Global Information Assurance Certification Paper

Contingency Planning for ACE/Server 5.0
T.C. Chen
GSEC Practical Assignment version 1.4b
Option 2 – Case Study in Information Security

[Editor's Note: Shortly after this paper was submitted, RSA Security released ACE/Server 5.1, which provides enhanced disaster recovery capabilities. Additional details are available at http://www.rsasecurity.com/products/securid/whatsnew.html.]

Abstract

"Chance favors the prepared mind." - Louis Pasteur

When most people think about contingency planning, the tendency is to focus on disaster recovery related activities. However, without proper business continuity planning as well, a business entity may find itself quickly overwhelmed dealing with the aftermath of a disaster. With executive management support, a thorough Business Impact Analysis should be performed to ensure that all critical business systems/ processes and interdependencies are identified, and that each one is assigned a tolerable loss/disruption threshold. Ideally, the findings of the BIA should be used to proactively reduce the business' risk profile, not just to develop and implement processes that respond to the current state.

As a facilitator of strong authentication for enterprise applications and services, the ACE/Server application plays a key role in the areas of availability and confidentiality. Due care should be taken to ensure that its underlying infrastructure is secure and sufficiently distributed to maximize authentication service availability. Since administration functions can only be performed on the master/primary ACE/Server, its host will always be a potential single point of failure. Documented recovery procedures must be tested and maintained to ensure that administration capabilities can be restored within an acceptable timeframe, so that they will be available to support the recovery of dependent applications and services.

Before Snapshot: Opportunities for Improvement

Background

The events of September 11, 2001, had a profound effect on the business world. As "the biggest disaster of our time", it brought to the forefront the many challenges facing business continuity, and forced the management of businesses large and small to reassess their vulnerability to natural and man-made disasters (Hanning). At my company (hereafter referred to as ACME), executive management decided to fund a previously tabled business continuity option. Rather than continue to rely on a contracted East Coast disaster recovery services provider, ACME embarked on a yearlong project to migrate all production application hosts from its corporate headquarters in Silicon Valley, California, to a new Data Center facility in the more geographically stable Southwest. In addition to reducing operational risk, this decision

would allow ACME to leverage its existing California Data Center facility – left with supporting distributed/redundant production and development application hosts – as a hot failover/recovery site.

September 2001 holds additional significance for ACME in that it marked the beginning of a significant improvement in strong authentication service reliability and manageability. Although RSA Security released ACE/Server 5.0 for customer ship on June 4, 2001, it was not until Patch 5.0.1 was released in September that a concerted effort was undertaken to evaluate this major application revision in the ACME environment.

RSA Security's ACE/Server application is the brain behind the widely used SecurID authentication protocol. This protocol provides strong authentication through the combination of two-factors – a Personal Identification Number (What You Know) and a pseudo-random value generated on a time-synchronous token (What You Have) to provide more reliable user authentication than one-factor passwords. It is most commonly used to protect perimeter devices supporting network access – i.e., client VPN or dial-up RAS – and sensitive web applications from unauthorized access.

ACME ACE/Server Infrastructure – September 2001

One of the major issues with ACE/Server 4.x (and previous versions) is its architecture design of one master server and one slave server per realm. Because the slave server only responds to authentication requests when the master server is off-line (or specifically, when it cannot communicate with the master), only one server is available to facilitate authentication at any given time. This design, combined with a network security requirement prohibiting extranet hosts from directly accessing intranet resources, meant that a second ACE/Server realm had to be established in the ACME DMZ to function as an authentication proxy.

A "one-way trust" relationship[1] established between the two realms was sufficient to allow users in the Corporate/Intranet Realm to authenticate through agent hosts registered in the Extranet/Proxy Realm. This configuration permitted ACME to adhere to the established security requirement and avoid having to manage user and token records in two realms. However, two additional servers were required to provide redundancy for the one additional authentication point; and having two realms also meant that in the event of a disaster, both of the two master servers would have survive or be recovered in order to restore full administration capabilities. The minimum survival of the two slave servers is sufficient to maintain authentication service availability for existing users and agent hosts; however, the addition of new users or agent hosts and modification of existing user profiles or agent host configurations would not be possible.

[1] Establishing a cross-realm link between two realms does not in itself set up any trust relationships. The ability to trust or allow users from a remote realm to authenticate against a local agent host is configured on an individual agent host basis.

Figure 1 illustrates the ACME production ACE/Server 4.x infrastructure.

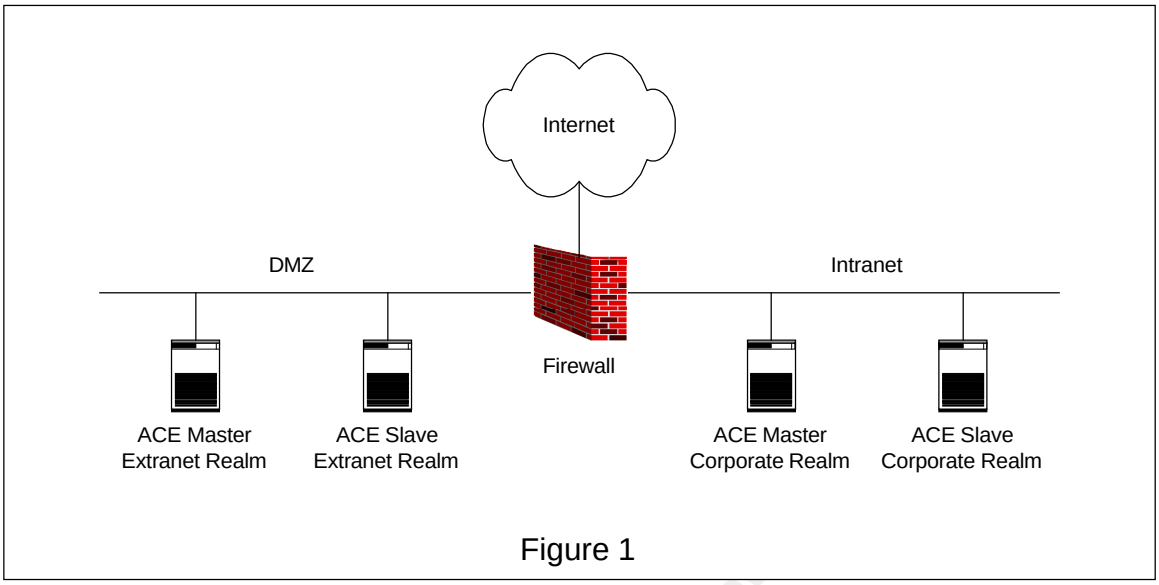

Figure 1

ACME Disaster Recovery Capabilities – September 2001

For several years, ACME relied on an East Coast disaster recovery services provider to have available the necessary facilities and hardware required to recover its mission critical applications. While the terms of its contract met the typical criteria for a Hot Site – a highly available facility, fully configured equipment ready for use within hours notice, exclusive use of the contracted portions of the facilities, testing available at least once a year (Harris 613) – the time required to assemble and fly ACME personnel and ship backup tapes from California to the East Coast would significantly delay the start of any recovery effort.

Previous Hot Site test results had proven that various application clusters could be recovered within a 48–64 hour window. However, the beginning of on-site data recovery was used for the start of recovery window instead of the time of Disaster Declaration[2]. Originally, for a large-scale regional disaster (like a high-magnitude earthquake), it was estimated that an additional 3-4 days would be required to handle the logistics involved. This meant that business operations could be impacted for up to 5-7 days. But then September 11 occurred, with its unprecedented impact on travel safety and availability. As a result, all travel dependent time estimates had to be reassessed, with corresponding adjustments to disaster recovery expectations.

While all "mission critical" applications were included on the Hot Site recovery list, a small number of arguably critical applications that either had reasonable workarounds or did not support easily recoverable architectures were omitted. ACE/Server was one of

[2] The Disaster Declaration is the formal executive management directive to execute the Disaster Recovery Plan that occurs after a period of time spent confirming and evaluating disaster details and implications.

these applications, with management's approval that dependent applications and services would be permitted to use native password instead of stronger authentication.

Had the ACE/Server been included, an estimated 4-6 hours recovery time, per master server, would be required to perform system installation and configuration, tape backup utility installation and configuration, data restoration, and application validation. Even then, the recovered ACE/Server database(s) would be missing at least 1-2 days worth of pre-disaster production data due to the off-site storage transfer cycles for archive backup tapes. Additional time would also be required to perform ACE/Agent validation or support agent host configuration changes.

During Snapshot: Redesigning the ACE/Server Infrastructure

Business Continuity Planning

Although the revised Security Essentials curriculum with CISSP Common Body of Knowledge material was not yet available when this case study was started, numerous good sources of information on Business Continuity Planning was easily found. The All In One CISSP Certification Exam Guide provided an easily accessible review of disaster recovery and business continuity best practices (Harris 591-633). Contingency Planning & Management Online provided a more industry-specific view, and granted access to its BCP Handbook and Knowledge Base after free user registration.

All of the sources that I consulted recommended similar basic approaches for the development of a comprehensive Business Continuity Plan. Gan Chee Syong's Introduction to Business Continuity Planning outlined the following phases – Project Initiation, Business Analysis, Design and Development (Designing the Plan), Implementation (Creating the Plan), Testing, and Maintenance (Updating the Plan).

ACME BCP Approach

With the clear mandate of executive management to develop and implement a comprehensive BCP in place (often the most difficult prerequisite to satisfy), the project was formally initiated with the establishment of a Steering Committee, retainer of an IT consulting services firm and selection of a cross-functional project management team. ACME then proceeded to have not just one, but a few Business Impact Assessments performed by different third parties – including one by its disaster recovery services provider. Since the overall objective of the project had already been set by executive management (as is often the case with large corporate projects), the purpose of the multiple BIA were to ensure that all current critical business processes were identified and assessed – not just those that had been reviewed in the past. Although I did not get an opportunity to contribute to or even see any of the BIA, as a production application owner, I was involved during the Design and Development and subsequent phases.

The overall objective of ACME's BCP contained two components, one tactical and one strategic – (1) migrate all production applications to a new Primary Data Center in the Southwest, and (2) enable 48-hour production failover/recovery capability to the existing, now classified as Secondary, Data Center in California. For distributed application architectures, only the master or primary server had to be migrated. Since hundreds of application hosts would have to be migrated without prolonged impact to business operations, it was clear that simply relocating the existing production hosts would involve too much risk and downtime. Instead, new hardware that met or exceeded existing application requirements would be purchased, installed and configured. This would allow the production applications to be migrated individually or in clusters at a scheduled time.

This migration approach was a perfect fit for ACE/Server. There was already an identified need to migrate the application to new, dedicated hardware to eliminate resource contention issues with another application on its current shared host. And, there is no safer way to perform a major application upgrade than to have two separate sets of hardware to facilitate a clean cutover. This allows the original production application installation to remain untouched and facilitate a simple fallback strategy in the event some unforeseen issues should happen to arise.

ACE/Server 5.0 New Features

ACE/Server 5.0 provides many new features including a new Web-based administration interface, LDAP v3 support, and High Availability hardware support. However, the most significant improvement is a new realm architecture that supports database replication from a primary server to up to ten (10) replica servers – each capable of supporting concurrent authentication. Whereas a separate ACE/Server realm was previously required to support the addition of an authentication point, a single ACE 5.0 replica server can replace an ACE 4.x master and slave pair. Additionally, since each replica server maintains a complete, nearly up-to-date, copy of the realm database through database replication with its primary server, it is now possible to accomplish disaster recovery with minimal data loss without reliance on backup tapes. More detailed information about these new features and benefits can be found on the RSA Security Website.

ACME ACE/Server Infrastructure Design Considerations

Because other production application and services are directly dependent on ACE/Server for strong authentication services, it is considered by ACME to be a supporting infrastructure component. As such, under the new business continuity model, it must be recoverable within six hours of Disaster Declaration[3]. While this recovery time requirement has significant implications, the basic business requirement

[3] Two key assumptions are required for such a short recovery window – (1) local staff trained and available to perform recovery activities at the Secondary Data Center (thereby minimizing travel time), and (2) advanced notification to designated recovery staff (prior to the Disaster Declaration) to maximize preparation time.

to provide robust strong authentication services for dependent applications and services is what drives the infrastructure design.

First, there needs to be a primary server, which must be placed in the Southwest (Intranet) to satisfy the location requirement for a primary application server. While the primary ACE/Server can also support authentication, it is preferable to limit its authentication duties to maximize resource availability for administration and database replication functions. Placing the first replica server also in the Primary (Southwest) Data Center is sufficient to relieve the Primary server of most authentication duties and provides a dedicated authentication point for the future production application hosts. While agent hosts running ACE/Agent 5.0 or later will periodically poll all the authentication servers to dynamically determine which one provides the best response time – usually a replica since the Primary will always be busier performing its other duties – a legacy ACE/Agent[4] should be statically configured to authenticate only against particular replica servers.

Next, two replica servers are required in the Secondary (California) Data Center to replace the existing ACE/Server 4.x realms supporting California Intranet and DMZ/ Extranet hosts. Since a new Internet connection will also be established in the Primary Data Center to provide direct Internet access, a fourth replica server is required to support authentication requests from new Primary DMZ hosts and redirected authentication requests from extranet hosts. Placing the four replica servers in these locations will provide a local primary authentication point for the majority of the dependent agent hosts on the ACME enterprise network. This will minimize the amount of authentication traffic that has to cross the WAN, and provide up to three secondary authentication points for remote agent hosts should its assigned local replica server experience a failure. Unfortunately, legacy ACE/Agents can only have one secondary authentication server defined – its Acting Slave server.

Finally, although it will normally not be a part of the production ACE/Server infrastructure, one more server – a development host running on production-class hardware – is also required. Traditionally, ACE/Server development and test activities are performed on workstation-class hardware, as this is usually sufficient for the functional testing of patches, hot fixes and configuration changes. However, having a development server, running on production-class hardware, configured to mirror the production Primary ACE/Server provides a capability to replace the lost Primary server with no more than network access to a surviving replica server. And provided that the replica servers have been distributed to different geographic locations, the availability of network connectivity should be the only limiting factor.

Recommended Application Configurations

Beyond the standard ACE/Server 5.0 installation, three application configurations are recommended to provide greater administrative control of the application. The last two

[4] An ACE/Agent prior to version 5.0 that is unaware of all the realm authentication servers, which has to be configured to authenticate against specific Acting Master and Acting Slave servers.

configurations require a couple of patches and hot fixes be installed, but all of these were incorporated into the current ACE/Server release – version 5.0.03.

1. Disable the System Parameter option – Allow Push DB Assisted Recovery.

This prevents the Primary server from automatically distributing the latest database files to a replica server. This is only applicable during installation or database recovery, and does not affect normal database replication. This ensures that large database file transfers do not go out over the network without administrator involvement.

2. In the ACE/Server configuration (for UNIX), choose **No** to "Resolve hosts and services by name?" so that the resulting sdconf.rec file shows for Addresses: **By IP address in RSA ACE/Server database**.

This not only eliminates some name resolution issues but also allows an agent host to undergo a hostname change without impacting its node secret encryption key (provided that its IP address does not change). This is extremely useful if your failover strategy involves the use of development hosts to recover your production hosts without an IP address change. For example, dev-host at IP 10.0.0.1 is reconfigured to be prod-host at 10.0.0.1 in order to facilitate the recovery the lost prod-host at IP 192.168.0.1.

3. On all (UNIX) ACE/Server hosts (Primary and replicas):
   a. Configure /etc/syslog.conf to log user.info to a log file
   b. Define the environment variable SDI_ASD_SYSLOG

This allows detailed database replication information to be written to the specified log file. This information is very useful for identifying database replication issues, and can be used to determine which replica server last successfully replicated with the Primary server (RSA ACE/Server 5.0 Administration Manual 189). As data is written to the log files on both the Primary and replica server for each database replication pass (every 1-2 minutes), it is very important that your log file be located on a large enough partition to support file growth. This is especially important on a Primary server with numerous replicas.

Infrastructure Design Approval

Once the new ACE/Server infrastructure design was completed, it had to be approved by two groups. The first group was the IT Architecture Review Board. This governing body is chartered with the responsibility of ensuring that any proposed IT solution is designed with an architecture that is both compatible with the company's long-term strategic direction and flexible enough to meet current and emerging business requirements. Since the new ACE/Server infrastructure design is essentially a simplification of the original design – using 5.0 replicas to replace secondary 4.x realms – with the addition of two replica servers, there was no difficulty in obtaining approval for this Single ACE/Server Realm design.

In order to provide a baseline from which the Production Migration Project could work, the configuration for all production application infrastructures were frozen by executive management mandate. Any changes to existing configurations had to be approved by the Production Migration design team and integrated into both the project migration project plan and the BCP. Since the Primary Data Center facility would not be ready for occupation until after September 2002, but there were immediate tangible benefits that could be realized from an earlier implementation, approval was granted to perform the ACE/Server 5.0 upgrade/migration in May.

Deploying the New Infrastructure

Once the specific funding details were worked out, it was determined that three servers would initially be ordered and deployed to support the main application upgrade in May – a new Primary server in the Southwest, and two servers in California to replace the existing production ACE/Server realms. The remaining three servers would be added to this initial infrastructure after the new Primary Data Center is ready. However, one technical consideration still needed to be worked out – how to configure the new Primary server with an IP address for a facility that did not yet exist – so that a reinstallation of the Primary Server and resulting redeployment of the infrastructure would not be required later in the year.

The workaround was actually quite simple. The network management team set up a small, dedicated VLAN in an existing Southwest data facility that could be reconfigured to be part of the new Primary Data Center network space. This allowed a new three-server realm to be established onto which a copy of the production ACE/Server database was loaded on the day of the May cutover. And with the direction to system and application owners to reconfigure their existing ACE/Agents to authenticate against the new servers in both California and the Southwest, a significant reduction in business risk was immediately realized[5].

Phase II of the deployment had to wait for the new Primary (Southwest) Data Center to be completed. A logical move was then required to physically relocate the Primary server hardware and its VLAN to the new facility – but no system or application reconfiguration was required. The two new replica servers were then added to the infrastructure. Once this final infrastructure was completed, another round of ACE/Agent reconfiguration was performed to relieve the Primary server of its authentication duties[6]. And finally, the new development server was installed on production-class hardware, configured to mirror the production Primary server configuration, and made ready to support failover/recovery activities.

[5] Since the Primary server was the only available authentication server in the Southwest, it had to hold authentication duties until the Phase II replica servers could be deployed. Legacy ACE/Agents were configured to use it as either the Acting Master or Acting Slave server (depending on their location).

[6] Legacy ACE/Agents were reconfigured to use one replica in Primary Data Center and one replica in the Secondary Data Center for their Acting Servers.

After Snapshot: New Capabilities

ACME BCP Profile – January 2003

By the end of December 2002, all production applications were successfully migrated to the Primary Data Center in the Southwest, with only distributed/redundant production and development application hosts remaining at the Secondary Data Center in California. IT Management-level disaster declaration and response procedures were documented and tested in tabletop disaster scenario exercises involving cross-functional teams; and business continuity goals and expectations had been communicated to all application owners.

Although the processes used to perform the production applications migration had provided a detailed roadmap that could be leveraged for disaster recovery, all application owners were tasked with the creation of additional documentation to prove that their applications could be failed-over to the Secondary Data Center within 48-hours of disaster declaration (within 6-hours for supporting infrastructure components). This documentation had to include detailed failover/recovery procedures, human resources requirements, vendor dependencies, and time estimates; all of which would be compiled into a master BCP Disaster Recovery playbook. The results from actual unit recovery tests were also requested though not required.

ACME ACE/Server Infrastructure – January 2003

The ACE/Server 5.0 infrastructure was built exactly as designed, consisting of one Primary Server in the Primary Data Center dedicated to supporting administration and database replication, and four replica servers – two in the Primary Data Center and two in the Secondary Data Center – to support strong authentication requirements.

Figure 2 illustrates the ACME production ACE/Server 5.0 infrastructure.

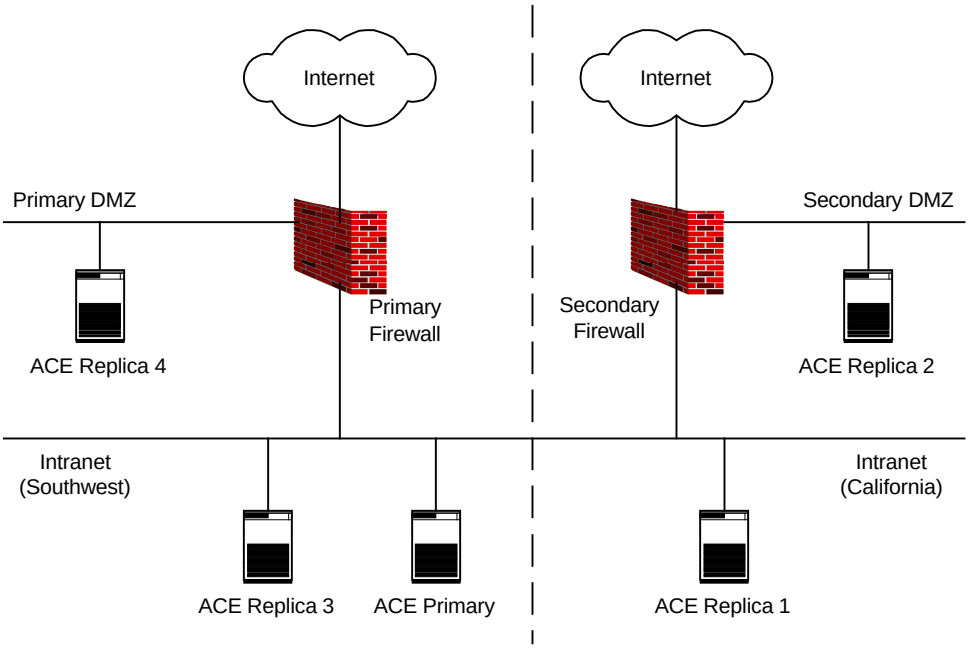

Figure 2

ACE/Server Disaster Recovery Capabilities

Since all agent hosts running ACE/Agent 5.0 or later are able to authenticate against any replica server and all legacy ACE/Agent hosts are specifically configured to only use replica servers as their Acting Master and Acting Slave servers, the lost of the production Primary server in itself does not warrant immediate execution of ACE/Server failover. Existing users will still be able to continue authenticating through existing agent hosts, with only administration capabilities impacted. Unless there is a critical need to recover administration functions, it may be preferable to bear with such a disability rather than invest the effort (and assume the associated impact and risks) required to perform the fail-over and subsequent fail-back activities.

With the finalization of the production ACE/Server 5.0 infrastructure, and the configuration of the development Primary server in the Secondary Data Center to match the production Primary server in the Primary Data Center, only a final review of the recovery requirements and key assumptions was needed to ensure that the ACE/Server Failover Procedures would meet expectations.

Primary Requirements: Fail-over the Primary ACE/Server and restore administration capabilities in the Secondary Data Center, and update all surviving replica servers within six hours of Primary Data Center disaster declaration.

Key Assumptions: Survival of the development Primary server and at least one of the replica servers in the Secondary Data Center, and available network connectivity between both servers.

Although there are no vendor dependencies, one additional proactive change is recommended to ensure that the failed-over Primary server (in the Secondary Data Center) can update the two replicas located in the Primary and Secondary DMZ. One additional access rule must be added each firewalls to allow the development/failed-over Primary server to communicate via the ACE/Server database replication ports to each replica server.

Based upon the time required for one trained administrator to perform similar tasks, it is estimated that execution of the detailed ACE/Server failover procedures should allow the Primary server to be recovered within one-hour of start time, followed by the update of one surviving replica server every half-hour afterwards (not including database file transfer time).

ACE/Server 5.0 Failover Procedures

Verification Phase

1. Confirm the directive to execute ACE/Server failover.
2. Verify that the production Primary server is inaccessible.
3. Determine which production replica servers have survived.

Primary Server Recovery Phase

4. Log on to each surviving replica server and determine which one last replicated with the production Primary server.
5. Decide which replica server will be used to recover the Primary server. The best choice is the one that last replicated. However, selecting a replica that is closer on the network may reduce file transfer time.
6. Stop the ACE/Server processes on the selected replica server.
7. Run **sddump –s** to create a dump file of the server database (sdserv.dmp).
8. Restart the ACE/Server processes on the replica server so that it can continue to support existing user and agent host authentication.
9. Log onto the development Primary server, and stop the ACE/Server processes.
10. Back up the existing $ACEHOME/ace directory.
11. Run **sdnewdb**, and create both new server and log databases.
12. Use a secure file transfer protocol to retrieve the server database dump file created in step 6 and the license.rec file (in binary mode) from the selected replica server.
13. Run **sdload –s –f sdserv.dmp –k license.rec** to load the database.
14. Go back to each surviving replica server, and stop the ACE/Server processes.
15. Go back to the now-production Primary server, and restart the ACE/Server processes.
16. Verify local administration functionality using sdadmin.
17. Go to your ACE/Server Database Administration Remote Mode client host.

18. Use a secure file transfer protocol to retrieve the sdconf.rec and server.cer files from the Primary Server (in binary mode) and update your Remote Administration client.
19. Verify Remote Administration functionality.

Surviving Replica Update Phase

20. On the Primary server, stop the ACE/Server processes.
21. Run **sdsetup –package** to create one replica package for every surviving replica.
22. Use tar (or something equivalent) to consolidate all the replica package files into one bundle.
23. Restart the ACE/Server processes, then run **tail –f <log file>** on the file that was configured to hold database replication information.
24. Perform steps 25 – 32 for each surviving replica server.

25. Go back to a surviving replica that you want to update.
26. Back up the existing $ACEHOME/ace directory.
27. Use a secure file transfer protocol to retrieve the replica package bundle created in step 21 (in binary mode) from the Primary server.
28. Extract the replica package files, and copy them to the $ACEDATA directory.
29. Run **sdsetup –config** and accept all the defaults
30. Restart the ACE/Server processes, then run **tail –f <log file>** on the file that was configured to hold database replication information.
31. Verify that the Primary server and this replica are able to reconcile databases.
32. Configure a test agent host and verify that you can successfully authenticate against this replica server.

Agent Host Recovery Phase

The steps required to recover an agent host will vary depending on the specific recovery approach used and the specific configuration of that particular ACE/Agent. In general, on the ACE/Server side, it will typically involve the modification of existing agent host entries – hostname/IP address or Acting Servers configurations – and possibly the need to clear and re-establish a new node secret encryption key.

Next Steps: BCP Testing and Maintenance

At the time of this submission, a master Disaster Recovery playbook is still in the process of being assembled by the ACME BCP team. The expectation is that within the next three months a comprehensive disaster recovery scenario test will be conducted to verify disaster recovery capabilities, and provide a baseline for plan maintenance and continuous improvement.

References

Hannig, Scott. “Recovering From Disaster: Implementing Disaster Recovery Plans Following Terrorism.” SANS Info Sec Reading Room. 21 Sept. 2001. 24 Sept. 2002 <http://rr.sans.org/recovery/terrorism.php>

Harris, Shon. All In One CISSP Certification Exam Guide. New York: McGraw-Hill / Osborne, 2002.

Syong, Gan Chee. “Introduction to Business Continuity Planning.” SANS Info Sec Reading Room. 1 Oct. 2001. 24 Sept. 2002 <http://rr.sans.org/recovery/continuity.php>

Contingency Planning & Management Online Home Page. 2 Oct. 2002 <http://www.contingencyplanning.com/>

RSA ACE/Server 5.0 Administration Manual. Bedford: RSA Security, 2001.

RSA ACE/Server 5.0.1 for Windows and UNIX README. Bedford: RSA Security, 2001.

RSA ACE/Server 5.0 for UNIX Patch 03 Readme. Bedford: RSA Security, 2002.

“RSA Security ACE/Server – What’s new in RSA ACE/Server 5.0?” RSA Security. 2002. 4 Oct. 2002 <http://www.rsasecurity.com/products/securid/whatsnewACE50.html>

“RSA Security Introduces Next-Generation Authentication Security Server.” RSA Security. 04 June 2001. 4 Oct. 2002 <http://www.rsasecurity.com/company/news/releases/pr.asp?doc_id=928>