IBM Information Technology Services

Lessons from a Windows to Linux/Samba Consolidation in a Large Public School District

John Janosik – jpjanosi@us.ibm.com

samba eXPerience 2006

April 26, 2006

Agenda

- Starting environment
- Architected solution
- File server migrations
- Domain migrations
- Performance issues encountered
- Lessons learned

Environment 2Q 2005

- **3 NT Domains with a total of ~70,000 user/computer accounts**
  - North, South, and Staff domains each with two-way trust to the other two domains plus an Active Directory admin domain not being migrated
  - Staff domain has a two-way trust to two other AD domains maintained by groups outside the IT department that will not be migrated
  - Win2k WINS servers were not being migrated

Environment 2Q 2005(Continued)

- ~4 terabytes of data on 36 NT and win2k servers
- Servers spread across 10 locations with no IT staff on site
- 100Mb network bandwidth to most locations
- ~30,000 client machines at ~100 sites. Mix of Win98, 2k, XP, and MacOS X

New Environment Goals

- Avoid MS license costs of moving to Win2k
- Eliminate extended hardware outages
- Improve efficiency of administrators
- Eliminate trips to remote servers
- Disaster recovery plan in event a datacenter is lost

New Environment

- **2 physical locations**
  - Network at the chosen sites already upgraded to redundant 1Gb backbone
- **At each location**
  - All machines running RedHat AS 3
  - IBM BladeCenter with 3 dual CPU HS20 blades for domain controllers
    - Samba 3.0.14a with OpenLDAP/bdb backend
  - 2 IBM x445 file servers (4 CPU, 8GB RAM)
    - Steeleye Lifekeeper for High Availability (HA)
    - Samba 3.0.14a configured as a domain member server
  - IBM DS4500 fiber channel attached storage for x445s

New Environment (Continued)

- **Normally server north1 has the shared1 resource**
  - Lifekeeper controls the following
    - Shared1 IP
    - /shared1 and /shared2 – 2TB filesystems
    - Samba smbd/nmbd processes with config file that binds only to the shared1 IP address and exports only the shared1 filesytem
- **Normally server north2 has the home1 resource**
  - Lifekeeper controls the following
    - Home1 IP
    - /home1 and /home2 – 2TB filesystems
    - Samba smbd/nmbd processes with config file that binds only to the home1 IP address and exports only the home1 filesystem
- **The north1 & north2 names are joined to the domain and winbindd is running on each machine outside of lifekeeper control**
  - Winbind is configured with the ldap idmap
  - Users do not know about, nor connect to the north1 and north2 names
- **The south datacenter has an identical configuration but holds the shared2 and home2 resources**

High Availability

- **An automated process keeps Samba configuration files for all four resources in sync on all file servers**
- **Rsync is used to backup the exported filesystems in one datacenter to the non-exported filesystems in the other datacenter**
- **During a failover, any of the resources can be made available from any of the 4 file servers**
- **Lifekeeper only automatically handles moving a resource between servers in the same datacenter**
- **Failing over between datacenters was a manual process**
  - DNS for the IP address of the resource needed updated because the north and south datacenters were on different subnets.
  - The data is not up to date since the rysnc backup is not done during the business day

File Server Migration Procedure

- Created top level directories under /sharedX and /homeX and set initial ACLs
- Used smbfs to mount Windows server and rsync to copy data during week
- Changed ACL on Windows server shares to disable user access and did a final rsync during the outage window
- “Net rpc share migrate” wasn’t used due to slow windows servers and poor network bandwidth causing the file migration time to exceed the outage window

NT to Samba Domain Migration

- Customer outage window again limited our migration options
- “net rpc vampire” directly into ldap backend would not complete in our testing
- “net rpc vampire” into tdb backend, followed by “pdbedit” to export to ldap backend worked but took hours
- “net rpc vampire ldif” patch obtained from the IBM Linux Technology Center
- Final Procedure
  - **“net rpc vampire” to ldif took about 5 minutes**
  - **“slapadd” from ldif took about 15 minutes on each ldap server.**
  - **Shutdown Windows BDCs/PDCs**
  - **Start ldap/smbd/nmbd on domain controllers**
  - **Re-establish trusts**

Testing phase #1 – Basic function

- Intermittent delays mapping drives
- The root cause was lookup of large groups not in the winbind cache
- Even with “ldapsam:trusted = yes” on the domain controllers it could take minutes to get the membership list of “domain users”
- High school groups had thousands of users and could take 10 seconds to look up
- This problem was worst when linux admins did “ls –l” in a directory where files were owned by “domain users” or when they were using getfacl to view acls that contained high school groups

Workarounds – Issue #1

- **Avoid looking up large groups**
  - Make group owner of all files a local group and use sticky bit on directories to keep new files and directories owned by the local group
  - Put users in different default groups
- **Eliminate winbind from nsswitch by building local passwd and group files.**
  - There were trusted windows AD domains outside our control
- **Selected workaround was to increase the winbind cache time to 1 day and use scripts to prime the cache with large groups nightly**

Testing Phase 2 – Load Tests

- **The customer wanted to support 12,000 concurrent active clients with drives mapped, with connections spread over 30 minutes**
- **Linux boxes as test clients with a distributed shell to kick off tests**
  - Domain controllers were tested with the smbtorture xplogin test from samba4
    - 20,000 domain logins in 13 minutes achieved
  - File server connections were tested with smbclient running a script getting and putting files
- **There were intermittent problems with logins failing. Calls to winbind to look up the users were failing when winbind was under load**
  - Winbind client code was updated to retry on failure, bugzilla 2736
- **The ext3 journal size had to be increased**
- **Memory had to be upgraded to 16GB**

Winbind performance in production

- **The migration went smoothly after testing and the solution was in production before the start of the 2005-2006 school year**
- **Performance problems were hit October 2005**
  - Winbind was CPU bound
  - Root cause was customer addition of “valid users” containing large groups into their smb.conf on home shares
  - IBM LTC provided improvements in winbind efficiency pulling entries from the winbind cache for Samba 3.0.21

Lessons

- **Three large trusted domains made it hard to get away from using winbindd in nsswitch for all domains**
- **Consolidation into one large domain and using “winbind trusted domains only = yes” on the member file servers might have worked**
  - The trusted AD domains were small so using winbind in nsswitch for those would not have caused performance issues
  - Local files/groups could have been built frequently from ldap or the file servers could have been OpenLDAP replicas
  - The domain controllers handled the load well. Reducing the number of domain controllers and increasing the number of file servers would a better use of resources

Questions?

John Janosik – jpjanosi@us.ibm.com