ADMINISTRATION RÉSEAU
ACCÈS AUX FICHIERS DISTANTS

A. Guermouche

Plan

Introduction

NFS

SAMBA

Plan

Introduction

NFS

SAMBA

Accès aux fichiers distants

Différences avec le transfert de fichier :

- l'accès aux fichiers distants est complètement transparent pour l'utilisateur
- tout se passe comme si le système de fichier distant était local
- l'utilisateur peut éditer le fichier, le modifier, . . . ; les modifications seront répercutées sur le système fichier distant

Les deux principaux protocoles :

NFS. Network File System (Unix/Sun-RPC)

SMB. Sserver Message Block (issu du monde Microsoft)

Les éléments d'accès aux fichiers

- Manipule des descripteurs, chemins, déplacements

Sytème d'exploitation

- Manipule des Files, Dentries, Inodes, déplacements
- Masque les différences à l'application (API uniforme, . . .)

Matériel

- Manipule des blocs
- Matériel-dépendant

Le choix entre NFS, EXT3,. . . se fait lors de l'ouverture du fichier.

source

Les supports suivants sont issus de

http://facweb.cs.depaul.edu/cwhite/TDC%20460/SAN.ppt

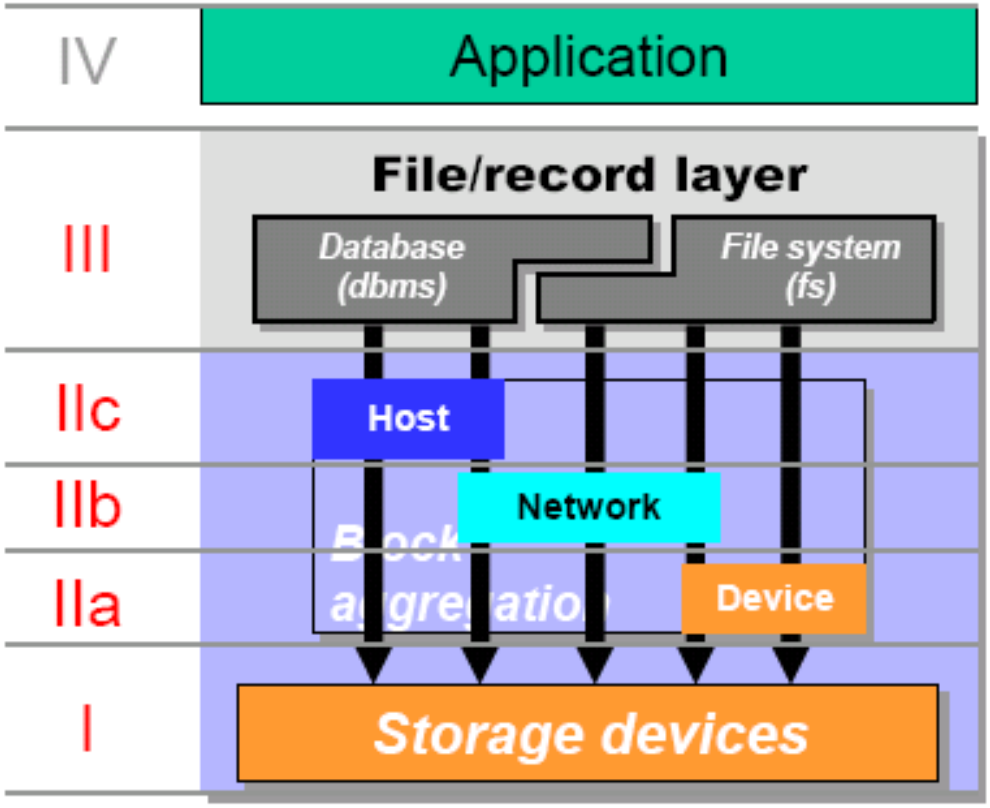
Outline

- Shared Storage Architecture
- Direct Access Storage (DAS)
  - SCSI
  - RAID
- Network Attached Storage (NAS)
- Storage Area Network (SAN)
  - Fiber Channel and
  - Fiber Channel Switch

The SNIA Model

- SNIA – Storage Networking Industry Association
- SNIA is a *framework* that captures the functional layers and properties of a storage system
- Trying to become an industry standard

The SNIA storage model
A layered view



The layers are as follows:

- IV. Application
- III. File/record layer
  - IIIb. Database
  - IIIa. File system
- II. Block aggregation layer, with three function-placements:
  - IIc. Host
  - IIb. Network
  - IIa. Device
- I. Storage devices

Three Basic Forms of Network Storage

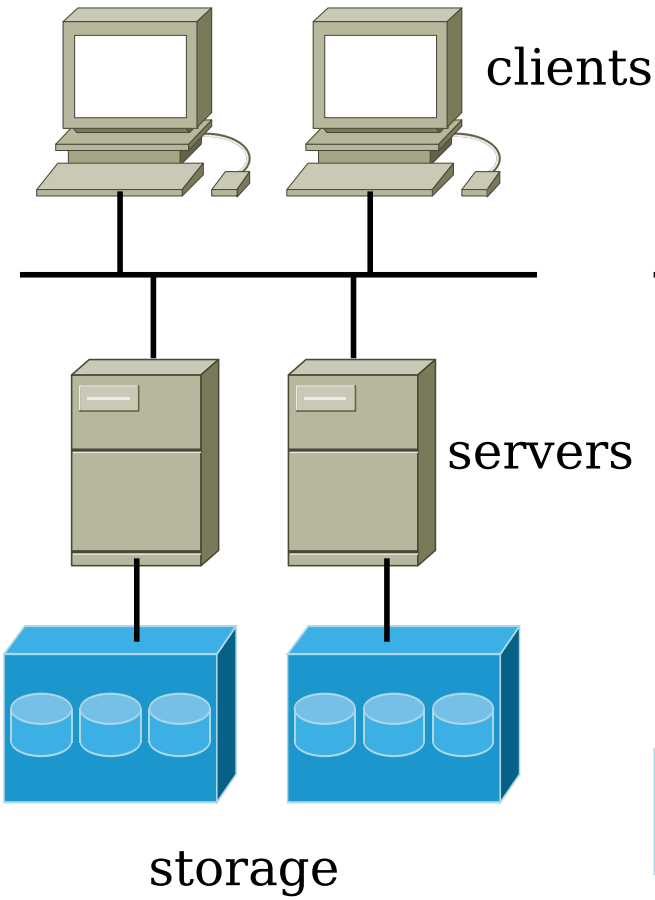
- Direct access storage (DAS)
- Network attached storage (NAS)
- Storage area network (SAN)

- And a number of variations on each (especially the last two)

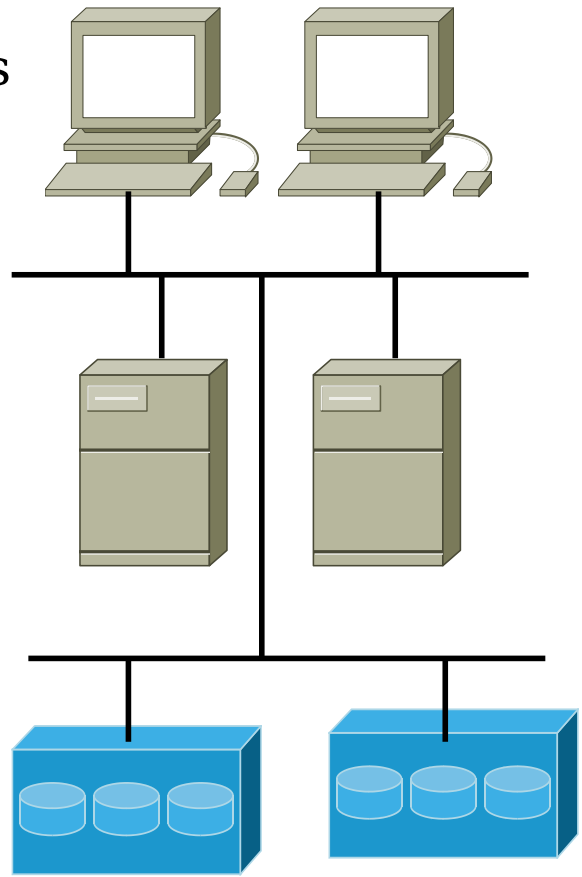
Quick Overview

	DAS	NAS	SAN
Storage Type	**sectors**	**shared files**	**blocks**
Data Transmission	**IDE/SCSI**	**TCP/IP, Ethernet**	**Fibre Channel**
Access Mode	**clients or servers**	**clients or servers**	**servers**
Capacity (bytes)	**10^9**	**10^9 - 10^{12}**	**>10^{12}**
Complexity	**Easy**	**Moderate**	**Difficult**
Management Cost (per GB)	**High**	**Moderate**	**Low**

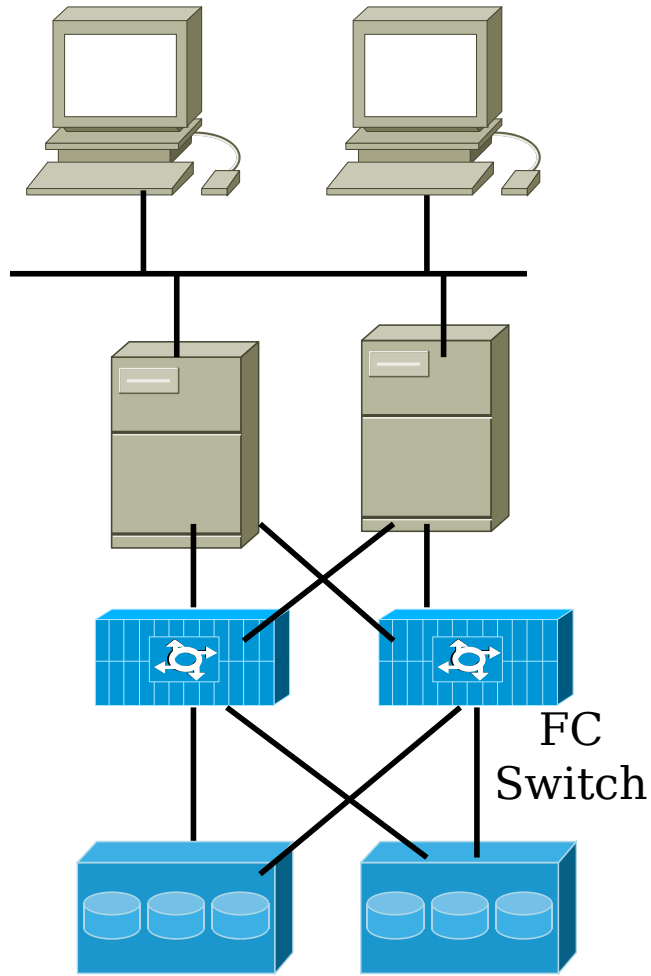
DAS



NAS



FC-SAN



Redundant Array of Independent Disks (RAID)

- A group of hard disks is called a disk array
- RAID combines a disk array into a single virtual device
  - called RAID drive
- Provide fault tolerance for shared data and applications
- Different implementations: Level 0-5
- Characteristics:
  - Storage Capacity
  - Speed: Fast Read and/or Fast Write
  - Resilience in the face of device failure

RAID Functions

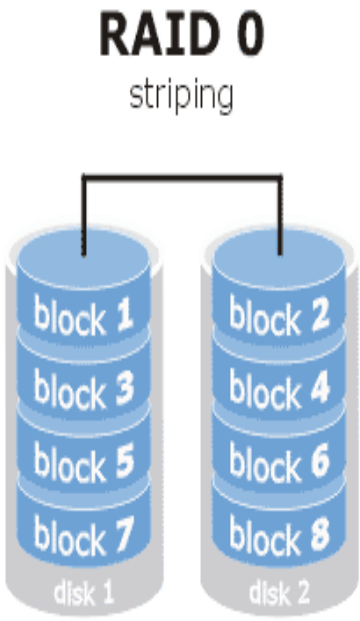
- **Striping**
  - Write consecutive logical byte/blocks on consecutive physical disks
- **Mirroring**
  - Write the same block on two or more physical disks
- **Parity Calculation**
  - Given N disks, N-1 consecutive blocks are data blocks, Nth block is for parity
  - When any of the N-1 data blocks is *altered*, N-2 XOR calculations are performed on these N-1 blocks
  - The Data Block(s) and Parity Block are written
  - *Destroy* one of these N blocks, and that block can be reconstructed using N-2 XOR calculations on the remaining N-1 blocks
  - Destroy two or more blocks – reconstruction is not possible

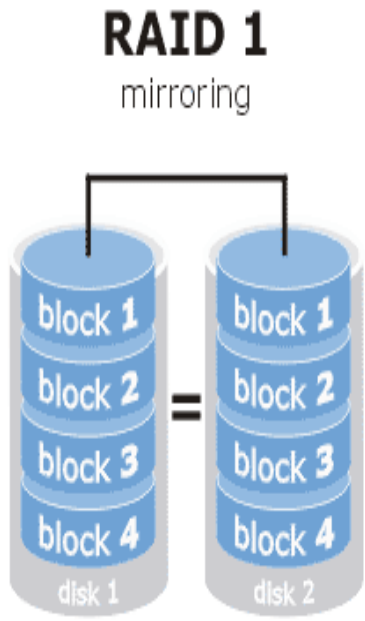
RAID Types

- **RAID 0**
  - Stripe with no parity (see next slide for figure)
- **RAID 1**
  - Mirror two or more disks
- **RAID 0+1 (or 1+0)**
  - Stripe and Mirrors
- **RAID 4**
  - Synchronous, Subdivided Block Access; Dedicated Parity Drive
- **RAID 5**
  - Like RAID 4, but parity striped across multiple drives

RAID 0

RAID 1

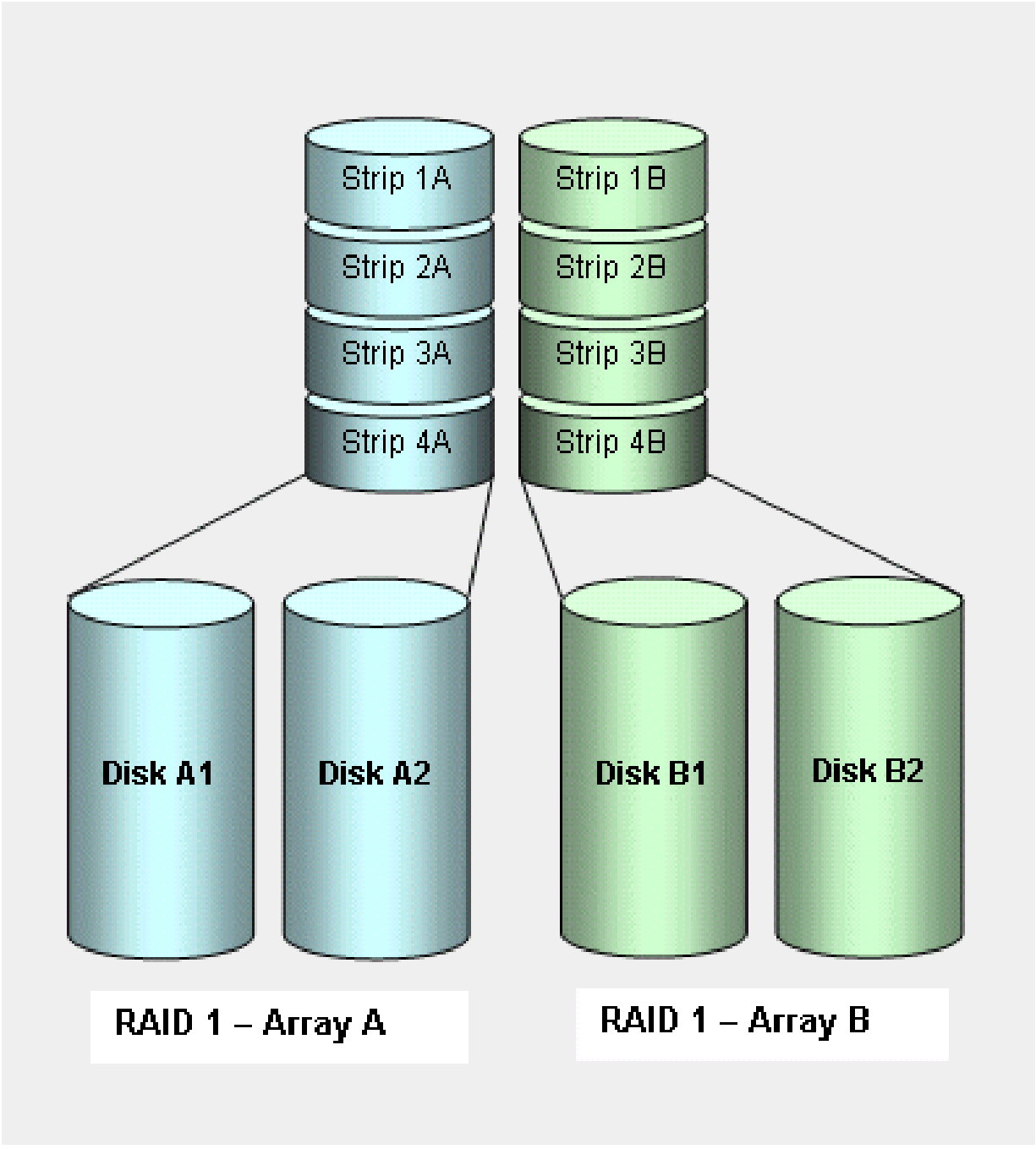




Disk Striping (no redundancy)

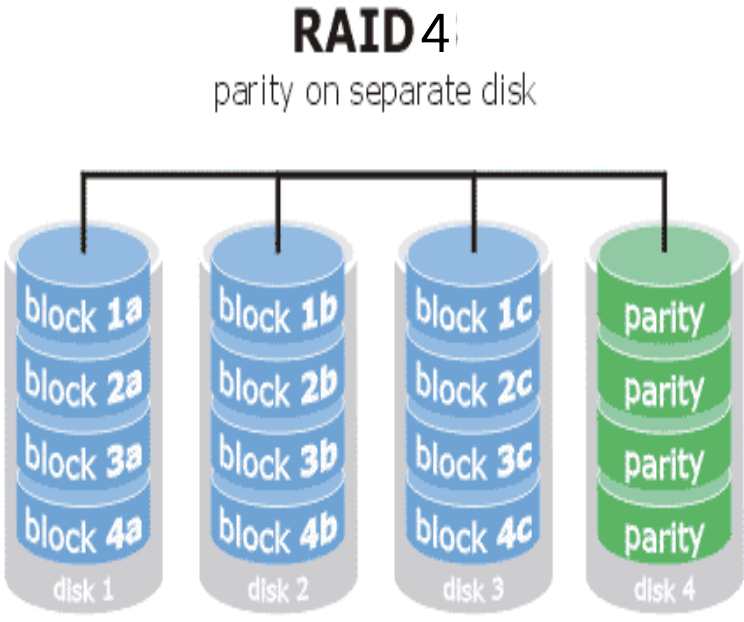
Disk Mirror

RAID 0+1 (or 1+0)

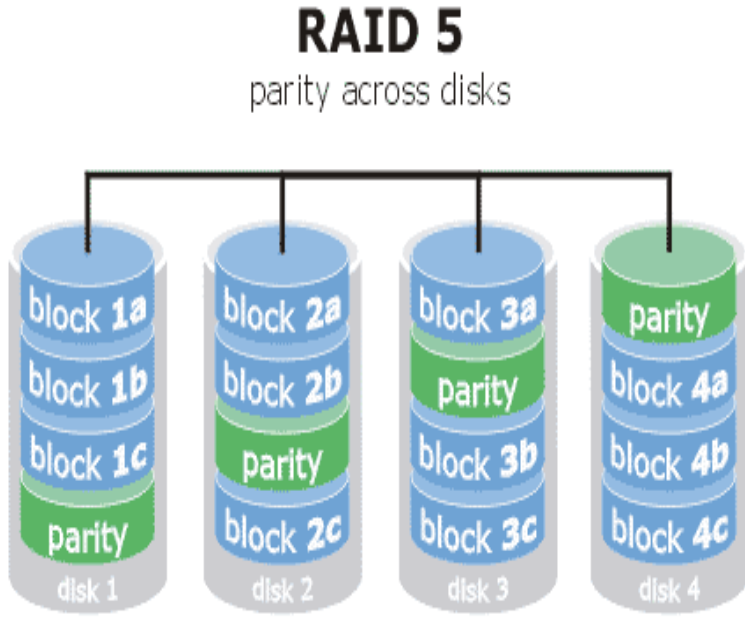

Strip 1A
Strip 2A
Strip 3A
Strip 4A
Strip 1B
Strip 2B
Strip 3B
Strip 4B
Disk A1
Disk A2
Disk B1
Disk B2
RAID 1 – Array A
RAID 1 – Array B

RAID 4

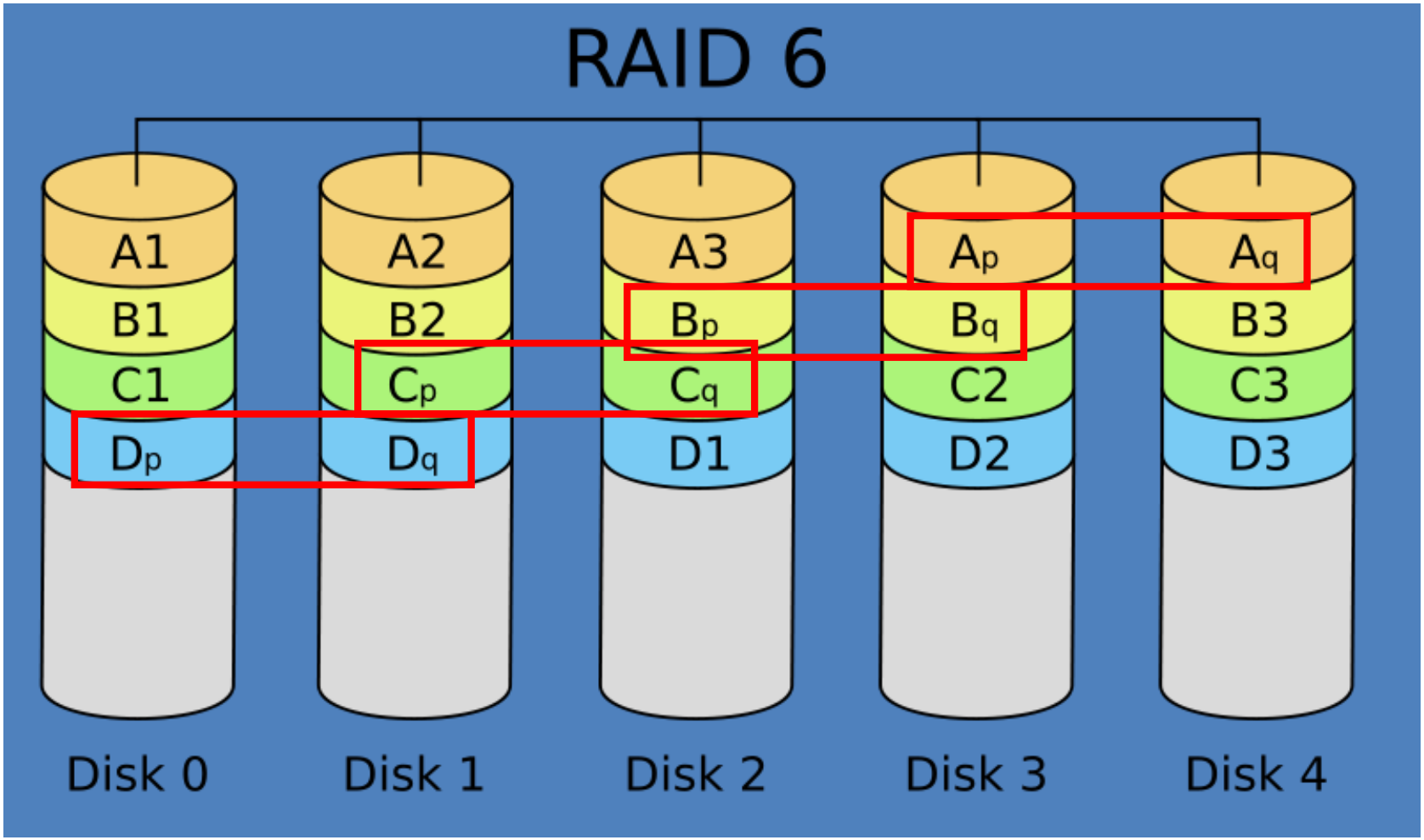
RAID 5



Disk striping with Dedicated Parity Drive



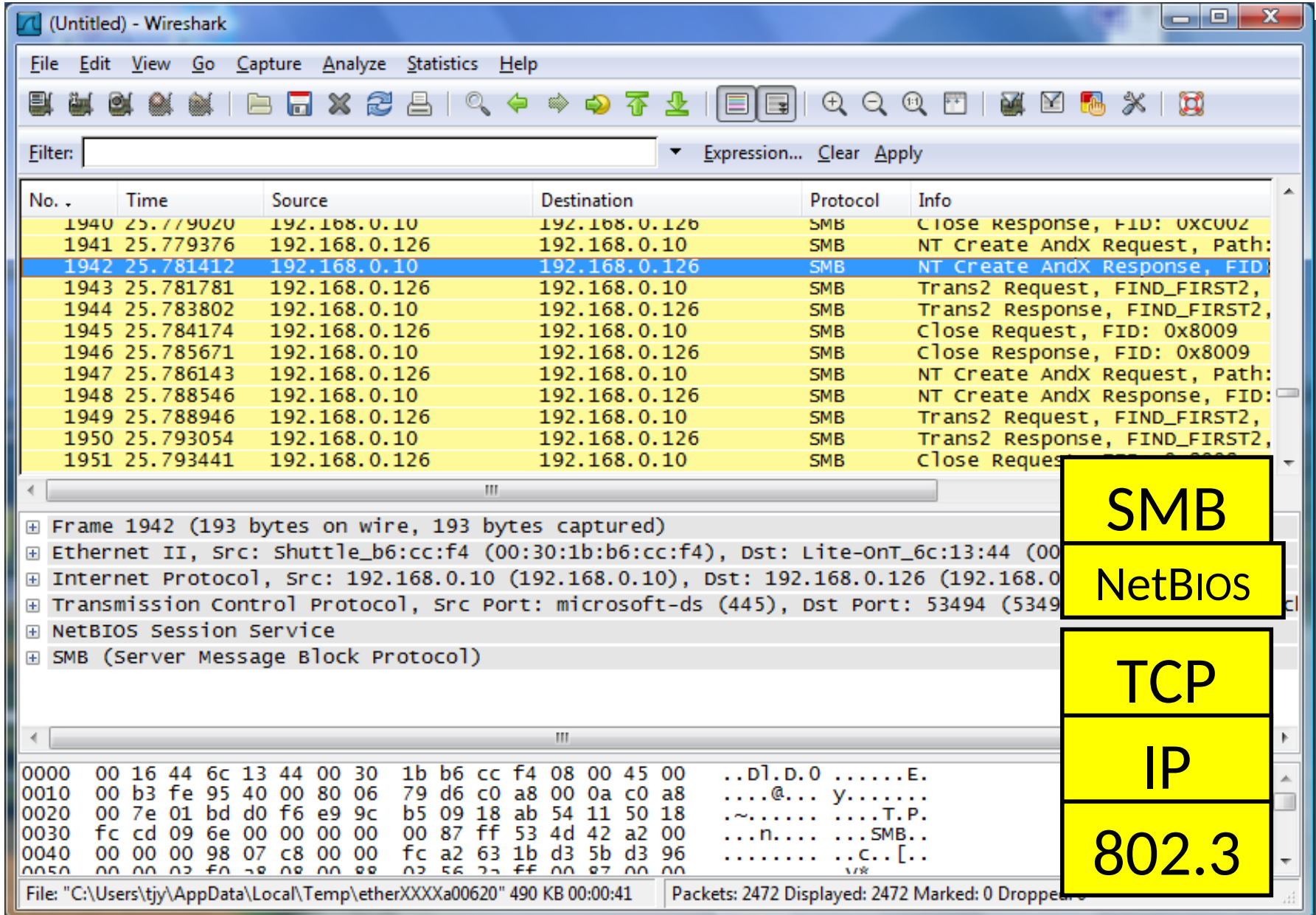
Disk striping with Distributed Parity Data

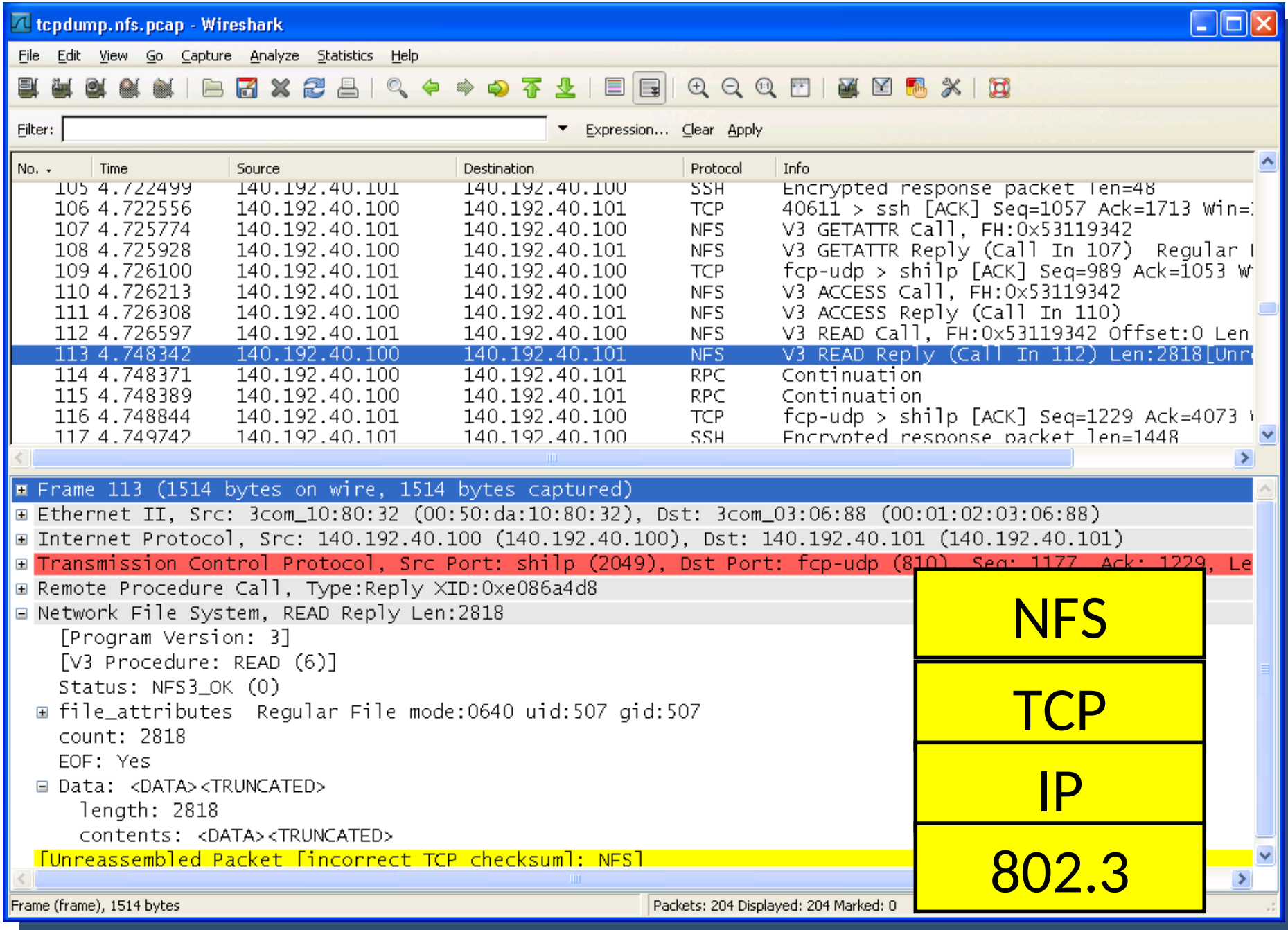


Striping (parity) data is duplicate.

Network Attached Storage (NAS)

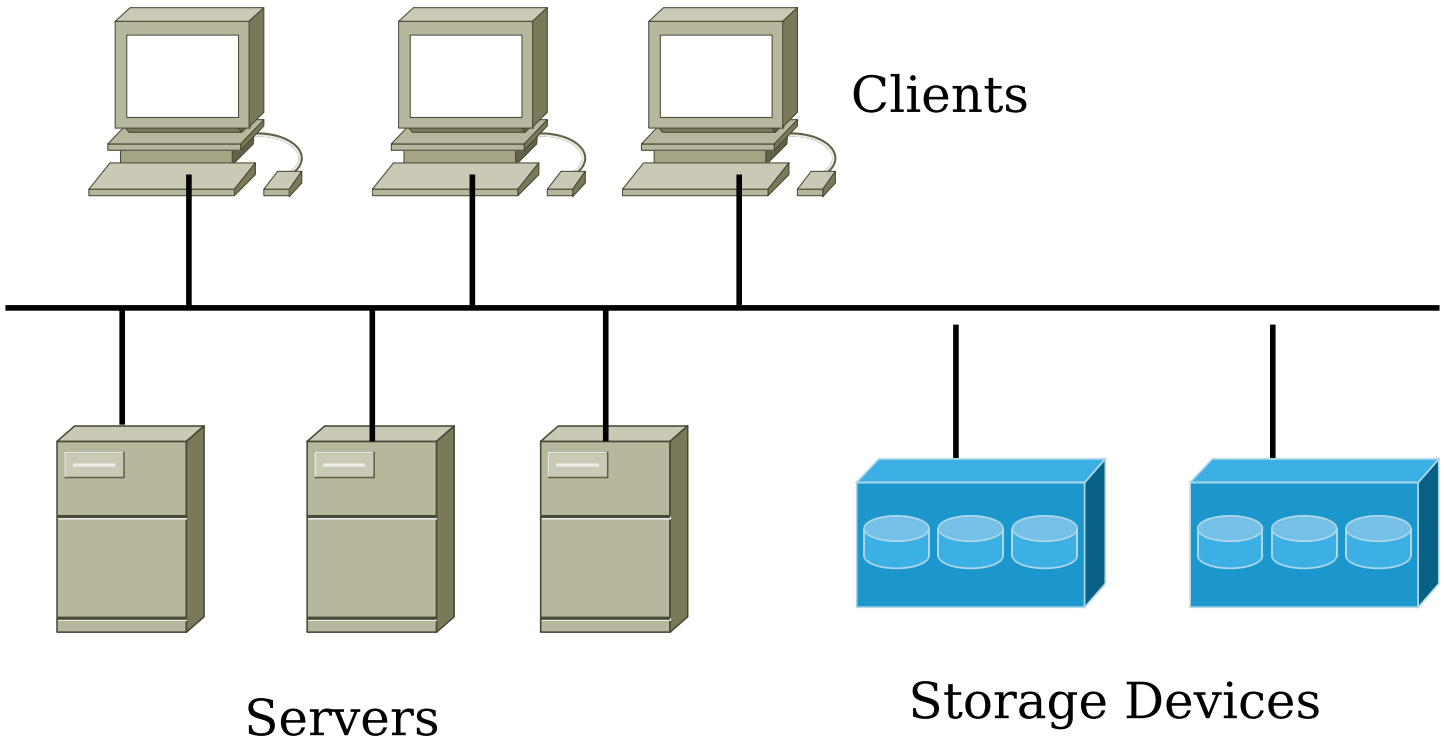
- **NAS is adedicated storage device, and it operates in a client/server mode.**
- **NAS is connected to the file server via LAN.**
- **Protocol: NFS (or CIFS) over an IP Network**
  - Network File System (NFS) – UNIX/Linux
  - Common Internet File System (CIFS) – Windows Remote file system (drives) mounted on the local system (drives)
    - evolved from Microsoft NetBIOS, NetBIOS over TCP/IP (NBT), and Server Message Block (SMB)
  - SAMBA: SMB on Linux (Making Linux a Windows File Server)
- Advantage: no distance limitation
- Disadvantage: Speed and Latency
- Weakness: Security


(Untitled) - Wireshark
File Edit View Go Capture Analyze Statistics Help
Filter: Expression... Clear Apply
No. Time Source Destination Protocol Info
1940 25.779020 192.168.0.10 192.168.0.126 SMB Close Response, FID: 0xc002
1941 25.779376 192.168.0.126 192.168.0.10 SMB NT Create AndX Request, Path:
1942 25.781412 192.168.0.10 192.168.0.126 SMB NT Create AndX Response, FID:
1943 25.781781 192.168.0.126 192.168.0.10 SMB Trans2 Request, FIND_FIRST2,
1944 25.783802 192.168.0.10 192.168.0.126 SMB Trans2 Response, FIND_FIRST2,
1945 25.784174 192.168.0.126 192.168.0.10 SMB Close Request, FID: 0x8009
1946 25.785671 192.168.0.10 192.168.0.126 SMB Close Response, FID: 0x8009
1947 25.786143 192.168.0.126 192.168.0.10 SMB NT Create AndX Request, Path:
1948 25.788546 192.168.0.10 192.168.0.126 SMB NT Create AndX Response, FID:
1949 25.788946 192.168.0.126 192.168.0.10 SMB Trans2 Request, FIND_FIRST2,
1950 25.793054 192.168.0.10 192.168.0.126 SMB Trans2 Response, FIND_FIRST2,
1951 25.793441 192.168.0.126 192.168.0.10 SMB Close Reques
Frame 1942 (193 bytes on wire, 193 bytes captured)
Ethernet II, Src: Shuttle_b6:cc:f4 (00:30:1b:b6:cc:f4), Dst: Lite-OnT_6c:13:44 (00
Internet Protocol, Src: 192.168.0.10 (192.168.0.10), Dst: 192.168.0.126 (192.168.0
Transmission Control Protocol, Src Port: microsoft-ds (445), Dst Port: 53494 (5349
NetBIOS Session Service
SMB (Server Message Block Protocol)
0000 00 16 44 6c 13 44 00 30 1b b6 cc f4 08 00 45 00 ..Dl.D.0 ......E.
0010 00 b3 fe 95 40 00 80 06 79 d6 c0 a8 00 0a c0 a8 ....@... y.......
0020 00 7e 01 bd d0 f6 e9 9c b5 09 18 ab 54 11 50 18 .~...... ....T.P.
0030 fc cd 09 6e 00 00 00 00 00 87 ff 53 4d 42 a2 00 ...n.... ...SMB..
0040 00 00 00 98 07 c8 00 00 fc a2 63 1b d3 5b d3 96 ....c..[..
File: "C:\Users\tjy\AppData\Local\Temp\etherXXXXa00620" 490 KB 00:00:41
Packets: 2472 Displayed: 2472 Marked: 0
SMB
NetBIOS
TCP
IP
802.3


tcpdump.nfs.pcap - Wireshark
File Edit View Go Capture Analyze Statistics Help
Filter: Expression... Clear Apply
No. Time Source Destination Protocol Info
105 4.722499 140.192.40.101 140.192.40.100 SSH Encrypted response packet len=48
106 4.722556 140.192.40.100 140.192.40.101 TCP 40611 > ssh [ACK] Seq=1057 Ack=1713 Win=
107 4.725774 140.192.40.101 140.192.40.100 NFS V3 GETATTR Call, FH:0x53119342
108 4.725928 140.192.40.100 140.192.40.101 NFS V3 GETATTR Reply (Call In 107) Regular
109 4.726100 140.192.40.101 140.192.40.100 TCP fcp-udp > shilp [ACK] Seq=989 Ack=1053 W
110 4.726213 140.192.40.101 140.192.40.100 NFS V3 ACCESS Call, FH:0x53119342
111 4.726308 140.192.40.100 140.192.40.101 NFS V3 ACCESS Reply (Call In 110)
112 4.726597 140.192.40.101 140.192.40.100 NFS V3 READ Call, FH:0x53119342 Offset:0 Len
113 4.748342 140.192.40.100 140.192.40.101 NFS V3 READ Reply (Call In 112) Len:2818[Unr
114 4.748371 140.192.40.100 140.192.40.101 RPC Continuation
115 4.748389 140.192.40.100 140.192.40.101 RPC Continuation
116 4.748844 140.192.40.101 140.192.40.100 TCP fcp-udp > shilp [ACK] Seq=1229 Ack=4073
117 4.749742 140.192.40.101 140.192.40.100 SSH Encrypted response packet len=1448
Frame 113 (1514 bytes on wire, 1514 bytes captured)
Ethernet II, Src: 3com_10:80:32 (00:50:da:10:80:32), Dst: 3com_03:06:88 (00:01:02:03:06:88)
Internet Protocol, Src: 140.192.40.100 (140.192.40.100), Dst: 140.192.40.101 (140.192.40.101)
Transmission Control Protocol, Src Port: shilp (2049), Dst Port: fcp-udp (810), Seq: 1177, Ack: 1229, Le
Remote Procedure Call, Type:Reply XID:0xe086a4d8
Network File System, READ Reply Len:2818
[Program Version: 3]
[V3 Procedure: READ (6)]
Status: NFS3_OK (0)
file_attributes Regular File mode:0640 uid:507 gid:507
count: 2818
EOF: Yes
Data: <DATA><TRUNCATED>
length: 2818
contents: <DATA><TRUNCATED>
[Unreassembled Packet [incorrect TCP checksum]: NFS]
Frame (frame), 1514 bytes
Packets: 204 Displayed: 204 Marked: 0
NFS
TCP
IP
802.3

Network Attached Storage (NAS)

- Specialized storage device or group of storage devices providing centralized fault-tolerant data storage for a network



Storage Area Network (SAN)

- A Storage Area Network (SAN) is a specialized, dedicated high speed network joining servers and storage, including disks, disk arrays, tapes, etc.
- Storage (data store) is separated from the processors (and separated processing).
- High capacity, high availability, high scalability, ease of configuration, ease of reconfiguration.
- **Fiber Channel** is the de facto SAN networking architecture, although other network standards could be used.

SAN Benefits

- Storage consolidation
- Data sharing
- Non-disruptive scalability for growth
- Improved backup and recovery
- Tape pooling
- LAN-free and server-free data movement
- High performance
- High availability server clustering
- Data integrity
- Disaster tolerance
- Ease of data migration
- Cost-effectives (total cost of ownership)

NAS vs. SAN ?

- Traditionally:
  - NAS is used for low-volume access to a large amount of storage by many users
  - SAN is the solution for terabytes (10^{12}) of storage and multiple, simultaneous access to files, such as streaming audio/video.
- The lines are becoming blurred between the two technologies now, and while the SAN-versus-NAS debate continues, the fact is that both technologies complement each another.

Fibre Channel

- Fiber Channel is well established in the open systems environment as the underlining architecture of the SAN.
- Fibre Channel is structured with independent layers, as are other networking protocols. There are five layers, where 0 is the lowest layer. The physical layers are 0 to 2. These layers carry the physical attributes of the network and transport the data created by the higher level protocols, such as SCSI, TCP/IP, or FICON.

FC Standard – ANSI T11

- T11 (technical committee) has been producing interface standards for high-performance and mass storage applications since the 1970's.
  - http://www.t11.org/index.htm
- Designed to transport multiple protocols, such as HIPPI, IPI, SCSI, IP, Ethernet, etc.
- Full duplex medium
- Channels are established between the originator and the responder.
- Transfer rate from 100MB/s to Gigabits/s
- Distance >10 km (single mode fiber)
- Multi-layer stack functions (not mapped to the OSI model)

Plan

Introduction

NFS

SAMBA

NFS : Network File System

Présenté par SUN en 1985 pour permettre à ces stations sans disque d'accéder à un système de gestion de fichiers distants (RFC 1904).
Utilise les appels de procédures distantes Sun-RPC (qui sont issues des travaux sur NFS)

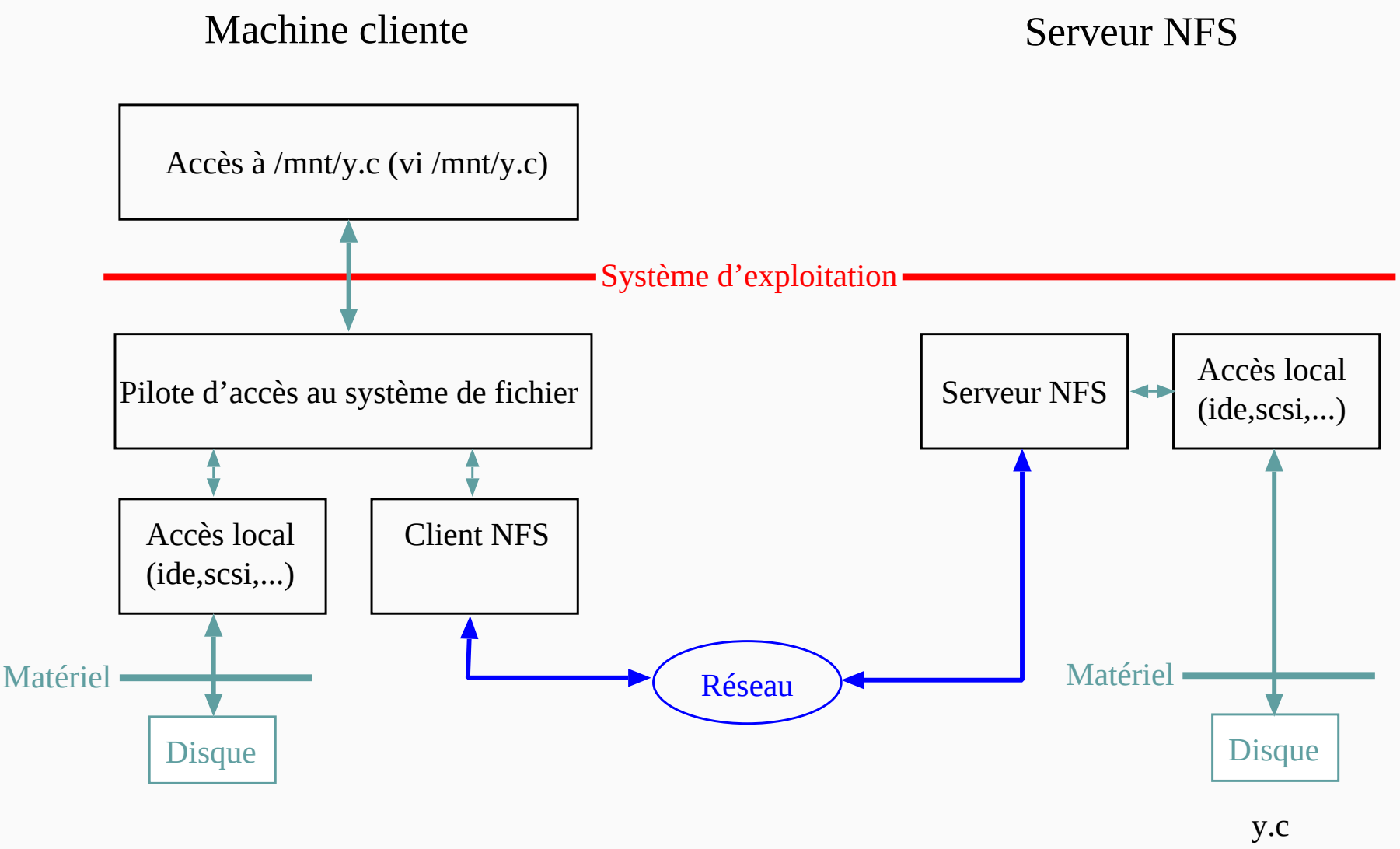
- à priori, les clients et serveur NFS devraient être des processus utilisateur s'exécutant au-dessus de RPC/XDR/UDP/IP.
- en fait, le client et le serveur NFS s'exécutent dans le noyau
  - le client pour rendre transparent l'accès à un fichier via NFS
  - le serveur pour des raisons d'efficacité

NFS : Network File System

Présenté par SUN en 1985 pour permettre à ces stations sans disque d'accéder à un système de gestion de fichiers distants (RFC 1904).
Utilise les appels de procédures distantes Sun-RPC (qui sont issues des travaux sur NFS)

- à priori, les clients et serveur NFS devraient être des processus utilisateur s'exécutant au-dessus de RPC/XDR/UDP/IP.
- en fait, le client et le serveur NFS s'exécutent dans le noyau
  - le client pour rendre transparent l'accès à un fichier via NFS
  - le serveur pour des raisons d'efficacité

Principe de fonctionnement



NFS et les RPCs

NFS repose sur les RPC (Remote Procedure Calls)

- Utilisation du portmapper (programme portmap de Linux)
- Portmapper = conversion des n° de prog RPC en numéro de port

Déroulement d'une RPC :

- Serveur RPC :
  - → Indique à portmap le port qu'il utilise et les numéros de programme RPC qu'il gère
- Envoi d'une requête RPC par un client :
  - → Il contacte portmap du serveur pour connaître le numéro de port du programme souhaité
  - → Il envoie les données au port correspondant

Un serveur sans état

Dans un accès à un système de fichier local :

- Les accès reposent sur un pointeur de fichier maintenu au niveau du système d'exploitation

NFS est basé sur une connexion réseau :

$\Rightarrow$ Probabilité d'une panne importante

Solution :

- Le serveur NFS ne conserve aucune information sur les accès/opérations effectuées (fichiers ouverts, accès précédents au fichier,...)
- Le système d'exploitation du client NFS se charge de maintenir les informations concernant les fichiers

Intérêts :

- Simplifie le redémarrage du serveur en cas de crash.

Sécurité

Principe d'authentification :

- $(\texttt{uid}, \texttt{gid})_{\texttt{local}}$ "mappé" sur $(\texttt{uid}, \texttt{gid})_{\texttt{distant}}$
  $\rightarrow$ équivalence entre les droits locaux et les droits distants
- Problème pour `root` :
  - Quels droits possède le *root* d'une machine cliente sur les fichiers exportés par un serveur NFS?
  - par défaut `root` (coté client) correspond a l'utilisateur `nobody` (coté serveur) pour des raisons de sécurité (sinon il faut mettre l'option `no_root_squash` dans `/etc/exports`)

Règle de non transitivité :

- Si A exporte `/home` à B; Si B monte `A:/home` dans `/home2` et exporte `/home2` à C alors C n'aura pas accès au `/home` de A

Liens symboliques :

- Les liens symboliques relatifs sont interprétés par rapport au système de fichier du client.

NFS en pratique (1/2)

Démons importants utilisés par NFS :

- **portmap.** Gestion des connexions des applications utilisant le mécanisme de RPC.
- **nfsd.** Authentification + Création, recherche, lecture et écriture de fichiers
- **mountd.** Montage des systèmes exportés (mount et umount)
- **statd.** Surveillance des nœuds du réseau (redémarrages...)
- **lockd.** Section critique (lock les fichiers utilisés)

NFS en pratique (2/2)

coté client. le fichier `/etc/fstab` doit contenir le chemin vers le point de montage et le chemin sur le serveur NFS.

```
192.168.0.1:/home /nfs nfs defaults,noauto
0 0
```

coté serveur. le fichier `/etc/exports` contient le chemin vers les dossiers à exporter ainsi que la liste les machines autorisées à y accéder. `/home`

```
192.168.1.0/255.255.255.0(rw,no_root_squash)
```

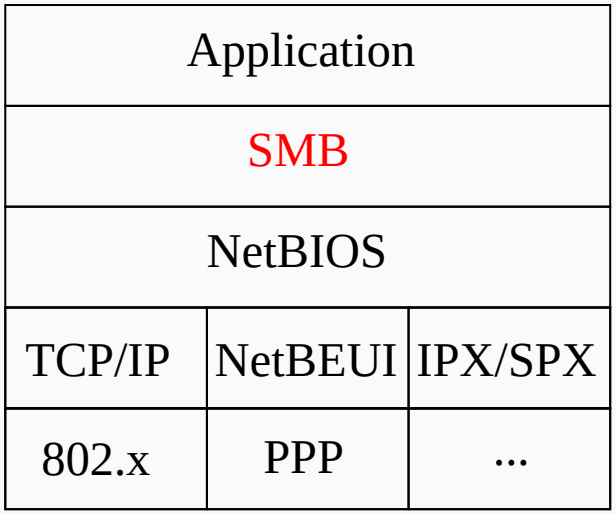
Après chaque modification de `/etc/exports` il est nécessaire :

- soit d'exécuter `exportfs` pour transmettre les modifications au serveur nfs
- soit de relancer le serveur NFS

Plan

SMB : Server Message Block

- Protocole de Microsoft et Intel permettant le partage de ressources (disques, imprimantes, . . .) à travers un réseau (1987)
- SMB est prévu pour être utilisé au dessus de l'interface NetBIOS
  - Utilisation des noms NetBIOS (15 caractères + 1 pour le type)
  - Utilisation du mécanisme de datagramme de NetBIOS par *broadcast* comme service de nommage (nom $\rightarrow$ MAC, pas d'adresse de niveau 3)



SMB (1/2)

- Chaque machine client ou serveur possède un nom sur 15 caractères
- SMB ajoute un 16ème caractère pour distinguer les serveurs de fichiers, les clients, les imprimantes, ...
- Notion de domaine
  - un ensemble d'utilisateurs (avec nom et mot de passe) et de serveurs (avec des droits d'accès)
  - un *primary domain server* contient la base de données des utilisateurs et de leur mot de passe
- Un serveur une ou plusieurs ressources
  - fichiers, imprimantes, ...
  - à chaque triplet (domaine, serveur, ressource) correspond un nom unique : `\\serveur\ressource`

SMB (2/2)

Deux niveaux de protection :

- au niveau de chaque utilisateur : basé sur le nom des utilisateurs, permet de gérer l'accès aux ressources voire aux éléments d'une ressource
- au niveau de chaque ressource : un mot de passe commun à tous les utilisateurs est associé à une ressource pour y autoriser l'accès

Résolution de noms : 4 méthodes utilisées

broadcast. résolution par diffusion d'une requête dans le réseau

lmhost. résolution en utilisant des associations prédéfinies entre noms NetBIOS et IP

host. utilisation de DNS

wins. utilisation d'un serveur WINS (*Windows Internet Name Server*). À chaque machine est associé un serveur WINS à qui elle envoie ses requêtes et auprès duquel elle s'enregistre.

SAMBA (1/2)

Samba : Implémentation de SMB sous UNIX qui permet le partage de ressources entre les mondes UNIX et Windows
Samba permet de :

- partager un disque UNIX pour des machines Windows
- accéder à un disque Windows depuis une machine UNIX
- partager une imprimante UNIX pour des machines Windows
- utiliser une imprimante Windows à partir d'un hôte Linux.

Le serveur Samba sur la machine Unix émule un domaine SMB

SAMBA (2/2)

Serveur Samba :

- configuration via le fichier `/etc/smb.conf`
- travail partagé par deux démons :
  - `smbd`. pour le service “serveur”
  - `nmbd`. pour le service résolution des noms NetBIOS

Client :

smbpasswd. permet de changer le mot de passe d’un utilisateur SMB

smbclient. permet d’interroger un serveur Samba depuis UNIX `smbclient //host/ressource` permet l’accès à la ressource

Possibilité de monter une partition Windows distante à l’aide de Samba $\rightarrow$ utiliser le système de fichier `smbfs`

Exemple : (extrait du fichier `/etc/fstab`)

```
//serveur/ressource /commun smbfs defaults 0 0
```