IDIAP COMMUNICATION

The IDIAP Multimedia File Server

Frank Formaz [a],Norbert.Crettol [a]

IDIAP–Com 04-05

June 2004

[a] {Frank.Formaz,Norbert.Crettol}@idiap.ch

IDIAP Research Institute **www.idiap.ch**
Rue du Simplon 4 P.O. Box 592 1920 Martigny – Switzerland
Tel: +41 27 721 77 11 Fax: +41 27 721 77 12 Email: info@idiap.ch

Contents

1 Abstract **2**

2 Introduction **2**

3 Needs and Requirements **2**

4 Solution developed **2**
4.1 First settings . . . 2
4.2 New Solution Description . . . 3
4.2.1 RAID Adapter and disks . . . 3
4.2.2 Mainboard and memory . . . 3
4.2.3 Network adapter . . . 4
4.2.4 Case . . . 4
4.3 System Configuration . . . 4
4.4 Services . . . 4
4.4.1 www . . . 5
4.4.2 ftp . . . 5
4.4.3 rtsp . . . 5
4.4.4 jsp . . . 5
4.4.5 network access . . . 5
4.4.6 filemanager . . . 6
4.5 Performances . . . 6

5 Data Processing **6**

6 Conclusion **7**

7 Acknowledgements **7**

1 Abstract

The IDIAP Multimedia File Server is a system designed to store and share all the data generated by our Smart Meeting Room [10] and our partners using different ways. This document presents a detailed description of the system including the hardware and the services.

2 Introduction

In addition of our Smart Meeting Room we needed to setup a system to share the recorded data to our partners. Because of the size of those data (detail below) and the variety of required services, it wasn't possible to use one simple computer. That's why a completely new solution has been set.

3 Needs and Requirements

After a first set of scripted meetings [8] some details for the implementation have been fixed. The raw video data will not any more stored but only the original DV tapes are kept as backup. This will save a lot of disk space as DivX compression will be used to compressed the raw video data. The following parameters have been chosen for the DivX tunning: bitrate 2300 and keyframes 300 which reduce dramatically the data size (raw data 250MB/min/camera divx data 17MB/min/camera) but keep a enough good quality for further processing. For the audio, the original 48kHz recordings are downsampled to 16kHz and original files are not saved.

In order to share the data, various capabilities are needed. Firstly all the "public" data have to be available from the web and thought ftp. In addition, to allow people to have a first look on the data, realmedia files have also to be provided.

Finally, this system should be available to other partner though protected areas and has to be OS independent.

Internally the data have to be shared to either Unix or Windows people with restricted rights. Some security policies have been defined in order to recover from a possible lost of data. All the audio and video data are burned on DVD-R to allow a fast recovery of original data in case of disater of to share data to partners. Because the size of this system, a backup with tape will be as expensive as the complete system so we decided not to implement such a solution. That's why the system has to be hardware defect robust and a weekly backup should also be available to avoid user wrong manipulations..

4 Solution developed

For either cost or performances reasons, a linux platform has been chosen for this project with usage as much as possible to opensoftwares. The main reason for this choice is the the fact that people using various systems have to be able to deal with the data and we should not be limitated by specific licenses in term of users number.

4.1 First settings

A first system based on an old PIII-500 PC with 2 NAS of 320 GB mirrored once a week has been set with the rsync software [13]. Rapidly the limits in terms of performances and disk size were reached but all the required services were present for a really low cost as it was common hardware. The major problem for this system was the security because of the difficulty to manage access on the data. Due to this problem it was not possible to internally share data to people however everything was available on the web or throught ftp. For all those reasons it has been decided to develop a new dedicated system able to manage all those features.

4.2 New Solution Description

After looking on commercial products able to share TB of data and able to be flexible enough to provide easily new services for an affordable price we were in front of a problem because nothing was really available. So we decided to start from scratch in order to exactely match our needs. The goal was to obtain a system as much flexible as possible with high performances in term of disk access, speed of network shares access and local application running. Last but no least we started this project in middle of year 2003.

4.2.1 RAID Adapter and disks

As one of the most important point was the financial aspect of the project, we decided to have a look on IDE disks intead of SCSI or FC one however there are more often used for this type of application. The difference in term of price can be a coefficient of 4 to much more. Different manufacturers provide RAID adapters which support IDE disks. We choose finally 3ware for the following reasons:

- RAID 5 support with dedicated controllers supported on linux (up to 12 disks/controller)
- Hotswap and hotspare support.
- Good linux support

The fourth generation of those cards is supporting the new SATA disks which seems to be the future reference for low price storage. However the prices are a little bit more expensive (6than normal IDE disks,we have the big advantage of a better performance in addition to the much more smaller cables size which is better for the ventilation of a small rack case 4.2.4. The final choise has so be done on the 3ware Escalade 8506-8 SATA RAID Controller [3].

Figure 1: 3ware Escalade 8506-8 SATA RAID Controller

In order to avoid later problems if we have to contact the 3ware support we also have to choose a disk which is known as supported by 3ware [1]. The final choice has been done on a 250GB disk from Maxtor [9].

4.2.2 Mainboard and memory

As for the disk choice, the mainboard need also to be known as compatible with the 3ware card. A compatibility list is provided [2] and so the choice has to be done between an Intel or AMD architecture. The Intel Xeon architecture is finally chosen because of it's better performance treating I/O. The final choice has been done on the SuperMicro X5DPE-G2 motherboard [15].

To provide a good solution to local applications running at the same time than data sharing and other services like www, ftp, rtsp or filemanager, we need a computer with enough memory and powerfull CPU. That's why we choose to install two Xeon 2.66GHz with 2GB DIMM (2 x Kingston 1GB DDR 266MHz CL2.5).

Figure 2: SuperMicro X5DPE-G2 Motherboard

4.2.3 Network adapter

The two onboard gigabit network ports are based on an Intel chipset which is not only well known but also one of the better supported ethernet chipset on linux and one of the most efficient [12].

4.2.4 Case

As case we have to concider a rackmount system which can be used with a dual Xeon motherboard and at least 16 hotswap disks. The final choice has been done on a 3U rackmount case from AIC [4].

This system is robust to fan or power supply defect which was one of our requirement. This rack can also aquire a small 2.5 inches HDD which will store the system. This will simplify our life for the configuration and let us the 16*250GB only for the data.

4.3 System Configuration

From the raw capacity of 4T (16*250GB) we setup 2 RAID 5 volumes of 8 disks (1 parity, 1 spare, 6 data) using the ext3 filesystem. Those two 1.5GB volumes are then merged into one simple RAID 0 volume of 3TB. Using the LVM capabilities available on the system, we are able to set smaller logical volumes into this space and change their size on the fly. We finally set 3 volumes. One for the system backup, one for the mmm data and the last one for weekly backups in order to provide a user misusage reliability. This latest volume can be deleted at any time in case of new capacity needs for mmm. With this setting we are robust to disk crash and in case of problem we can replace defect hardware without system interrupt. In case of system disk crash, we setup a mirror on the data volume which can also be use to boot or correct eventual problems. We are also able to boot with a CD.

4.4 Services

After the hardware setting we started the services configuration. Based on a debian distribution, our linux box has been configured with following services:

Figure 3: The MMM server

4.4.1 www

Main way to access mmm data, this service is based on the well known apache [5] server. A simple interface has been designed to show the recorded meetings and powerfull tools are beeing developped to provide dynamic access to the whole data stored on the server (video, audio, white board, beamer, pen, meeting documents, ...).

4.4.2 ftp

The ftp protocol is provided by the very light but powerfull vsftp [17]. This server is not only one of the lightest on the market but also one of the most secure.

4.4.3 rtsp

In order to have a quick look on the data content without downloading the whole data (which is mandatory especially for partners who cannot download tenth of gigabytes over the internet), a video streaming system has been setup. Two main concurent are available on the market. The Windows Media Server from Microsoft and Real Server from Realnetworks [11]. The second has finally been chosen because of the availability of players on both unix and windoze world.

4.4.4 jsp

To develop some of the front end interfaces of the system, the use of java programming was necessary. That's why a tomcat5 [16] server has been setup on the system.

4.4.5 network access

The sharing of data is done by two ways. For unix people an internal NFS3 share is provided and for windows people the same data are shared using samba [14]. Different access rights such as read only or write only on some directories have been setup to allow clear distinction between different type of users (guest users, developer, data manager, administrator).

4.4.6 filemanager

For the M4 project partners, a protected area has been setup on top of the web server using a modified version of FileManager [7]. This interface allow authorized people not only to access private data but also to upload personal data on the server.

4.5 Performances

Perfs: some bench tests have been run on the system to check performances. In order to stress at maximum the system we used the bonnie++ benchmark suite [6]. We got so locally 280MB/s for read and 60MB/s for write (which correspond to services access such as web). Through the network we obtain 80MB/s for read and 15MB/s for write (which correspond to nfs or smb access).

5 Data Processing

After setting up the services, the use of them is the best way to be sure of their behaviour. In order to be usable on the final front end, the recorded data have to be prepared to deal with special formats. In addition data have to be provided in two modes. Previsualization mode and research mode. The first one is available on the web page and allows people to have a first look on the recorded data and so, the goal is to provided data the lighter as possible. But, in order to use those data for research purpose people need data the more detailed as possible and that's why we also have to provide data in high quality mode (wav for audio and DivX for video). The following picture 4 explains the whole preprocessing work which allow us to provide wav, divX, real and other files.

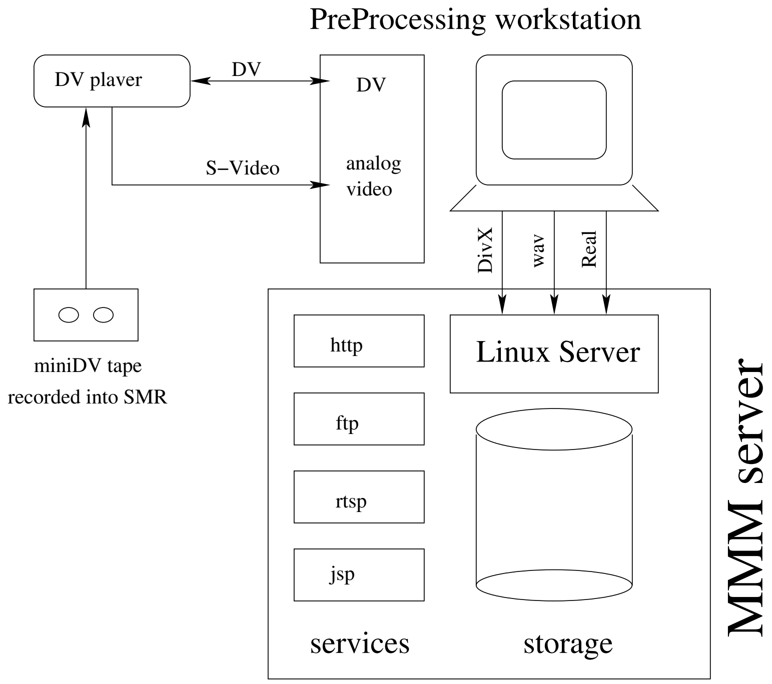

Figure 4: Data Processing

6 Conclusion

A fully icompact, usable and affordable 3TB system has been setup with hardware robustness and serious access control in order to manage the data recorded from our SMR in addition to other partners. Those data are shared to user using various services such as www, jsp, ftp, rtsp, nfs or smb and is flexible enough to deal with future extensions. In addition, as new products are coming on the market now using this technology (SATA RAID), it confirm us it was a good and efficient choice.

7 Acknowledgements

A big part of this project has been realized by Marcel de Riedmatten who managed all the initial setting of the System. Other people at IDIAP have also been involved in the installation and setting of the new Server. These people are: Pierre Wellner, Sbastien Marcel, Mael Guillemot, Mike Flynn and Thierry Collado.

References

[1] 3ware. *3ware PATA/SATA drive compatibility list.* http://3ware.com/products/pdf /Drive_compatibility_list.pdf.

[2] 3ware. *3ware tested motherboard compatibility list.* http://3ware.com/products/pdf /Motherboard_compatibility_list.pdf.

[3] 3ware. *Escalade 8506-8 SATA RAID Controller.* http://www.3ware.com/products/serial_ata.asp.

[4] AIC. *AIC RMC3E 3U Rackmount 16 x S-ATA Hot-Swap w/650W Redundant PS.* http://www.aicmicro.com.

[5] apache. *apache Project Page.* http://www.apache.org.

[6] Russell Coker. *Bonnie++ Benchmark Suite.* http://www.coker.com.au/bonnie++/.

[7] horsburg. *FileManager Project Page.* http://www.horsburgh.com/h_filemanager.html.

[8] D. Gatica-Perez I. McCowan and S. Bengio. Meeting data collection specifications. Technical Report IDIAP-COM-03-10, IDIAP, 2003.

[9] Maxtor. *Maxtor 250GB, 720RPM, 8MB cache SATA disk 7Y250MO.* http://www.maxtor.com/.

[10] D. Moore. The idiap smart meeting room. Technical Report IDIAP-COM-02-07, IDIAP, 2002.

[11] Real Networks. *Real Server.* http://www.real.com.

[12] A. Betz P. Gray. Performance evaluation of copper-based gigabit ethernet interfaces. In *27th Annual IEEE Conference on Local Computer Networks (LCN'02)*, Tampa, Florida, 2002.

[13] rsync. *Rsync Project Page.* http://rsync.samba.org/.

[14] samba. *Samba Project Page.* http://www.samba.org.

[15] SuperMicro. *Mainboard SuperMicro X5DPE-G2-O.* http://www.supermicro.com/PRODUCT/MotherBoards/E7501/X5DPE.htm.

[16] tomcat. *Tomcat Project Page.* http://jakarta.apache.org/tomcat.

[17] vsftpd. *vsftpd Project Page.* http://vsftpd.beasts.org/.