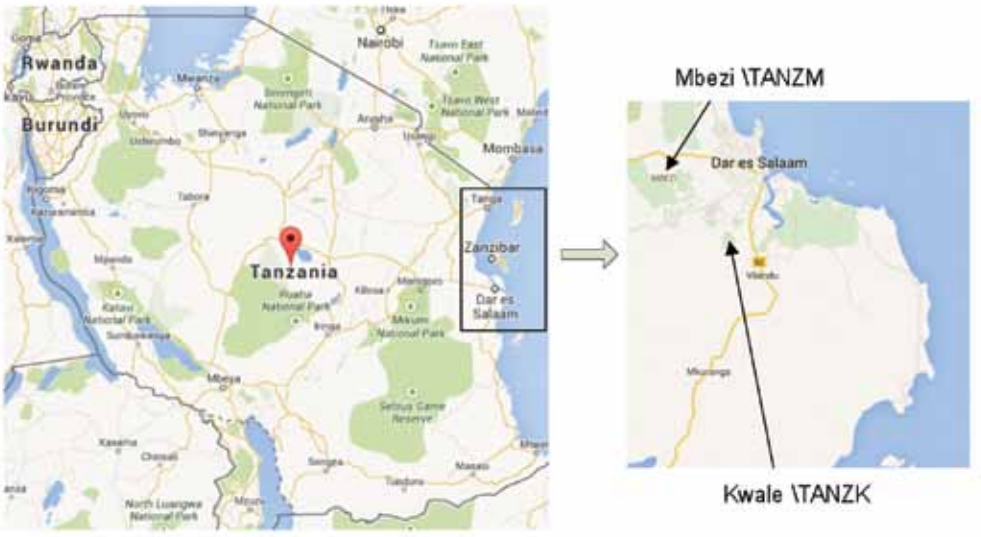

Figure 2 – Location of HP NonStop PPFE Servers

[5] For more information about Shadowbase and the Sizzling-Hot-Takeover business continuity architecture, please read the Gravic white paper: *Choosing a Business Continuity Solution to Match Your Business Availability Requirements*.

farm node requests to connect to the Mbezi IN database coming in and overloading the environment, causing timeouts and retries.

If the IN system failover to the backup systems installed in the Mbezi datacenter had been successful, subscriber calls to redeem prepaid vouchers could have continued with little, if any, interruption. Instead, cell phone requests to redeem prepaid vouchers for additional minutes could not be serviced by the backup PPFE system in Mbezi. Subscribers running out of minutes lost their cell phone service.

Lesson Learned: Avoid single points of failure. All network, system, application, and data components necessary to provide service must be redundant, must exist in geographically seperate locations, and need to be tested for failover periodically.

The Initial Recovery

By the next morning, the \TANZK node was once again operational from a hardware viewpoint. However, none of its software had yet been restored. There were no startup scripts, no subsystems, and no applications running. An attempt was made to start TMF, but it immediately ran into problems. Since the database had been left in an inconsistent state from the sudden node outage, TMF could not recover the database from the Audit Trail. The database had to be recovered from its backup, provided by ETI-NET's virtual tape facility, BACKBOX.[6]

By noontime Saturday, communication between the IN systems in the Kwale datacenter was established with the PPFE backup in the Mbezi datacenter via the USSD gateway system installed in Kwale. The USSD gateway ran on a Windows server that had not been damaged by the explosion. By this time, the Tanzanian prepaid voucher redemption service had been unavailable for a day.

The Long Database Recovery

The database backup strategy is that a full backup (via a TMF online dump) is done over the weekend with incremental backups (via dumping the generated audit trail files) during the week. Therefore, the backup system's corrupted database had to be restored, followed by rolling forward thru nearly a full week's worth of Audit Trail change data. This recovery was accomplished via a TMF Recover Files operation for the corrupt database partitions, and it had to be completed before the database could be made available for application access. The database comprised SQL/MP tables with SQL/MX aliases. The database recovery from the online dump was completed by Saturday afternoon. However, it took an additional two weeks to complete the Audit Trail roll forward recovery.

During this time, even though the PPFE service was restored to subscribers, the production node now in Mbezi (the original backup node) had no backup, and a failure of that system would have again rendered the PPFE service unavailable. When database recovery was finally completed, the Tanzanian PPFE system was restored to an active/passive SZT configuration with production accomplished by the \TANZM node and backup services provided by the recovered \TANZK node.

[6] See http://www.etinet.com/ for more information about the BACKBOX product line.

Lesson Learned: 1) Back up more than once per week – perhaps every few hours if you are using virtual tape. Otherwise, to recover the database, you may need to restore considerable amounts of data as you roll the database forward from the last backup point, which can take a long time, extending the service outage. 2) Parallelize the backup and restore operations as much as possible (and test the recovery). 3) Consider whether a single backup system/site is sufficient, since an extended outage of the primary system/site or the backup system/site will leave you vulnerable to another prolonged outage if either system/site also fails.

The Erratic Backup System

However, even though the PPFE was now running in an active/passive SZT mode, the backup node in Kwale proved to be somewhat unstable. Vodacom made many service calls to HP. HP was very helpful, but the service calls were for a backup system and not for a production system and hence were not of the highest priority Service Level Agreement (SLA). The service contract allowed for longer service intervals for non-production system issues. However, no matter what the time of day or night, HP was very helpful and consistently went above and beyond its contractual obligations under its support agreement.

Lesson Learned: Carefully consider the support SLA's required for backup systems for mission critical applications.

Vodacom finally asked HP to visit onsite and verify that the Kwale node hardware was in good condition. HP performed exhaustive tests on the hardware and determined that the server was unusable due to smoke, water, and other fire-related damage.

Replacing the Backup System

Vodacom turned to its insurance company to obtain funding for purchasing a new backup node. The insurance company at first balked because the damaged node was still functioning, albeit erratically. However, after an extended negotiation, Vodacom convinced the insurance company to finance a new node.

Lesson Learned: If you do not have a total loss, be prepared to fight with the insurance company to obtain coverage.

The new node funded by insurance was installed in a different building than the original \TANZK node. It was given the name \TANZKW. At the same time, Vodacom replaced its IN network with the IN Advantage system from Siemens. The multiple INs in the original configuration were replaced with just two of the more powerful Siemens' INs, freeing considerable space in the new datacenter.

Bringing Up the New Backup System

As shown in Figure 3, bringing the \TANZKW node into service was the next task. The first problem was the shipping

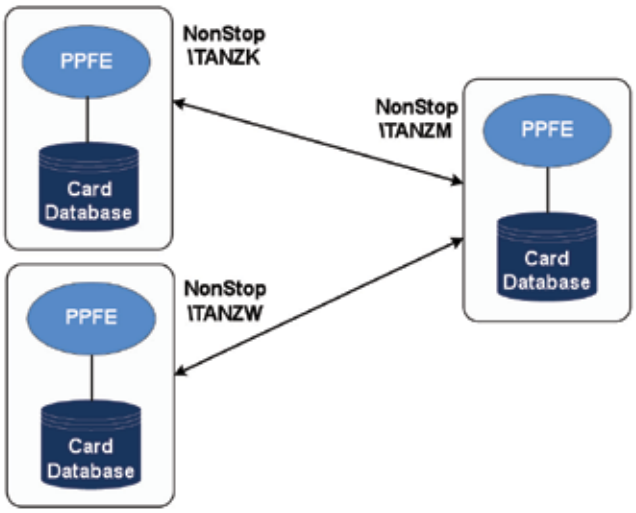

Figure 3 – Recovering \TANZKW as the New-Hot-Standby

of the replacement NonStop to the Kwale data center. When the new hardware was shipped, the cabinets had to be stripped and reassembled on site because the cabinets could not be accommodated standing up on the airplane. Every single device was unracked/de-installed from the cabinets in the factory, and the cabinets were shipped in the horizontal position. Devices – servers, switches, disk shelves – were all separately packed and shipped on pallets. All of this shipping resulted in a much longer time to get the new node hardware rebuilt and up and running.

Once the \TANZKW hardware was operational, the node's subsystems and communications were installed. Vodacom personnel were onsite for six days bringing up the \TANZKW node. At this point, further restoration procedures for the node could be accomplished from the Vodacom remote office via Telnet.

Next, the support team began the migration of the PPFE application to the new node. They started by loading the \TANZKW databases from the \TANZM production system. Several challenges slowed this procedure to a crawl. The first problem was that the software versions on the older \TANZM node and the newer \TANZKW node were different. The original \TANZM and \TANZK nodes used the J06.12 version of the NonStop operating system, and the \TANZKW node used J06.16. Furthermore, the original nodes used SQL/MX version 2.3.4, and the new node used SQL/MX version 3.2. The result was that the standard NonStop export/import utilities could not copy the database directly from the source environment to the target environment.

Lesson Learned: Understand the limitations of different hardware and software versions on systems that must interoperate, and be prepared to address the complexities.

The support team switched to NonStop's UNPAK2 and PAK2 utilities to perform block moves via FTP for the SQL/MX tables. Macros were created to generate SQL LOAD jobs to run in parallel for the large SQL/MP tables. Because the production system, \TANZM, had to be quiesced to PAK the SQL/MX tables, the PPFE application was quiesced at midnight when it was handling a minimum load and these tables were copied during overnight periods. Each morning, before the application was restarted, Shadowbase replication was initiated on the newly copied tables to keep them synchronized as the application updated them.

The Shadowbase SOLV Online Loader was employed to copy the smaller SQL/MP tables since it could replicate these tables while the application was running. It also kept these tables synchronized with the production system as the tables were being created and loaded. Thus, small-table migration proceeded continuously throughout the day.

Once the SQL/MP tables were loaded, their indices had to be built, which took an extraordinary amount of time. Because of the size of the tables and the amount of system resources available (disk and cpu), the indexing had to be undertaken one table at a time. Twelve terabytes of data had to be indexed, which took a total of 55 days. The support team worked long hours through March and into April to accomplish this task. Progress slowed considerably in April, which is vacation month in Tanzania. Progress was further gated by the lack of overtime authorization to complete the restoration task. All work had to be accomplished during normal office hours while the employees were still tasked with their regular support and systems administration jobs.

Lesson Learned: Allocate additional bedget for recovery operations after a failure to allow for the additional effort. Such factors should also be considered when developing or updating a business continuity plan, to ensure sufficient resource availability when it's needed, rather than after the fact when it only further slows the recovery process.

After the database was fully loaded onto \TANZKW with all of the indices created and Shadowbase replication keeping the database synchronized with the production database on \TANZM, the support team brought up the applications. To their horror, the disks immediately disconnected. It turned out that the logical unit locations for the various volumes were different from the ones that had previously been used on the older system's XP storage area network. It took another two weeks to fix these issues.

Once the database and disk storage was correctly configured, application environment issues arose with the OSS environment that affected the PPFE application, and further challenges with SQL/MP and the SQL/MX aliases came to light. These issues were finally resolved, and testing of the new \TANZKW node began in June, 2014.

Lesson Learned: It ain't over 'til it's over! In other words, even though you think the end is in sight, it often isn't...

Finally, on July 10, 2014, nearly a year after the original explosion, \TANZKW was brought into service as the backup system for \TANZM, and \TANZK was shut down. During all this time, the damaged system \TANZK had continued in an erratic fashion as the backup PPFE system. It even survived another fire that erupted in its data center in December, 2013. The fire caused further damage to the \TANZK system, but the support team was able to return it to service.

As of this writing, the systems have now been switched so that \TANZKW is the production system, and \TANZM is the backup system. The XP storage array used by \TANZM is currently being rebalanced to eliminate OSS hot spots that impose heavy loads on some disks in the array, which is necessary so that \TANZM will have the capacity to handle peak PPFE loads if a failover occurs.

Summary

Vodacom's PPFE is a critical system for its mobile services in Africa. The PPFE allows subscribers to top off their cell phones with purchased vouchers. Without the PPFE, cellular service will gradually grind to a halt as subscribers use up their minutes and cannot add additional minutes. To ensure continuous availability, the PPFE systems are duplexed across HP NonStop servers using a Shadowbase bi-directional SZT configuration that provides data access failover times measured in seconds.

An explosion in August, 2013, nearly destroyed one of Vodacom's Tanzanian PPFE systems. It took Vodacom nearly a year to rebuild the damaged system using new hardware. Challenges included hardware problems due to smoke damage, software version issues, lengthy database loads, and limited staff hours. During this time, the PPFE system limped along as an active/passive SZT system with the damaged node providing erratic backup to the production node. Fortunately during this period, the fault-tolerant HP NonStop production node experienced no problems that required failover to the erratic backup system.

Though prepaid voucher services to Vodacom subscribers should have been restored in seconds by a rapid SZT failover following the explosion, PPFE services were lost for a day because the IN communication channels linking the production site and the backup site were damaged. The IN environment did not have a backup system; probably the single most important lesson of this entire saga is to eliminate single points of failure.

Lesson Learned: If you want high availability, make sure that you can always failover to a backup environment quickly. Active/Passive environments can provide this level of avalability. If you want continuous availability, the "backup" environment must be up and running and in a known working state at all times. Sizzling-Hot Takeover and fully Active/Active environments can attain this level of availability. Pick your architecture to meet your business objectives.

A series of unfortunate events caused a mission-critical production environment to fail, with loss of services for over a day. The recovery of the failed environment took considerably longer than even the most pessimistic business continuity planner could have conceived. Failures happen. It is not a matter of if; it is only a matter of when. The Vodacom experience provides valuable insight into how to plan to avoid having a failure turn into a catastrophe. Consider whether any of the Lessons Learned might apply to your IT systems, and address them before disaster strikes.

The Shadowbase Data Replication Product Suite

The Shadowbase solution suite comprises several products addressing business continuity, data replication, data and application integration, zero downtime migration, and other utilities to deliver a true 24x7 "nonstop" enterprise. Shadowbase sales and support are now directly available globally from your HP NonStop account team, Business Connexion (Pty) Ltd in Africa, or contact Gravic, Inc. for more information for local resellers in your region.

Brett Dismore began his IT career in 1982 in the South African banking operations environment, moved through the RACF, TSO and JCL systems programmer ranks, and then worked as a DBA on the IBM mainframe. In 1995 he was offered the opportunity to implement and support the Tandem K Series platforms, which he did for four years, before moving to a NonStop third-party presales support career. Mr. Dismore has spent the last 13 years working as a principal engineer at Business Connexion (Pty) Ltd, providing support on the NonStop platforms for Vodacom SA, Vodacom Mozambique, Vodacom Tanzania and Vodacom DRC.

Paul J. Holenstein has direct responsibility for the Gravic, Inc. Shadowbase Products Group and is a Senior Fellow at Gravic Labs, the company's intellectual property group. He has previously held various positions in technology consulting companies, from software engineer through technical management to business development, beginning his career as a Tandem (HP NonStop) developer in 1980. His technical areas of expertise include high availability designs and architectures, data replication technologies, heterogeneous application and data integration, and communications and performance analysis. Mr. Holenstein holds many patents in the field of data replication and synchronization, writes extensively on high and continuous availability topics, and co-authored Breaking the Availability Barrier, a three-volume book series. He received his BSCE from Bucknell University, a MSCS from Villanova University, and is an HP Master Accredited Systems Engineer (MASE).