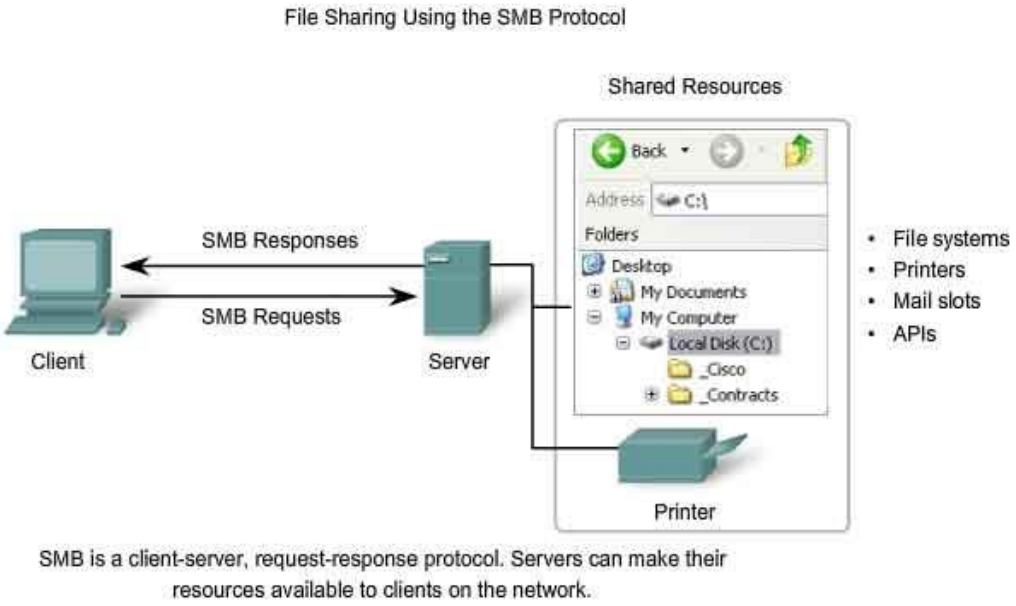
SMB NETWORK FILESYSTEM : In-Kernel SMB Server

Namjae Jeon
namjae.jeon@protocolfreedom.org

What is CIFS/SMB ?

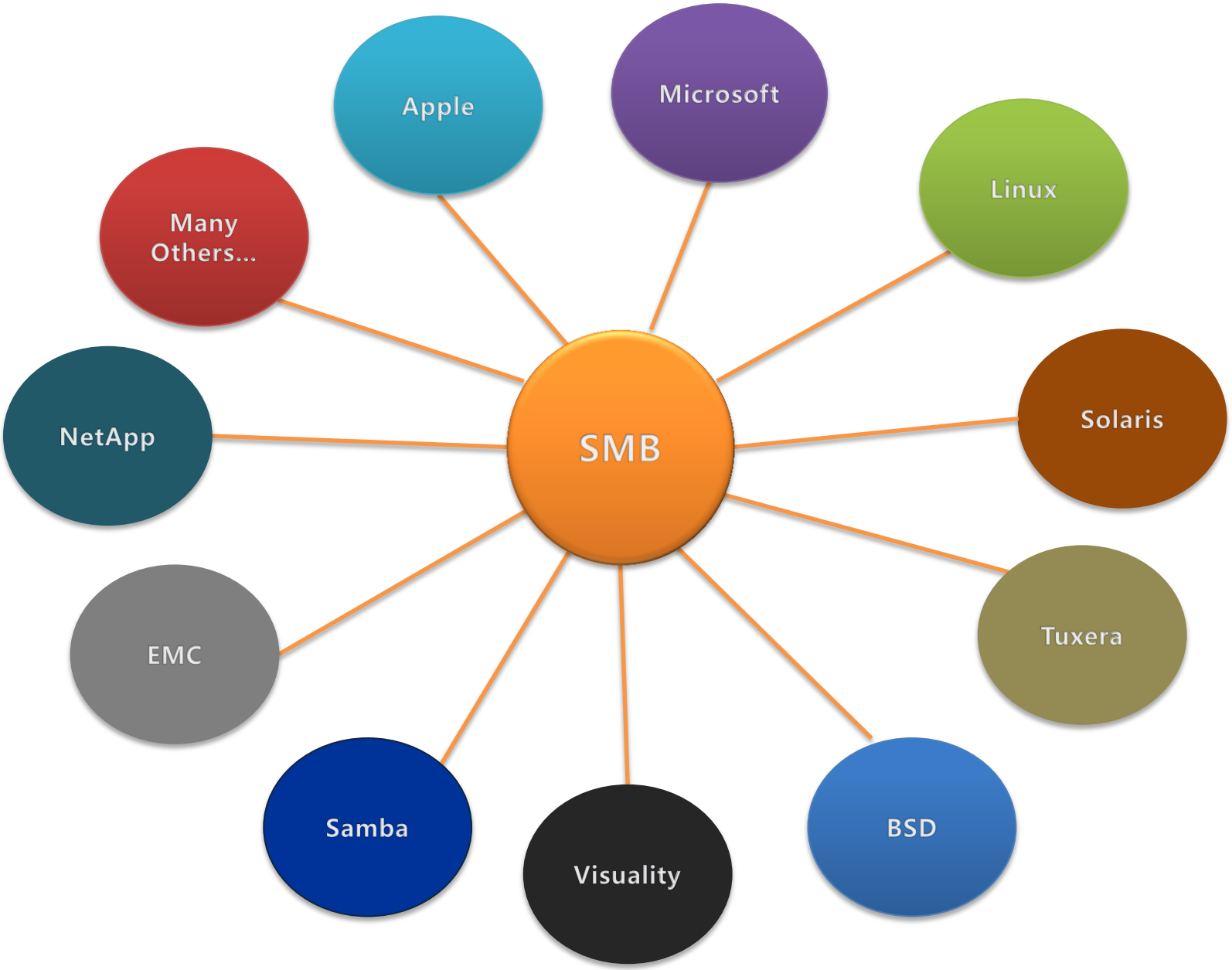
- SMB (Server Message Block).
- CIFS (Common Internet File System).
- SMB/CIFS is a protocol for sharing files, printers bet ween computers.
- You are probably using it every day.
  - Windows Network→Host→Shared Fold er→File
- “File Sharing”:
  - Expose your own files
  - View other’s files



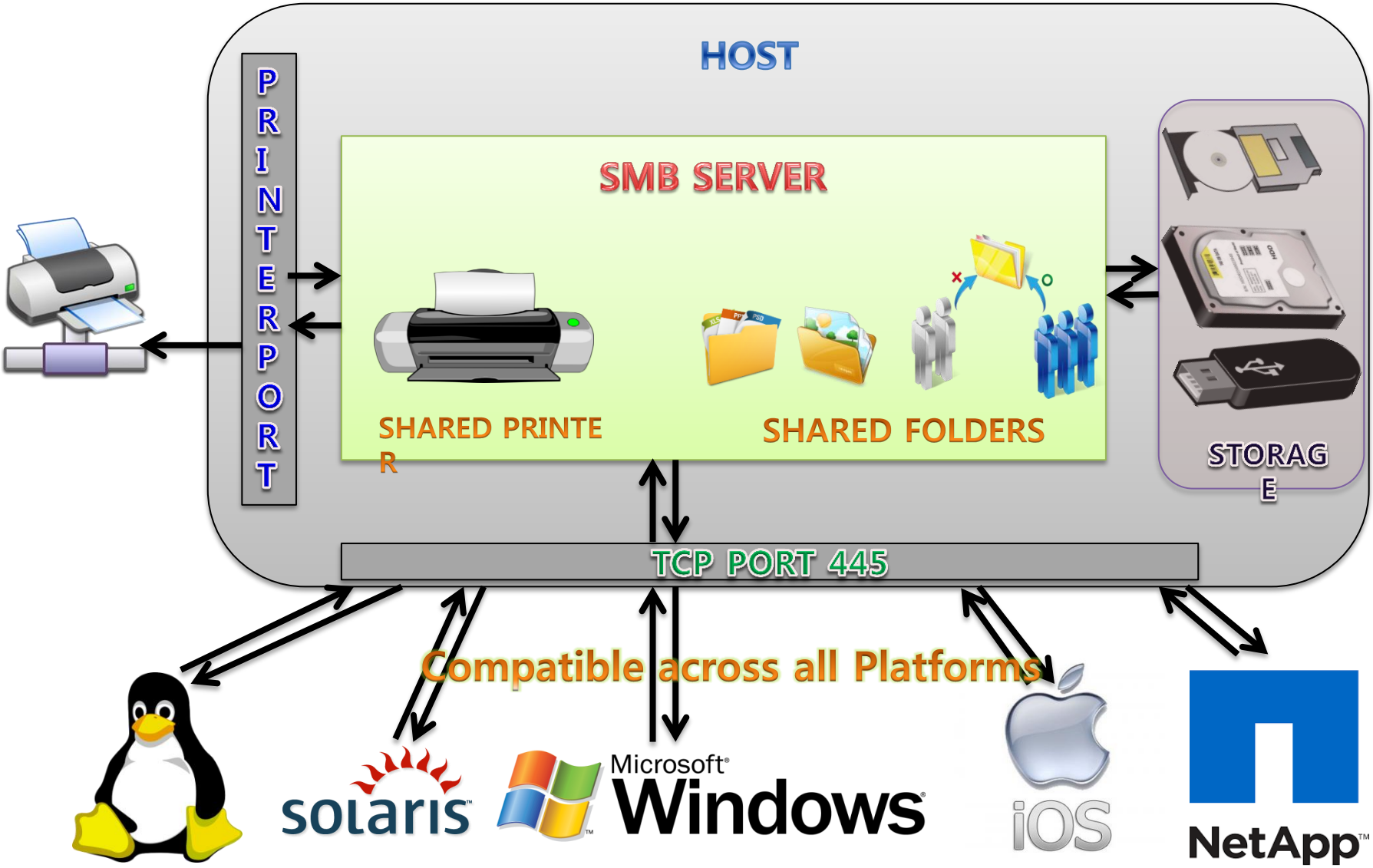
Brief history of CIFS/SMB

- SMB-1982 Designed by IBM (*Barry Feigenbaum)*
  - PC-DOS – 1984
  - LAN Manager – 1988
  - Unix and other operating systems (part of the OS or as a suite like Samba)
- CIFS Adopted by Microsoft as of Win95
  - Common Internet File System
  - Windows NT 4.0 – 1996
  - IETF draft – Common Internet File System – 1997
  - SNIA Technical Specification – 1999
- SMB (returned to the original name) - 2000
  - Windows 2000 Extensions – 2000
  - Extensions for other implementations of SMB
- SMB 2.0 (or SMB2) – 2008 Windows Vista
- SMB 2.1 (or SMB2.1) – 2010 Windows 7
- SMB 3.0 (or SMB3) – 2012 Windows 8
- SMB 3.1.1 – 2015 Windows 10

SMB implementers



CIFS/SMB universality and connectivity



Samba project

- An opensource project which impleme nts SMB protocol.
- Samba was originally developed by And rew Tridgell in 1992.
- The name "Samba" was derived by running the Unix command grep through the system dictionary (i.e. grep -i '^s.*m.*b' /usr/share/dict/words)
- Samba has been developed by Samba t eam who is consist of about 40 people.
  - File sharing, print service support.
  - SMB 1, 2, 3 version support and still developing on project.
- Samba Team have decided to adopt GPLv3 since 3.0.37 version(2007)

Andrew Tridgell

Samba Team

SMB Solutions of other company

- Apple : switched to its own SMBX solution after samba adopted GPLv3
- NetApp, EMC : user space SMB engine
- Oracle/Solaris : In Kernel SMB engine
- Visuality system, CIFS NQ : Samba GPLv3 is important marketing point for customers
- Tuxera, TSMB : Open SMB solution in shop

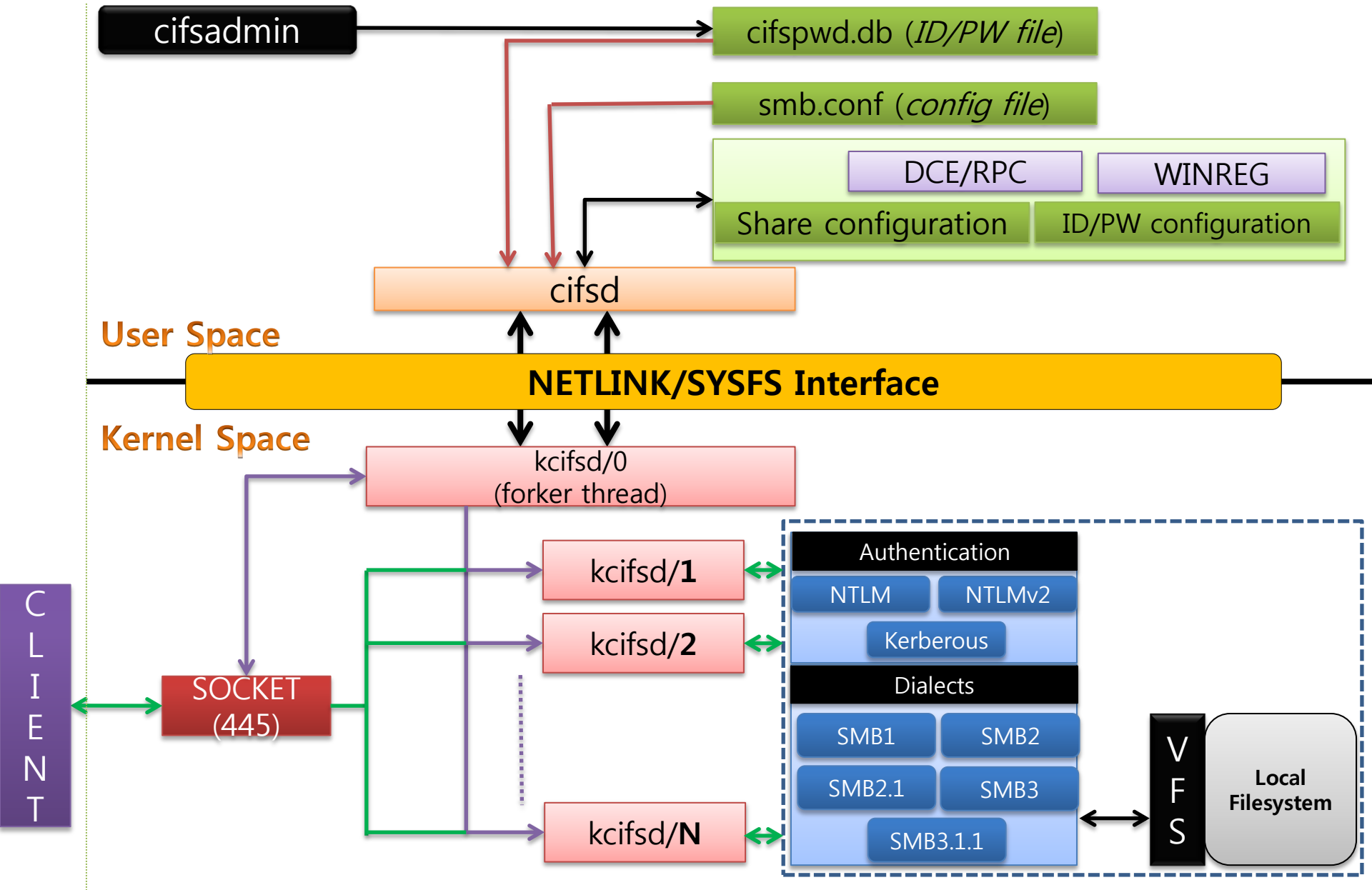
NEW SMB Server Needs

- Anti-GPLv3
- In-Kernel Performance Optimization
- High performance with HW accelerator
- Easy implementation features – RDMA, Multi-C hannel
- better integration with NFS server in kernel

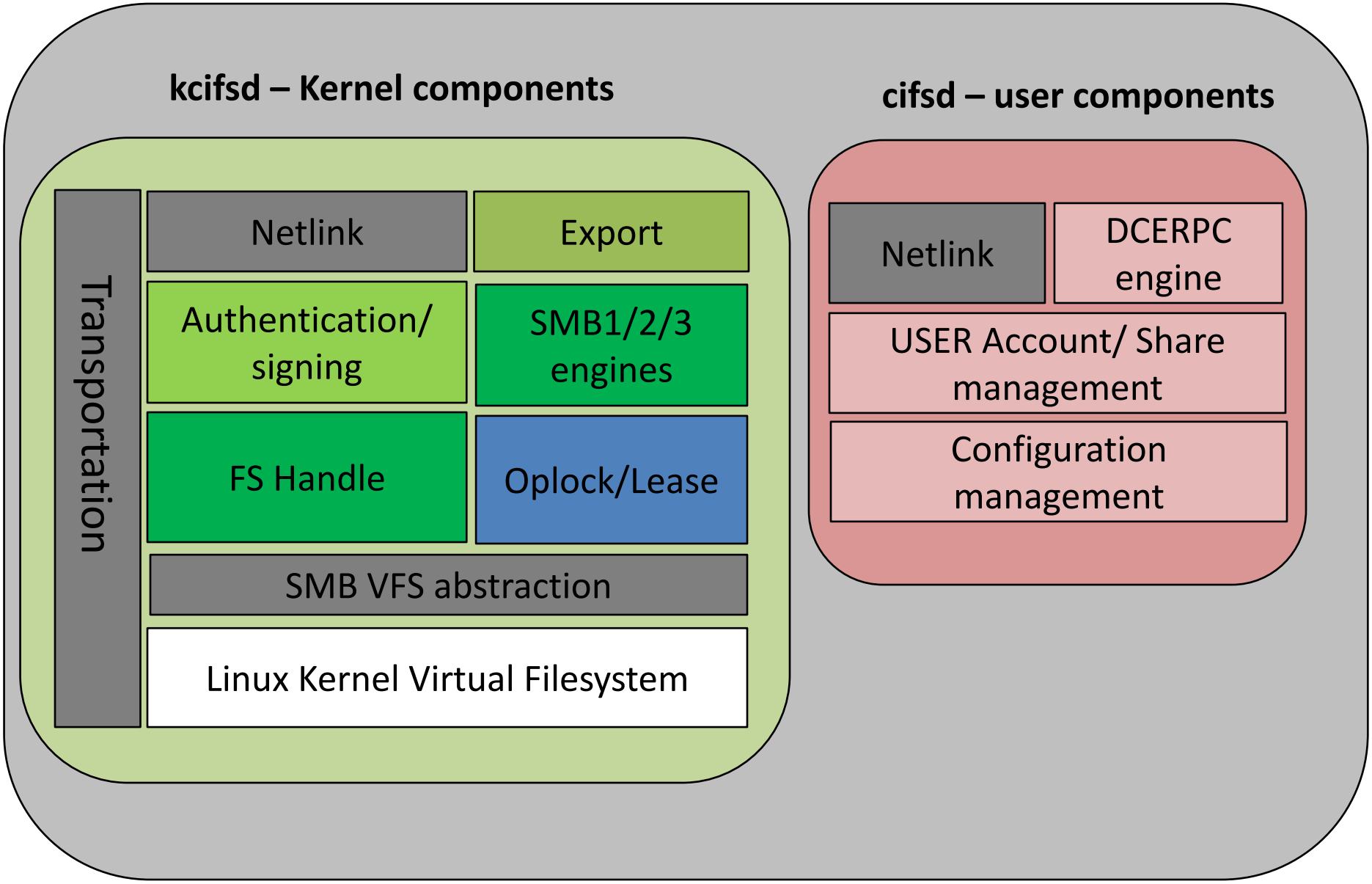
Introduce “cifsd” opensource project

- GPLv2 or later License
- Name “cifsd” to make a pair with “cifs” of linux
- Initial release(2015/12/25) version support SMBv1, SMBv2/2.1, and user level helpers
- Support SMBv3.0 (Win 8), SMB3.11(win 10) now
- Optimized performance in kernel space
- The subset of performance related operations => Kernel
- The other subset which are not really related with performance => Userspace
- Have not yet release stable version

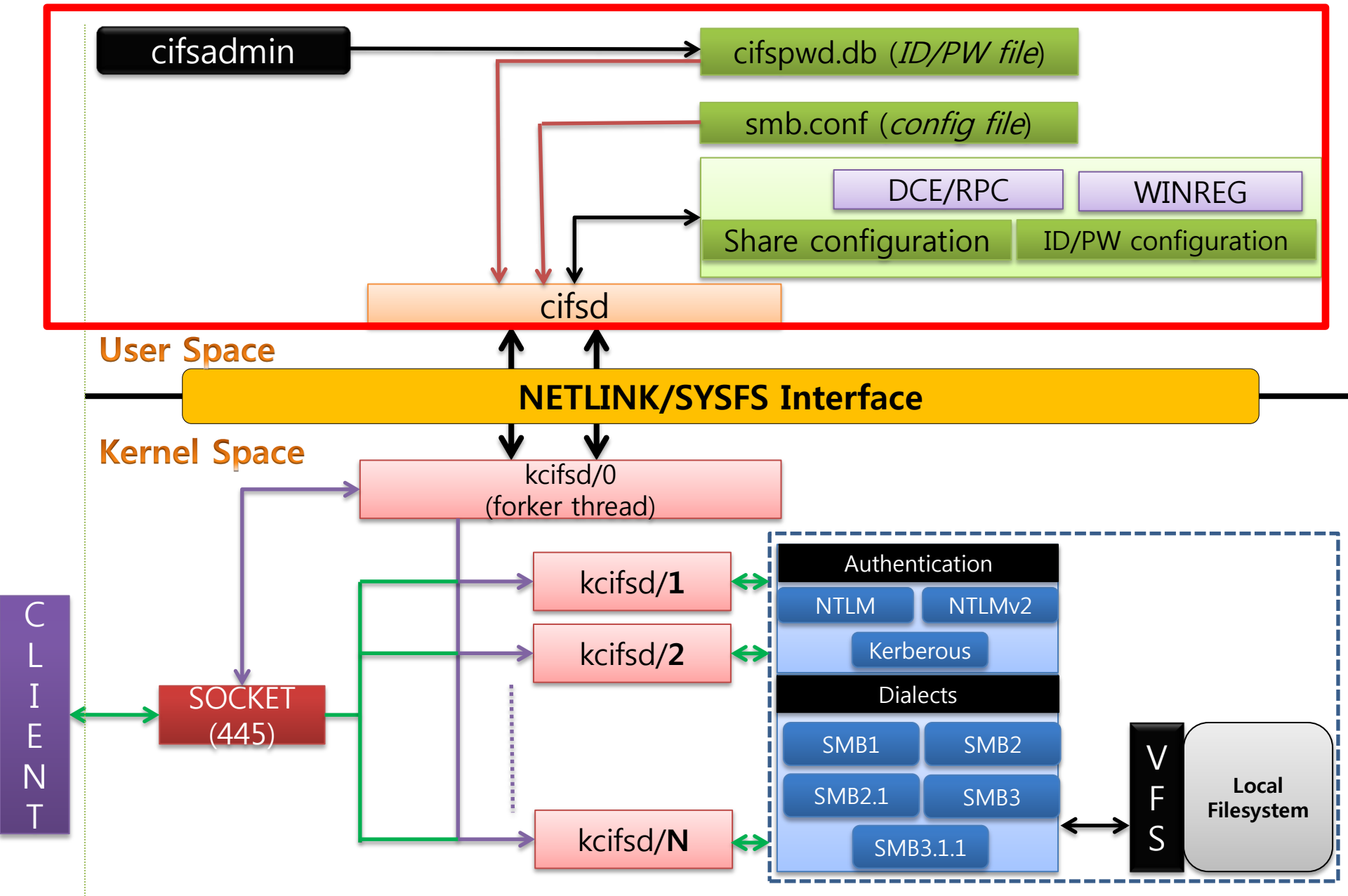
ARCHITECTURE



kcifsd/cifsd components



User level helpers



user level helper (cifsd)

- **cifsd is an userspace daemon to:**
  - transfer user account and password that are registered using cifsadmin (part of utils for user space) to kcifsd
  - allow sharing information parameters that parsed from smb.conf to smb export layer in kernel
  - work non-performance features((DCE/RPC, WINREG, all management operations) with cifsd through netlink or sysfs
- **cifsd provides the options for:**
  - Change smb.conf location(default : /etc/cifs/smb.conf
  - Change cifspwd.pwd location(default : /etc/cifs/cifspwd.db)

user level helper (cifsadmin)

- **cifsadmin is an administrative tool which allows support for:**
  - Management of user accounts
  - Set configuration parameters at runtime(TODO)
- **cifsadmin provides user level support for:**
  - Addition of User accounts
  - Deletion of User accounts
  - Query user account status

user level helper (cifstat)

- **cifsstat tool provides following statistical information**
  - Server uptime in seconds
  - Number of shares
  - Connection type – SMB1/SMB2.0/SMB2.1/SMB3.0 (*Dialects*)
  - Current open files count
  - Outstanding Request
  - Total Requests Served
  - Avg. duration per request
  - Max. duration request
- **cifsstat provides user level support for:**
  - Display of server statistics for all clients connected
  - Display of server statistics for individual selected client

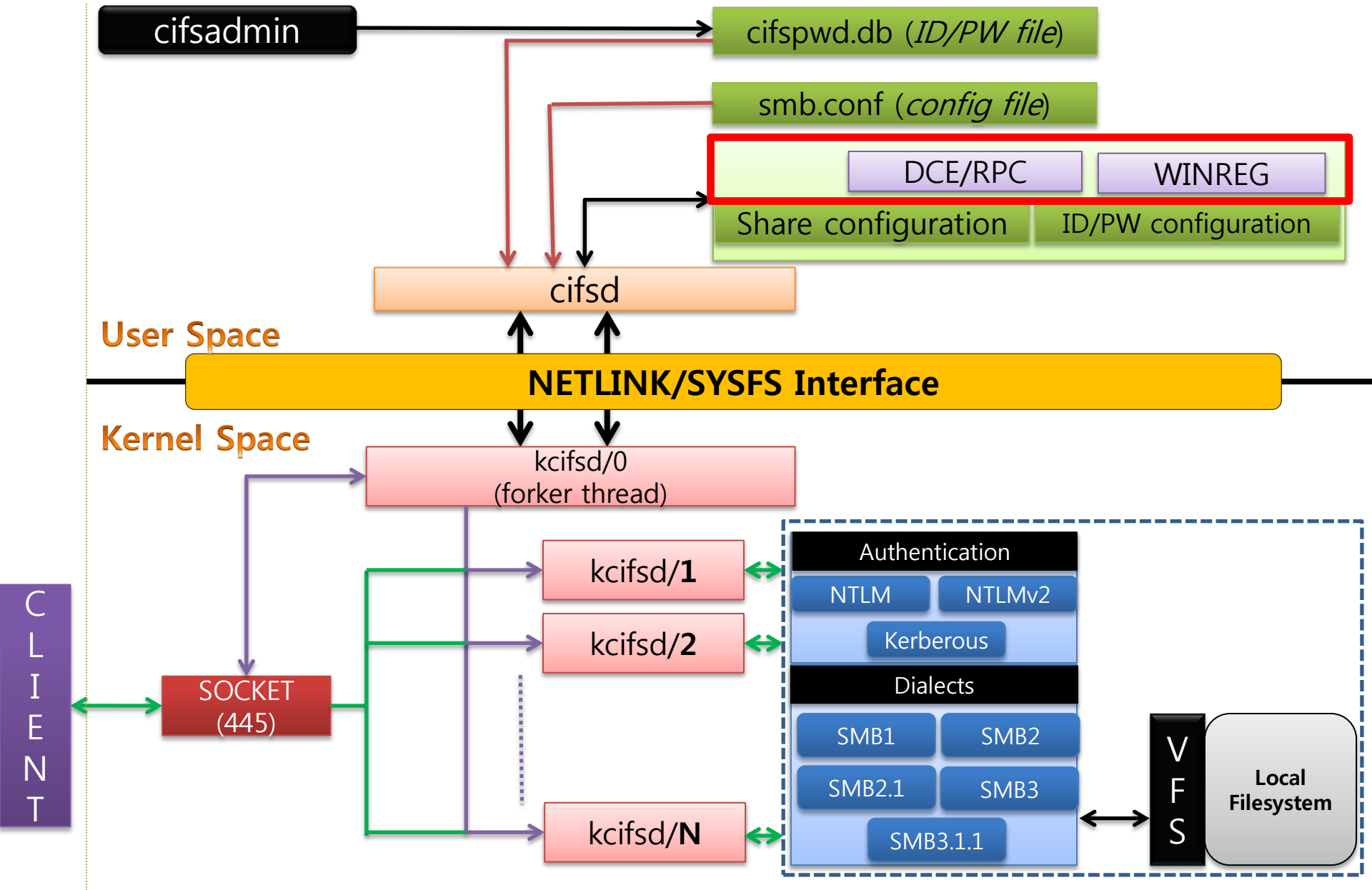
user level helper (smb.conf)

- **cifsd provide configuration file called “smb.conf”**
  - Compatible with command of smb.conf of samba
  - Will implement only what we need in many parameters
  - List up current supporting parameters in cifsd user guide v1.0

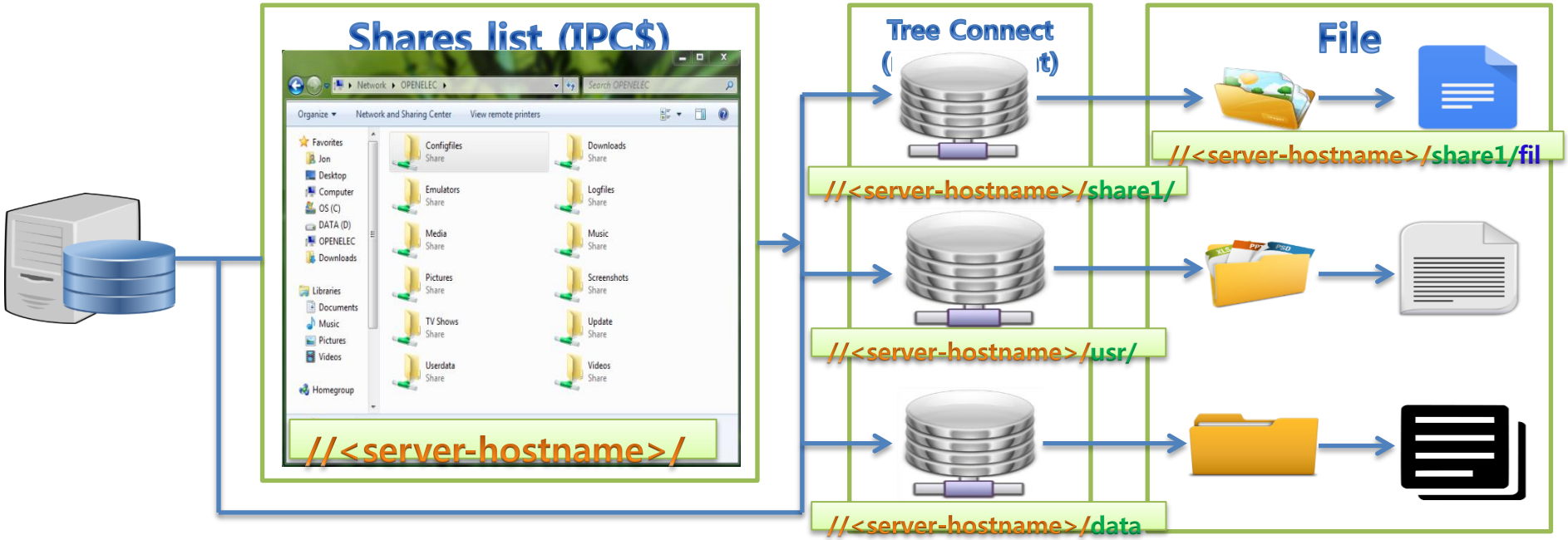
(https://github.com/namjaejeon/Documents/blob/master/CIFSD_User_Guide_v1.0.pdf)

```
[global]
        server string = my home
        server signing = yes
        server min protocol = smb2.1
[shares]
        comment = content server share
        path = /mnt
        guest ok = yes
        writeable = yes
        invalid users = root fred hayul
```

What is DCE/RPC



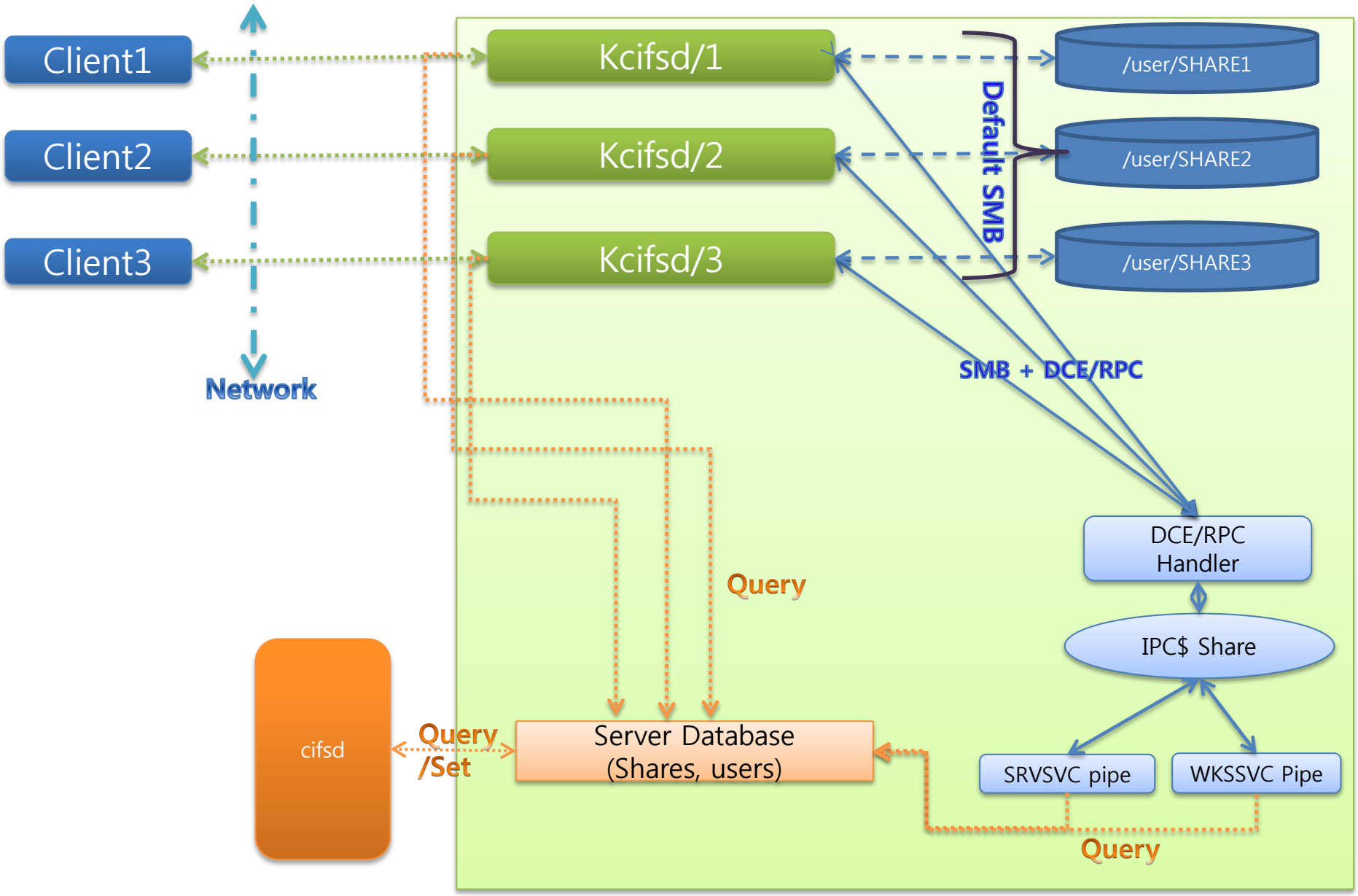
What is DCE/RPC



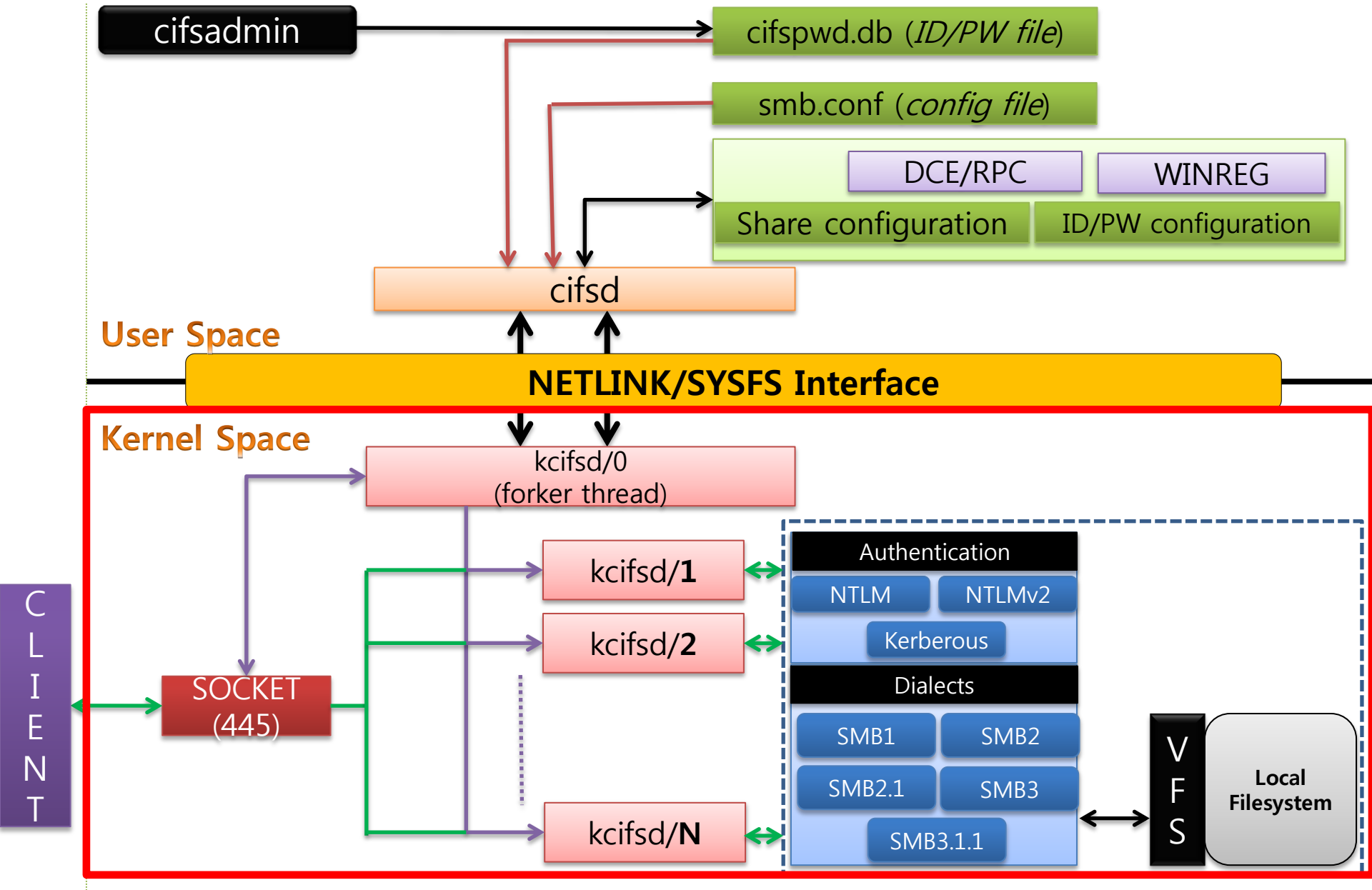
- All SMB over RPC communication happens on hidden $IPC share
  - $IPC : Special Hidden Share
    - Used for non-filesystem operations like account management, printing
    - Gives access to named pipes over the network
  - RPC over SMB uses ENDPOINT called NAMED PIPE e.g.
    - **\\pipe\srvsvc**
    - **\\pipe\wkssvc**
    - \\pipe\winreg
- DCE RPC Protocol Data Units(PDUs) are sent over named pipes, using SMB commands
- **CIFSD implements basic operations on pipes SRVSVC and WKSSVC**

  Basic operations like getting list of shares, server name & domain name.

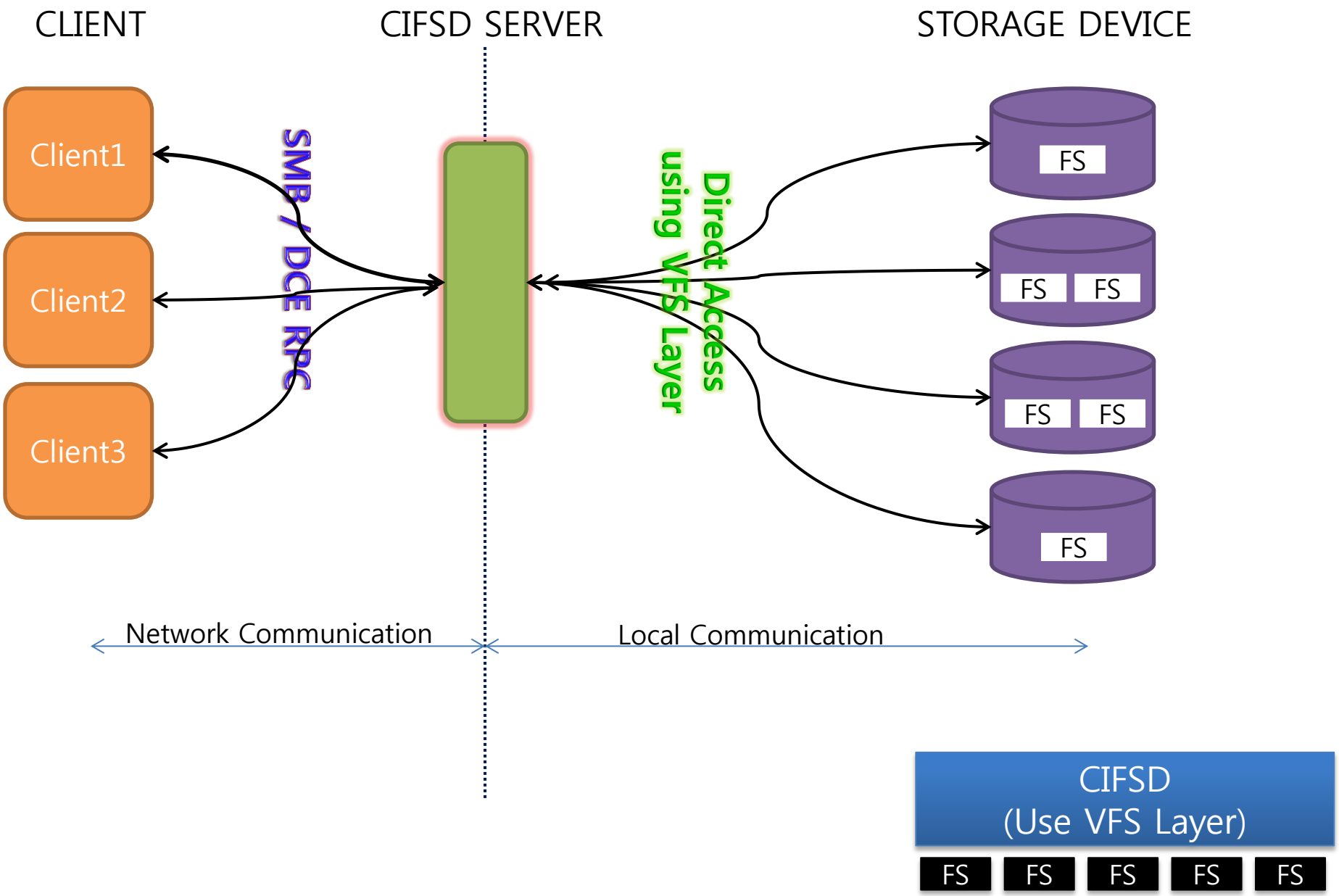
DCE/RPC : Processing



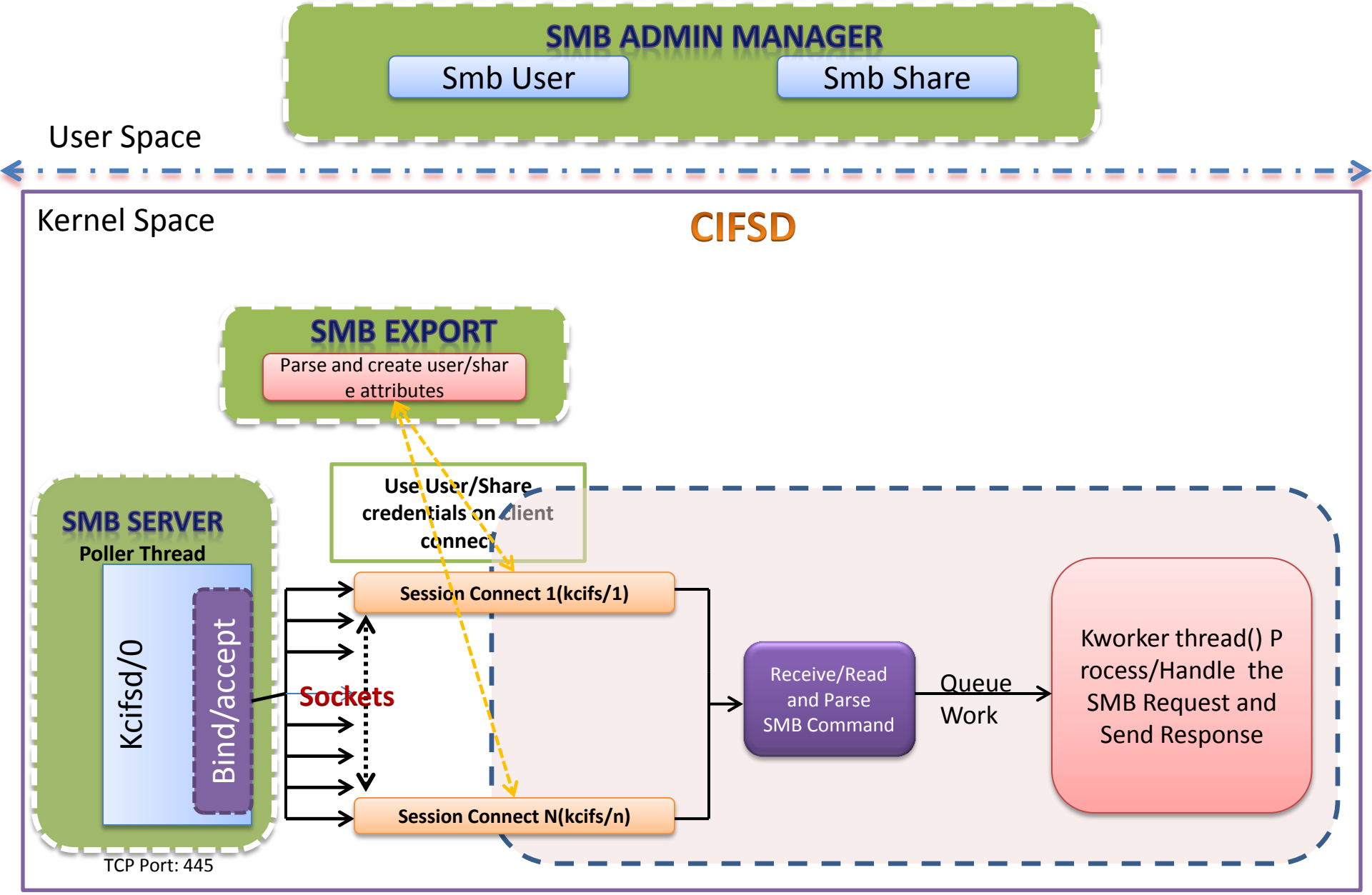
Kernel engines



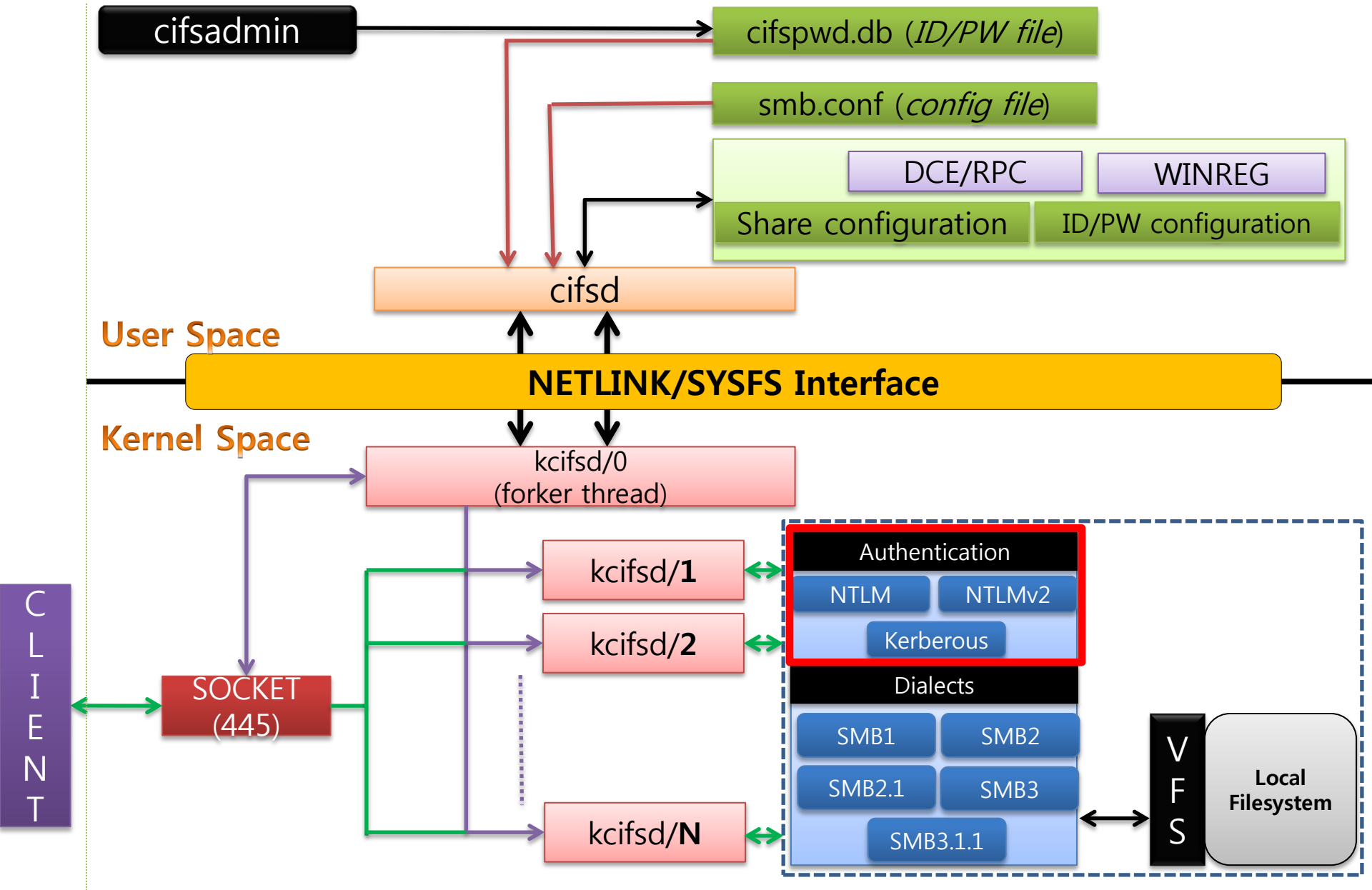
Overview of connection



Cifsd work flow



Authentication

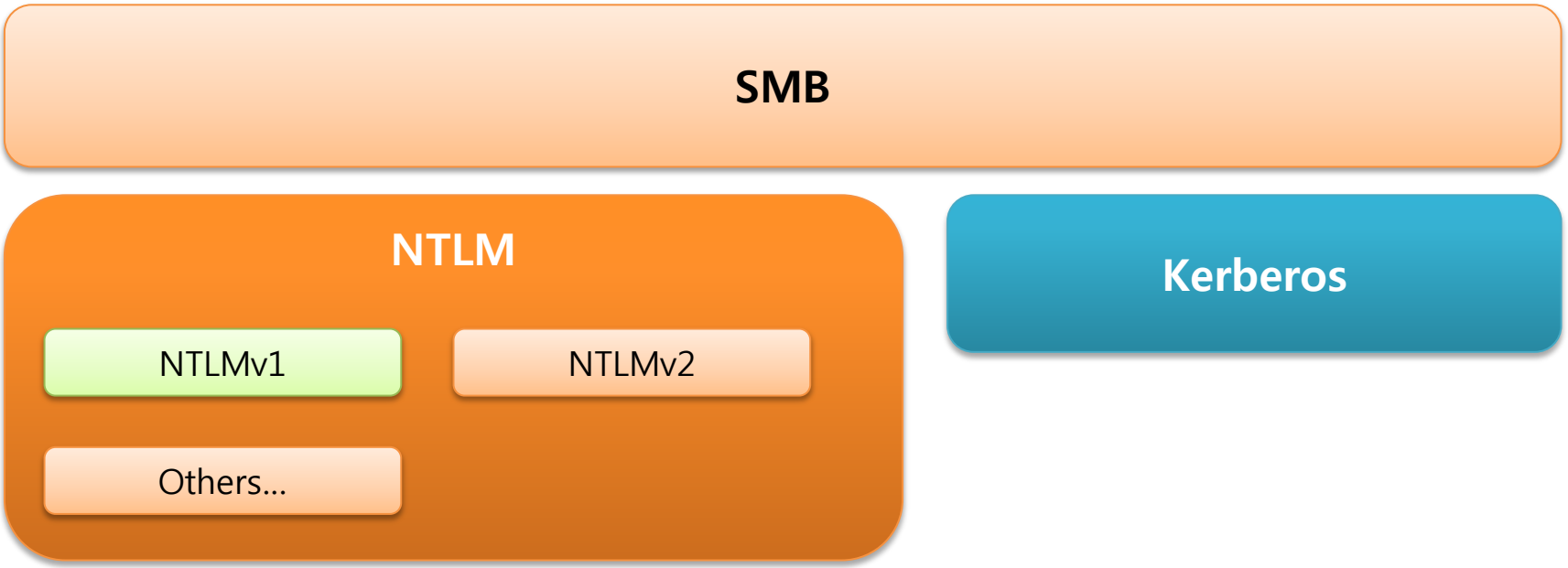


Windows Authentication

- Authentication is the process which allows a server and client to validate each other, thus enabling restricted access to it.
- User provide username and password to access files in shares.
- The password is encrypted before it is sent to the server.

Authentication protocols

- Windows Authentication adopt the following authentication protocols
  - NTLM
  - NTLMv2
  - Kerberos(TODO)
- Based on **Challenge-Response Authentication Protocol** consisting of three me ssages, commonly referred to as **Negotiation**, **Challenge** and **Response Auth entication**
- **NTLM is being replaced by Kerberos, but NTLM still very widely used.**



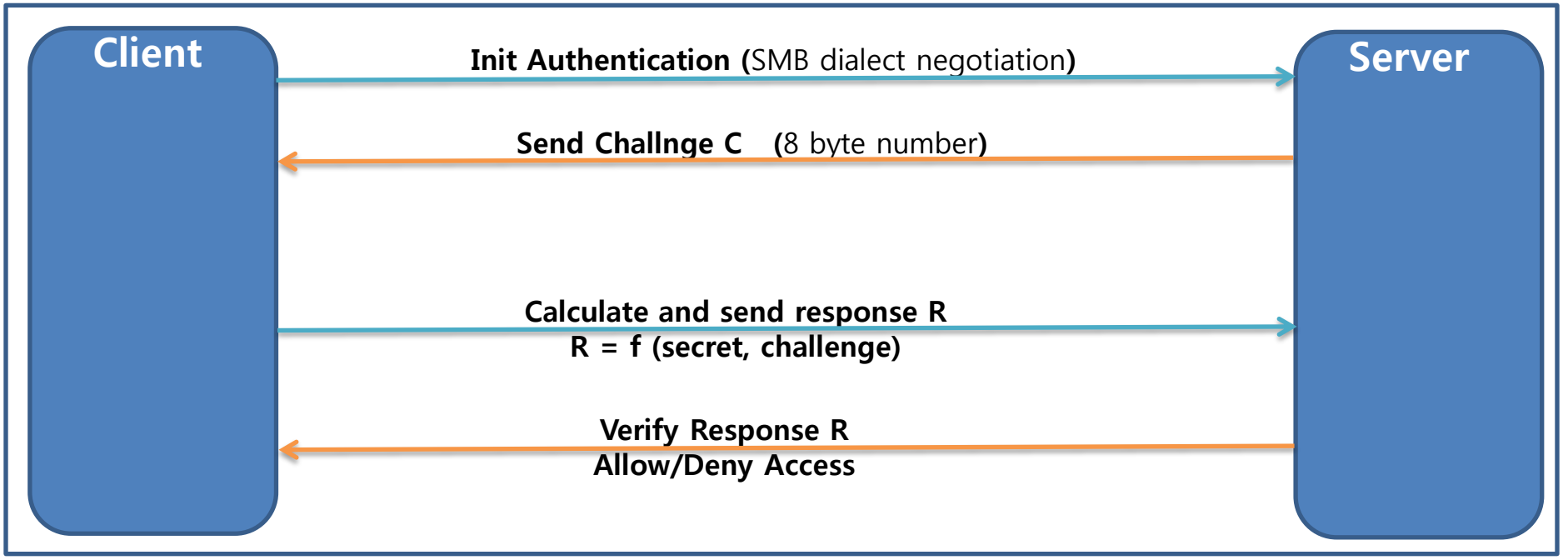
Challenge-Response Authentication

- Server sends "Challenge (C)" to client. Challenge is:
  - 8 byte random number generated by Server
  - Random number to prevent prediction attacks, replay attacks

- Client sends "Response (R)" to Server challenge.

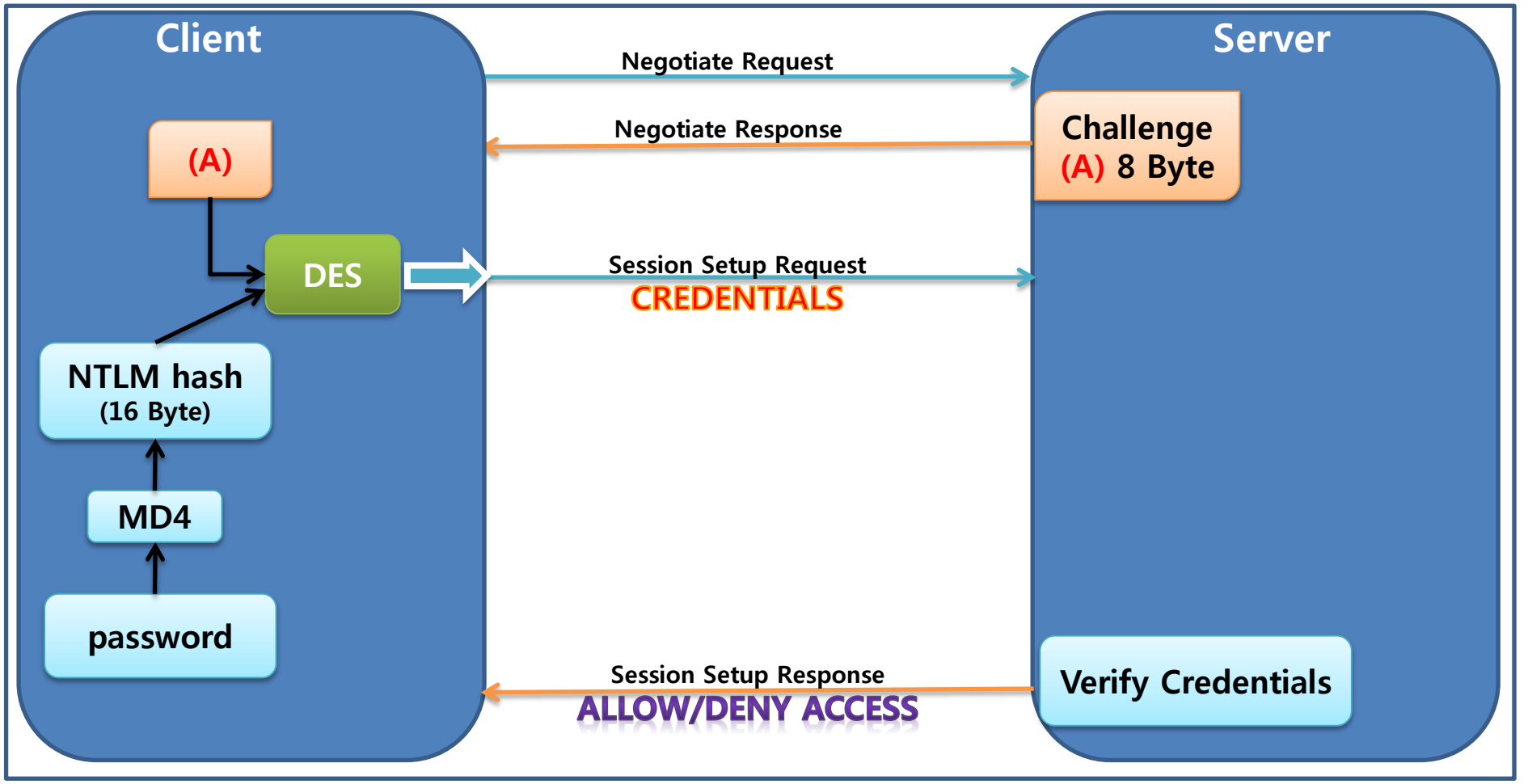
  R = f (secret, challenge)
  - secret: client login user and password
  - Response is encrypted (DES, MD4, HMAC-MD5 etc. algorithms)

- Server verifies client response – allow/deny access



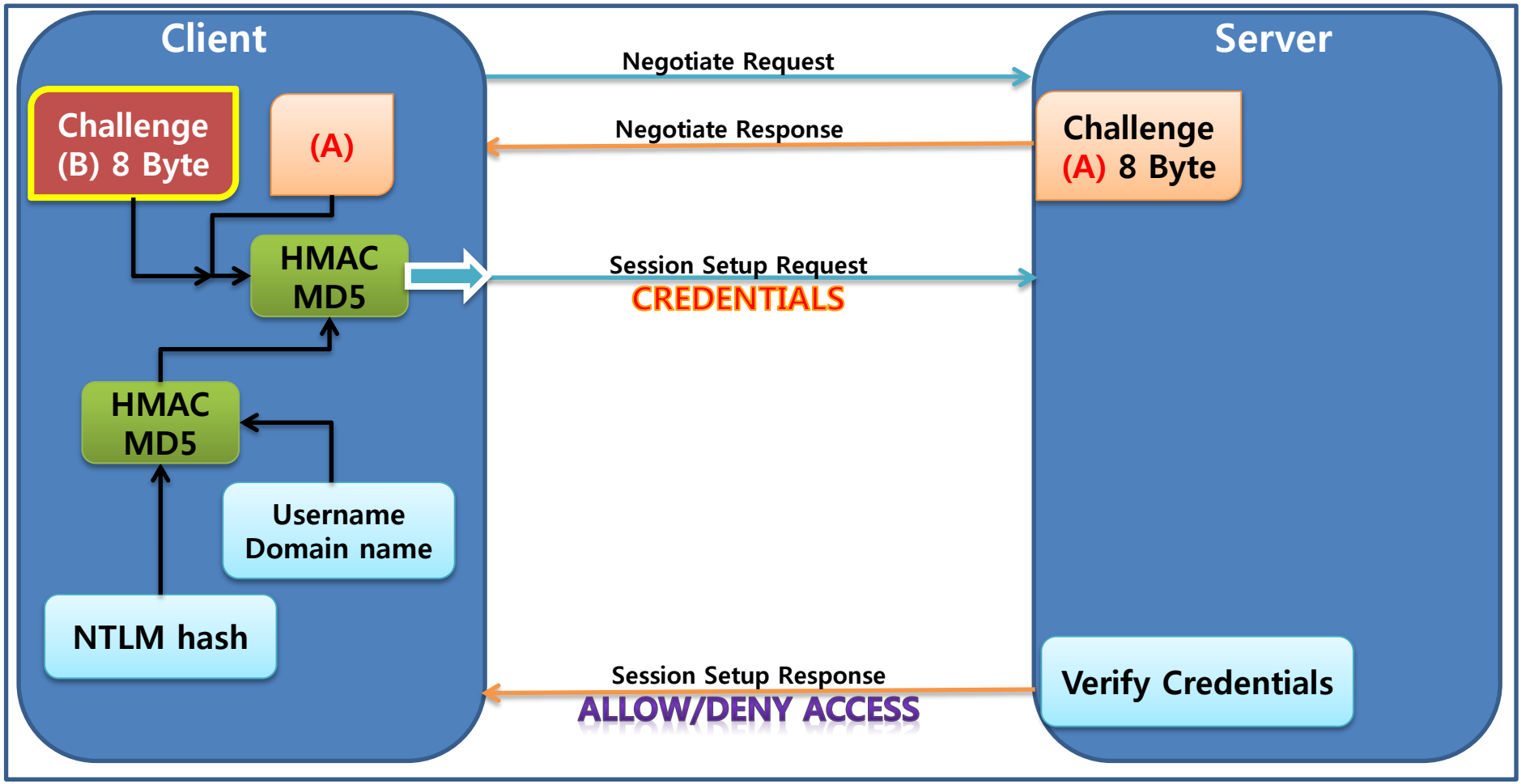
Authentication Methods – NTLMv1

- Only server side challenge – difficult to protects against pre-computed dictiona ry attacks

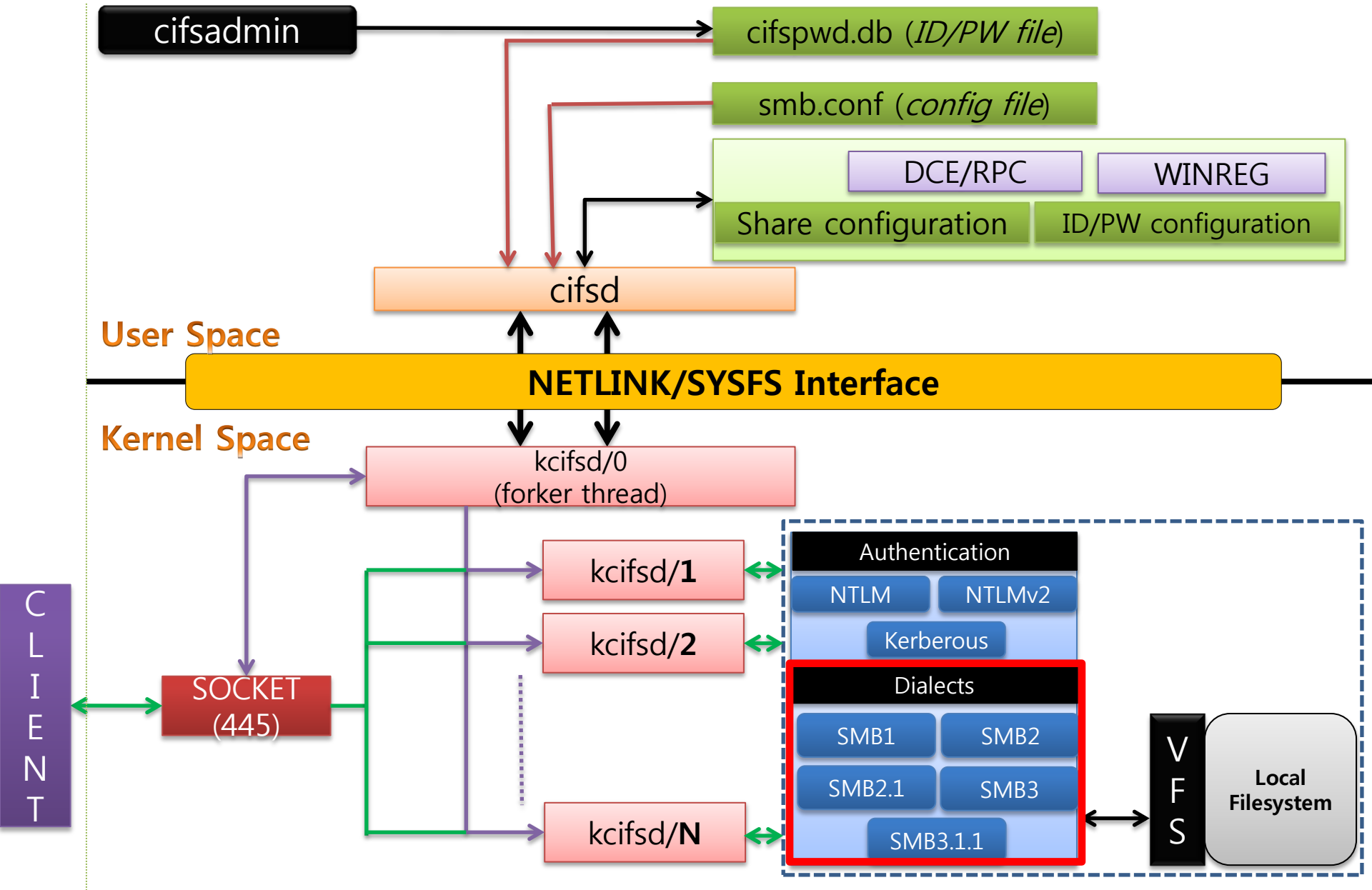


Authentication Methods – NTLMv2

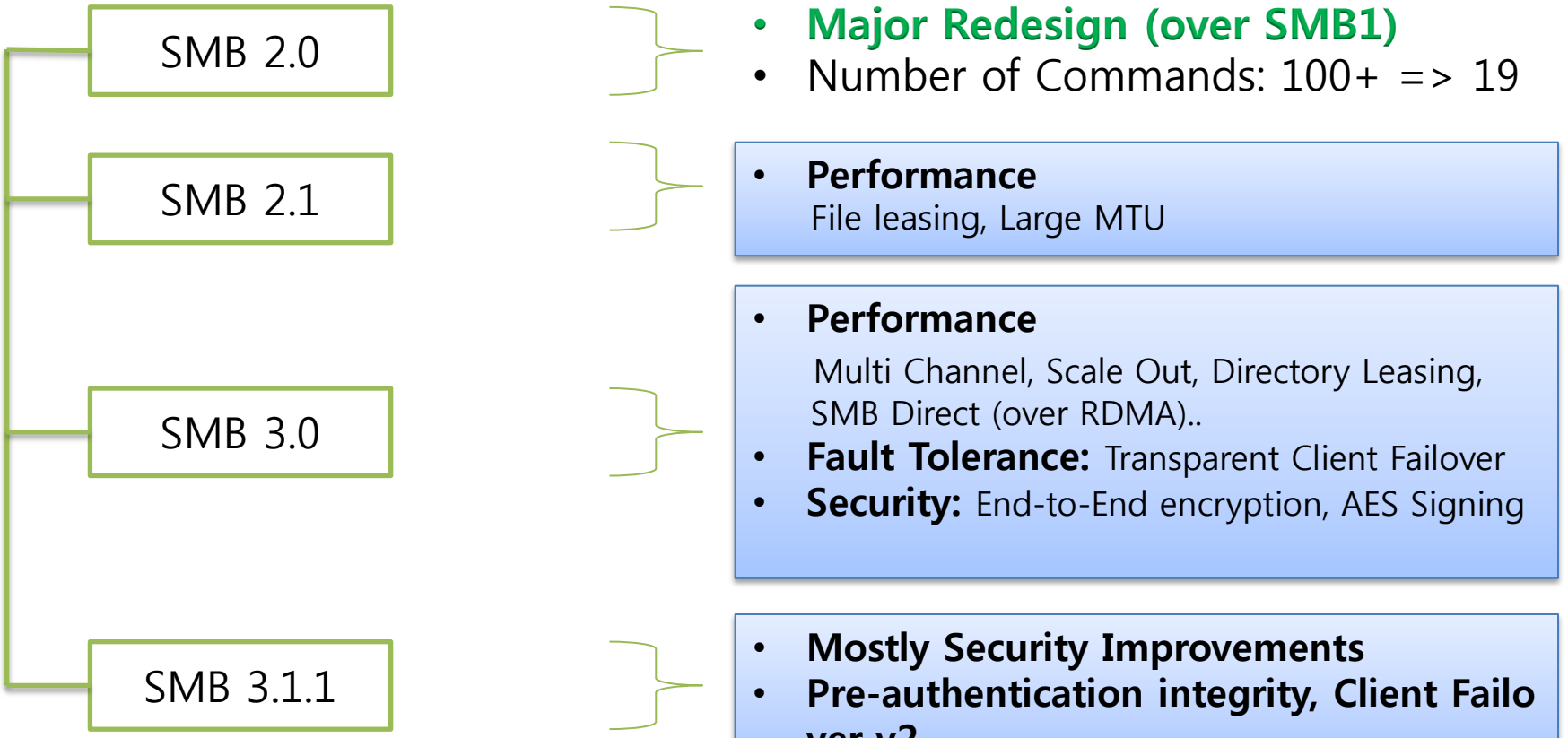
- Additional client side challenge –protects against pre-computed dictionary attacks
- Username/domain name is included in HMAC-MD5 cryptography - more secure than NTLMv1



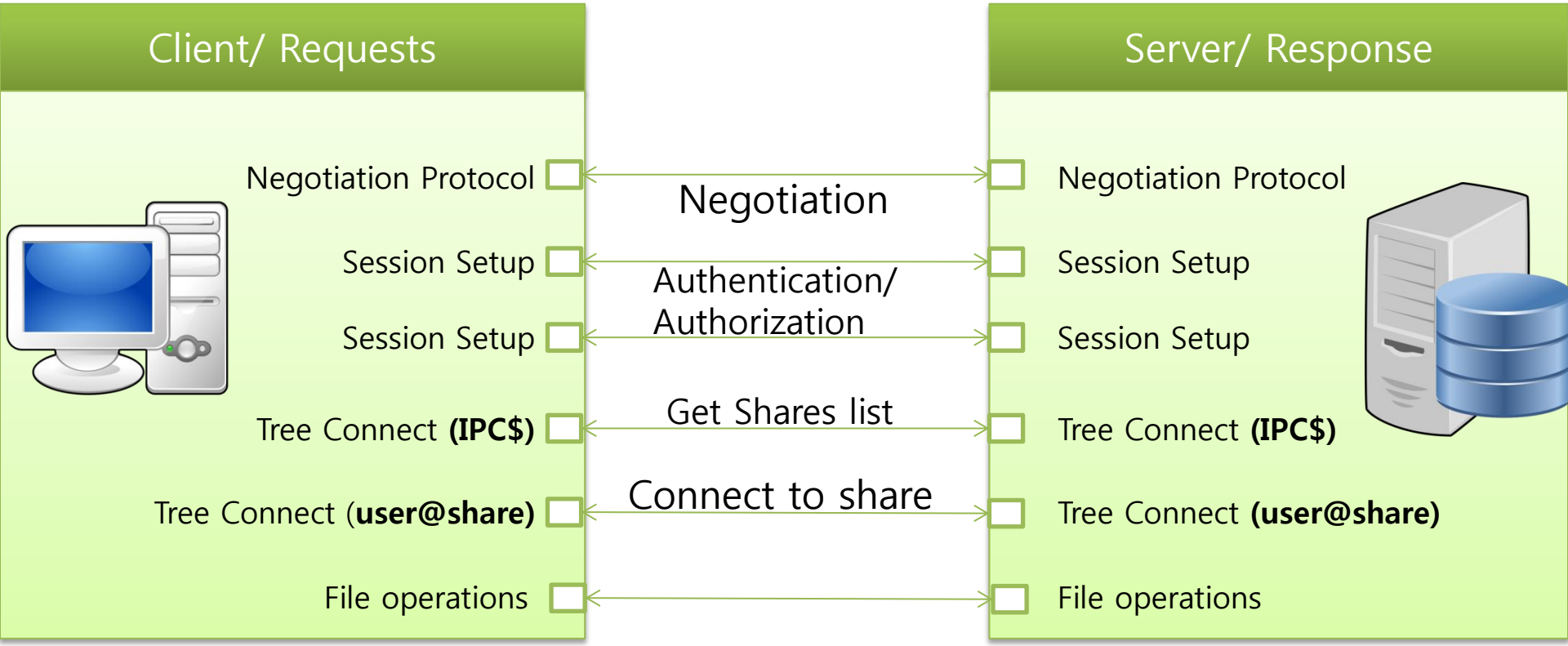
SMB engines


cifsadmin
cifspwd.db (*ID/PW file*)
smb.conf (*config file*)
DCE/RPC
WINREG
Share configuration
ID/PW configuration
cifsd
User Space
NETLINK/SYSFS Interface
Kernel Space
kcifsd/0
(forker thread)
kcifsd/1
kcifsd/2
kcifsd/N
Authentication
NTLM
NTLMv2
Kerberous
Dialects
SMB1
SMB2
SMB2.1
SMB3
SMB3.1.1
SOCKET
(445)
CLIENT
VFS
Local Filesystem

SMB protocol : SMBv2 Evolution



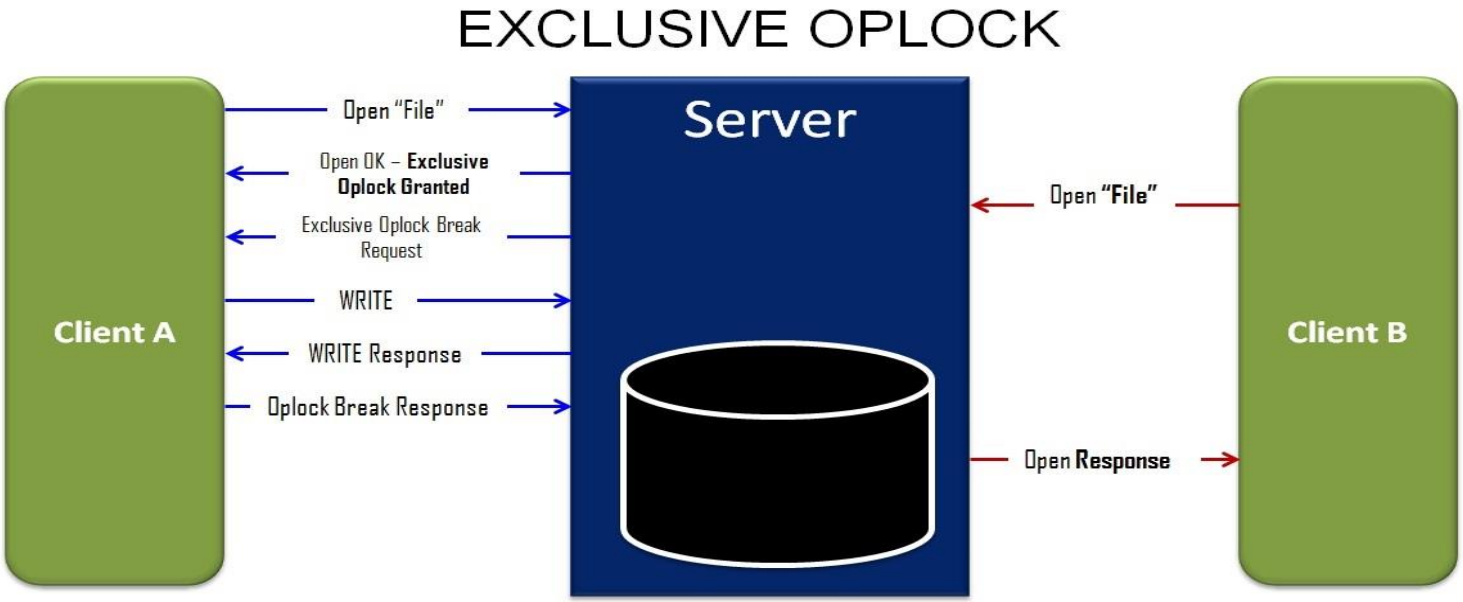
Example of SMB exchange



Opportunistic Locks (Oplocks)

- **Oplock purpose – Client local data caching**
  - Client requests an opportunistic lock during file open to cache data locally
  - Reducing network traffic and improving apparent response time
  - Improved throughput by local caching

- **Data Coherency with Oplock – Oplock Break**
  - Coherent data - data on the server and all the clients is synchronized
  - Oplock Break for data coherency – discard client local cached data



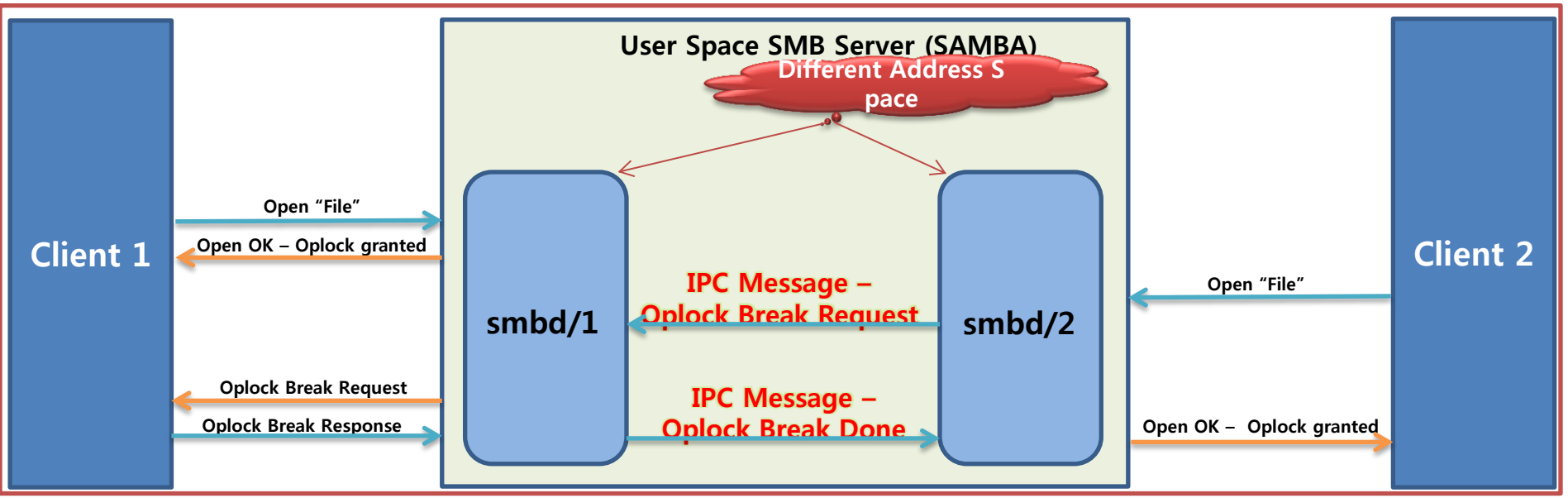
Opportunistic Locks (Oplocks)

- Oplock types in SMB/SMB2
  - Batch Oplock (Read, Write, File Handle caching)
  - Exclusive Oplock (Read, Write caching)
  - Level II Oplock (Read caching)

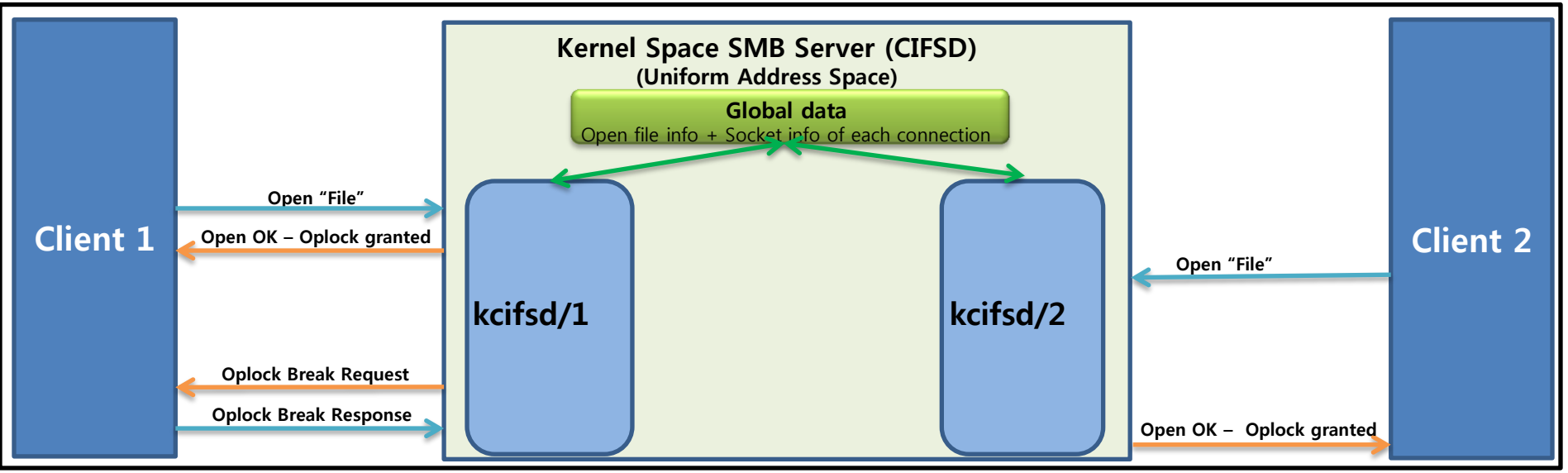
- Lease in SMB2.1
  - New oplock(Lease) added in SMB2.1 to further reduce network traffic by SMB/SMB2 oplock break
  - Uses ClientID(GUID) and Lease Key to uniquely identify a particular client – Oplock owner of a open file
  - Lease is not broken if file opened simultaneously from same client
  - Break lease if another client request lease

- **CIFSD Oplock Advantages**
  - Simplified implementation in kernel space – no need of any IPC communication between different server threads due to uniform kernel address space
  - In kernel space, integration of oplock/lease with NFS Server for data coherency (TODO)

Opportunistic Locks (Oplocks)

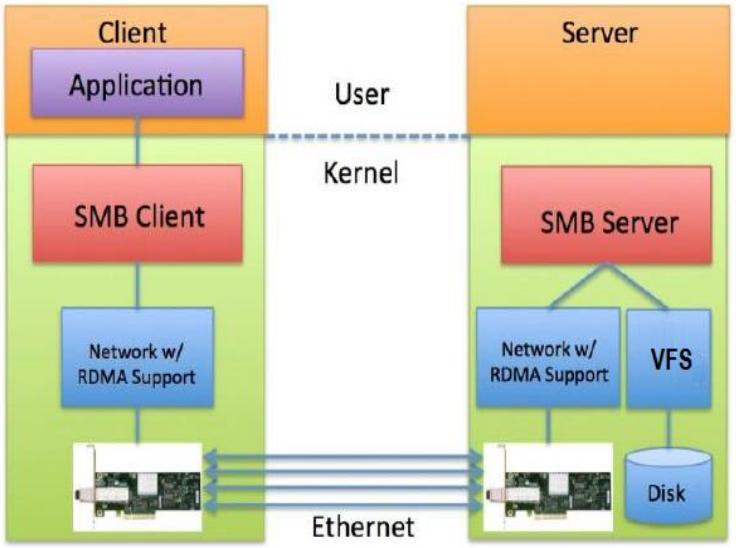


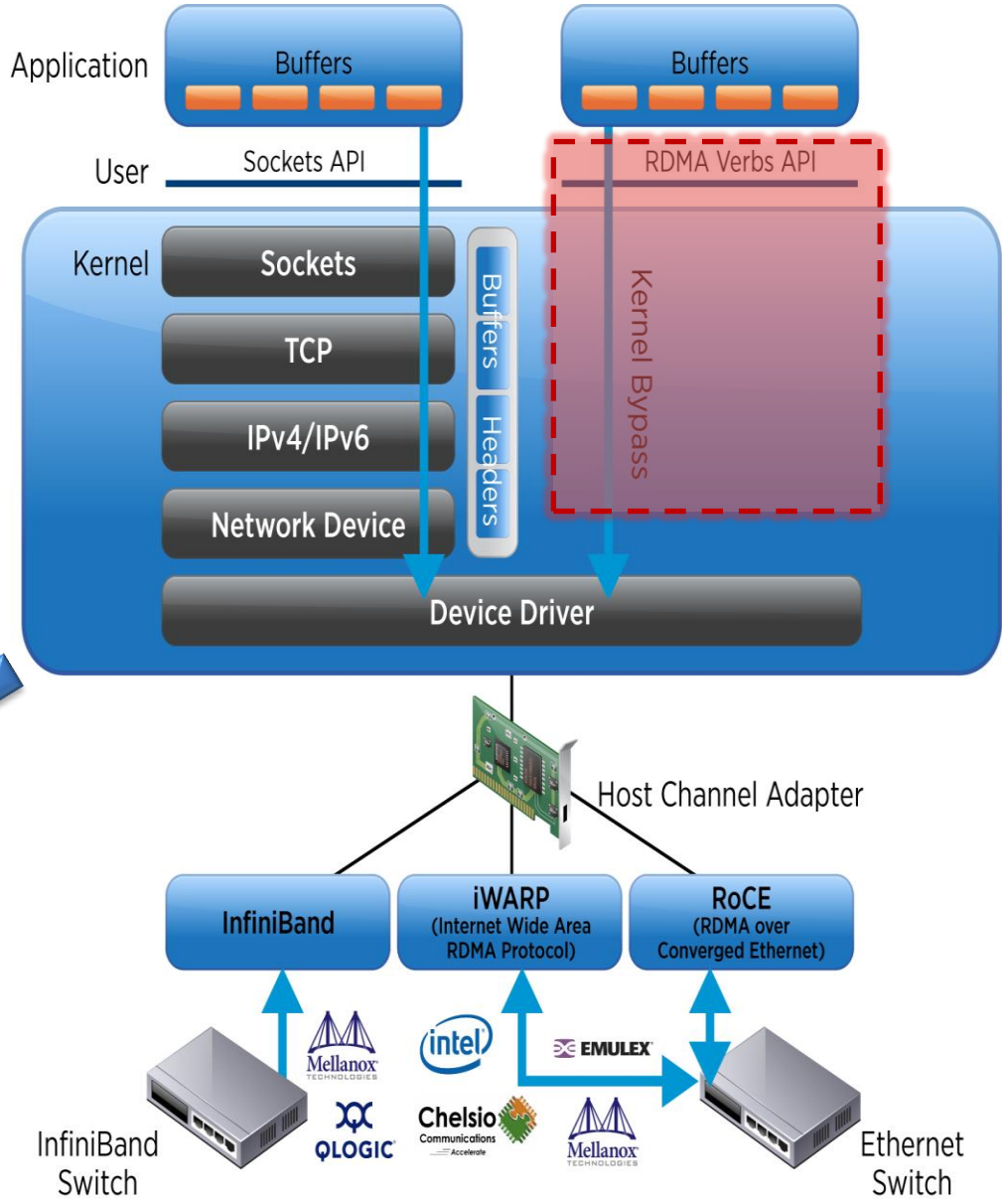
SAMBA versus CIFSD (Oplock Management)



SMB Direct(RDMA)

- **RDMA Main purpose**
  - ✓ Low Latency
  - ✓ Copy Avoidance
  - ✓ Reduces CPU Utilization
  - ✓ Reduces Memory Bandwidth
  - ✓ High Bandwidth utilization

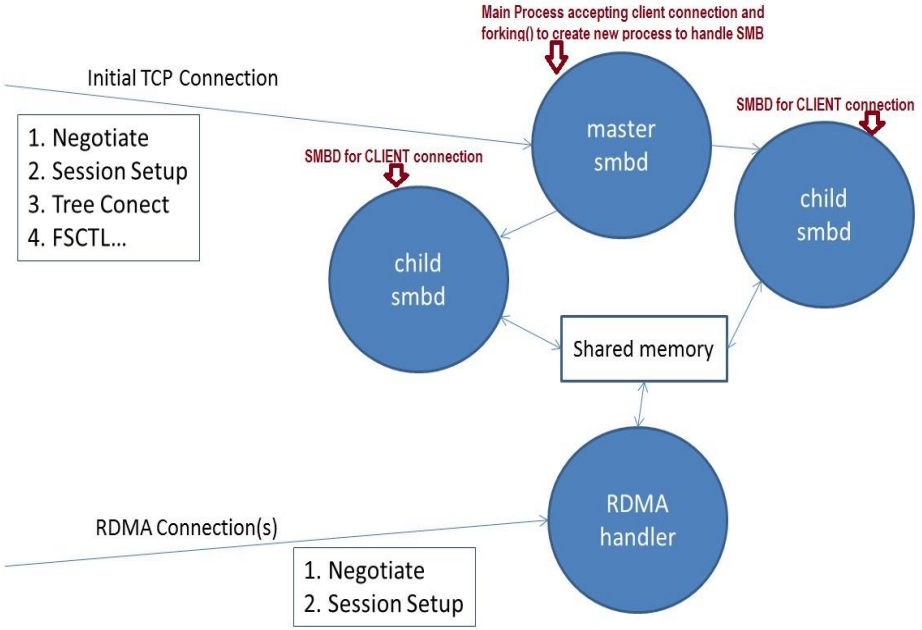




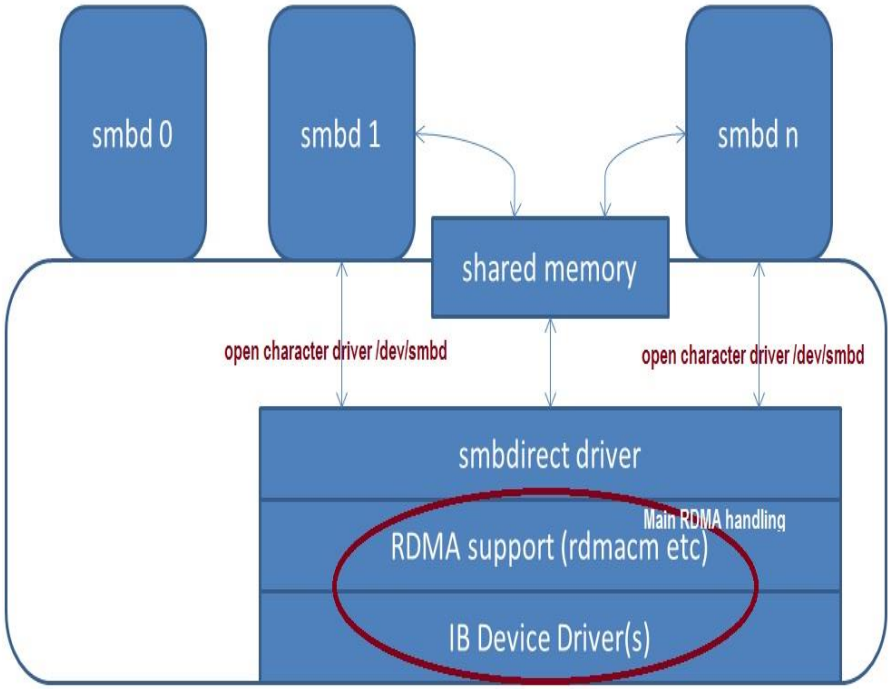
Samba SMB Direct

- 3 kinds of Design considerations to implement RDMA(SMB DIRECT) Support
  1. Convert Samba to a threaded model (Everything in one address space) => Need too many change, max open fd issue)
  2. Separate process to handle all RDMA connections and data transport => many context switch issue
  3. Kernel driver to handle RDMA => no kernel knowledge

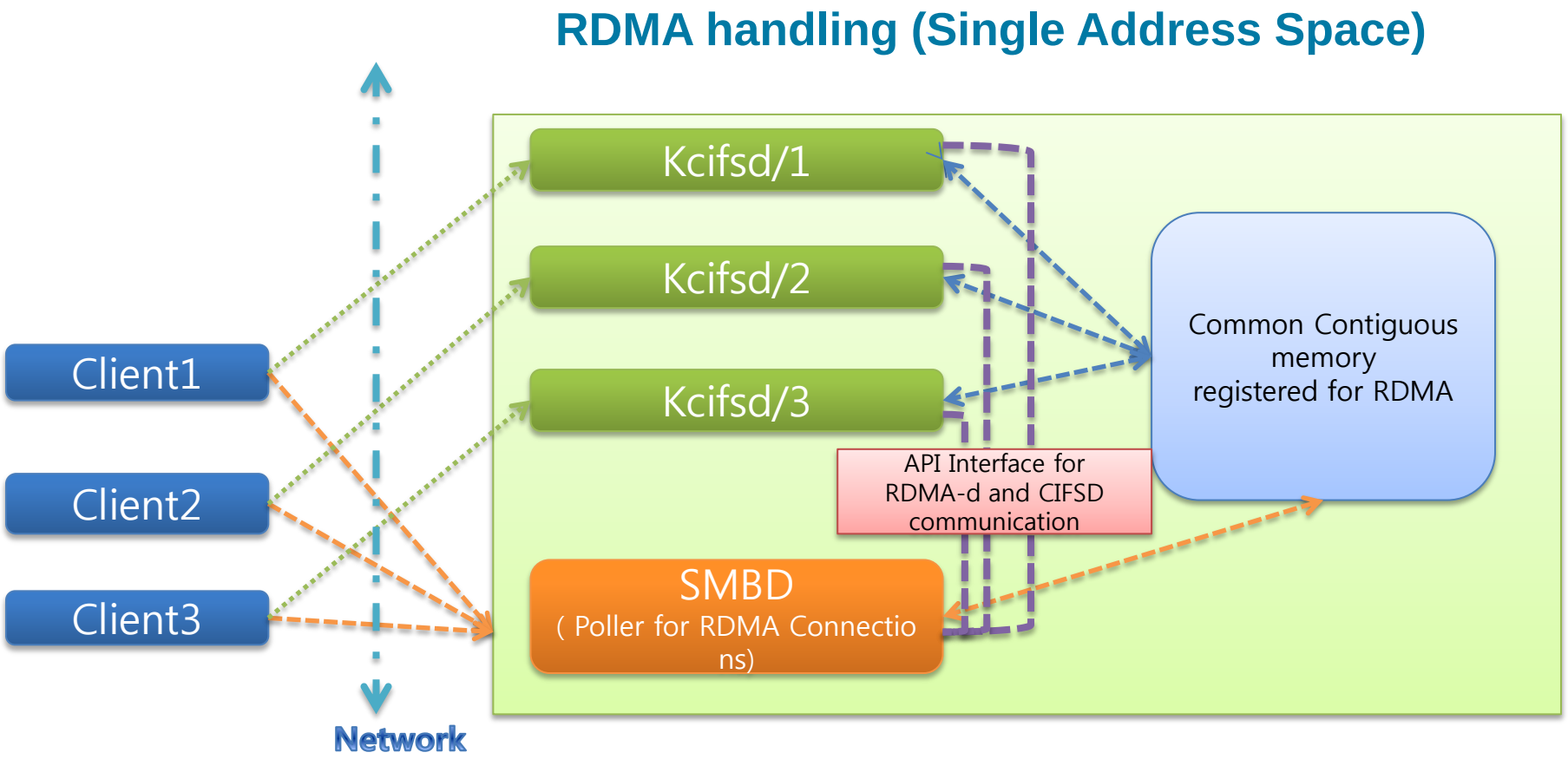
Separate RDMA handler process



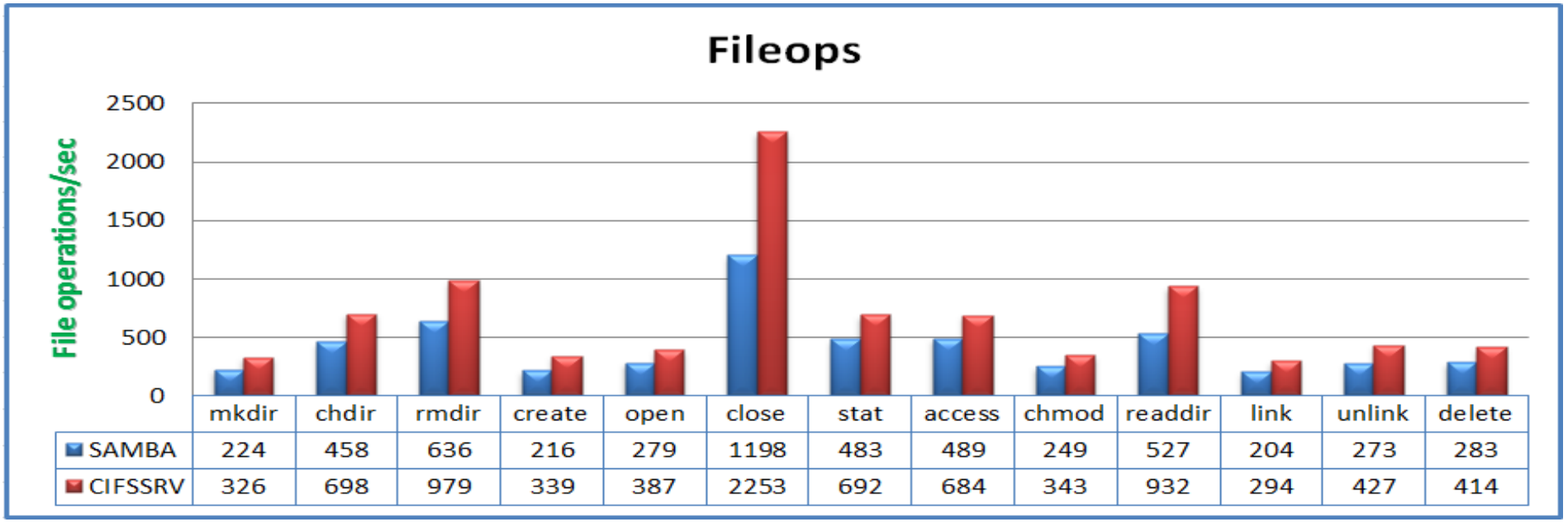
Kernel Driver to handle RDMA



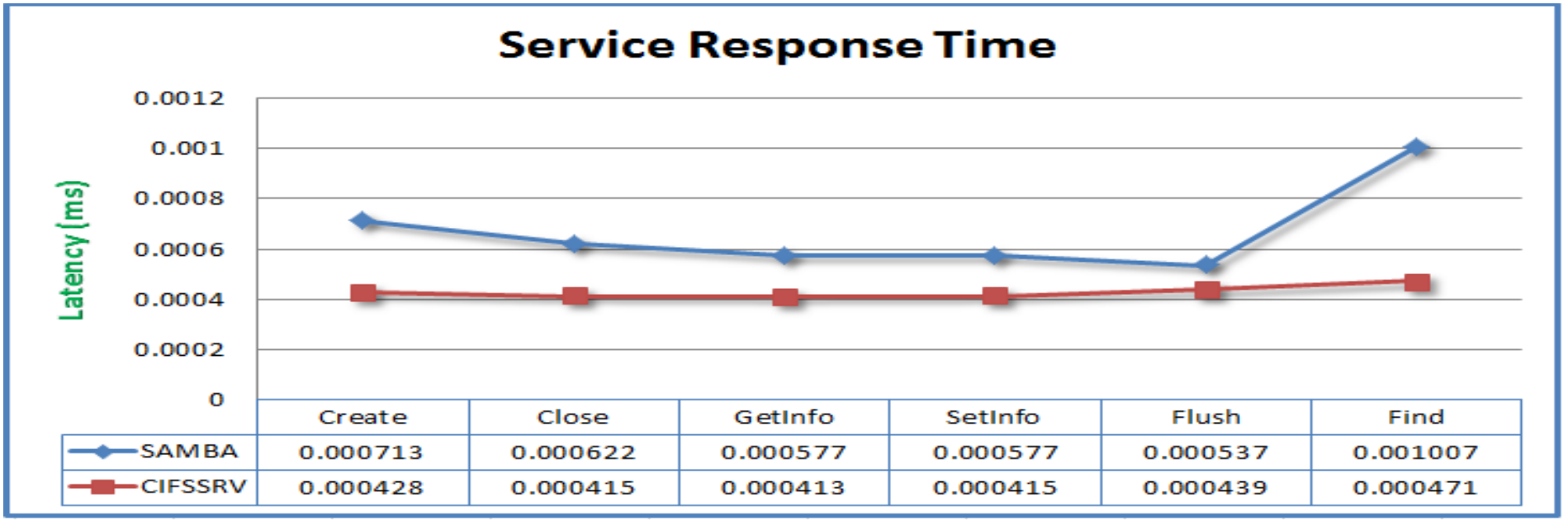
cifsd SMB Direct



Performance comparison



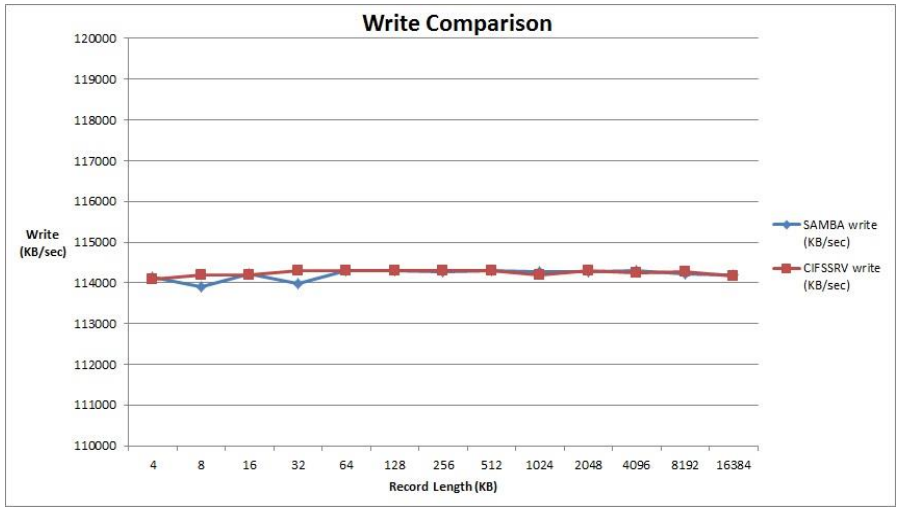
	mkdir	chdir	rmdir	create	open	close	stat	access	chmod	readdir	link	unlink	delete
SAMBA	224	458	636	216	279	1198	483	489	249	527	204	273	283
CIFSSRV	326	698	979	339	387	2253	692	684	343	932	294	427	414

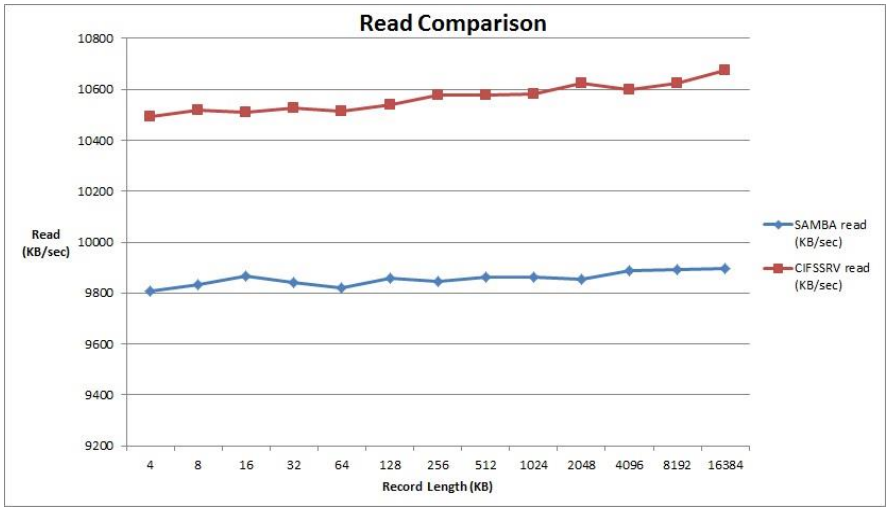


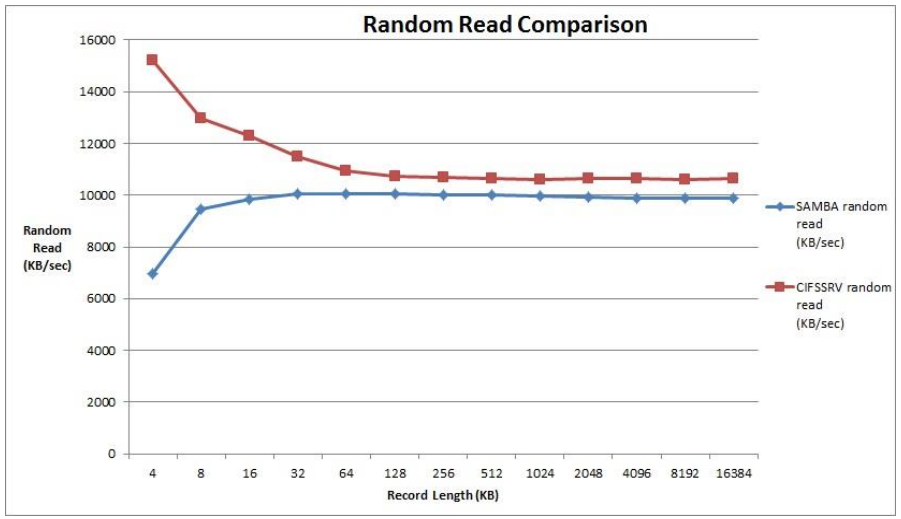
	Create	Close	GetInfo	SetInfo	Flush	Find
SAMBA	0.000713	0.000622	0.000577	0.000577	0.000537	0.001007
CIFSSRV	0.000428	0.000415	0.000413	0.000415	0.000439	0.000471

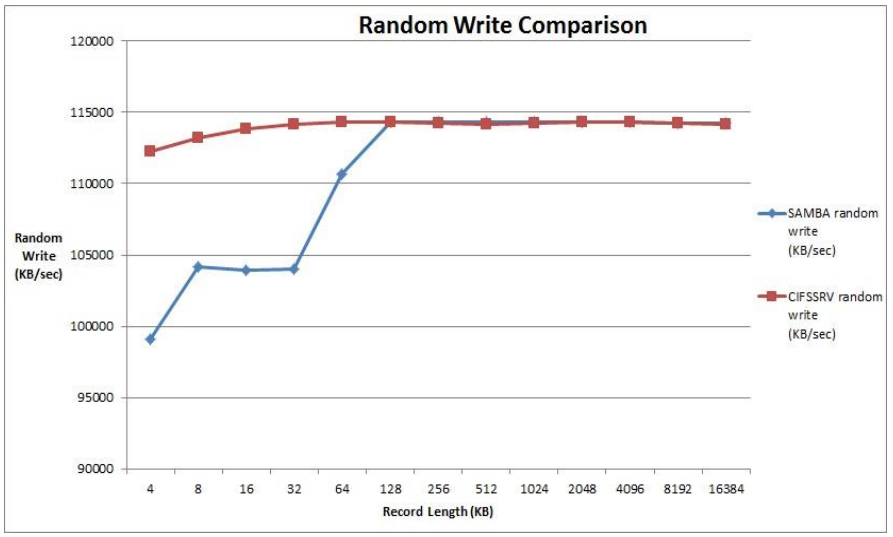
Performance comparison(read/write)

iozone READ/WRITE result



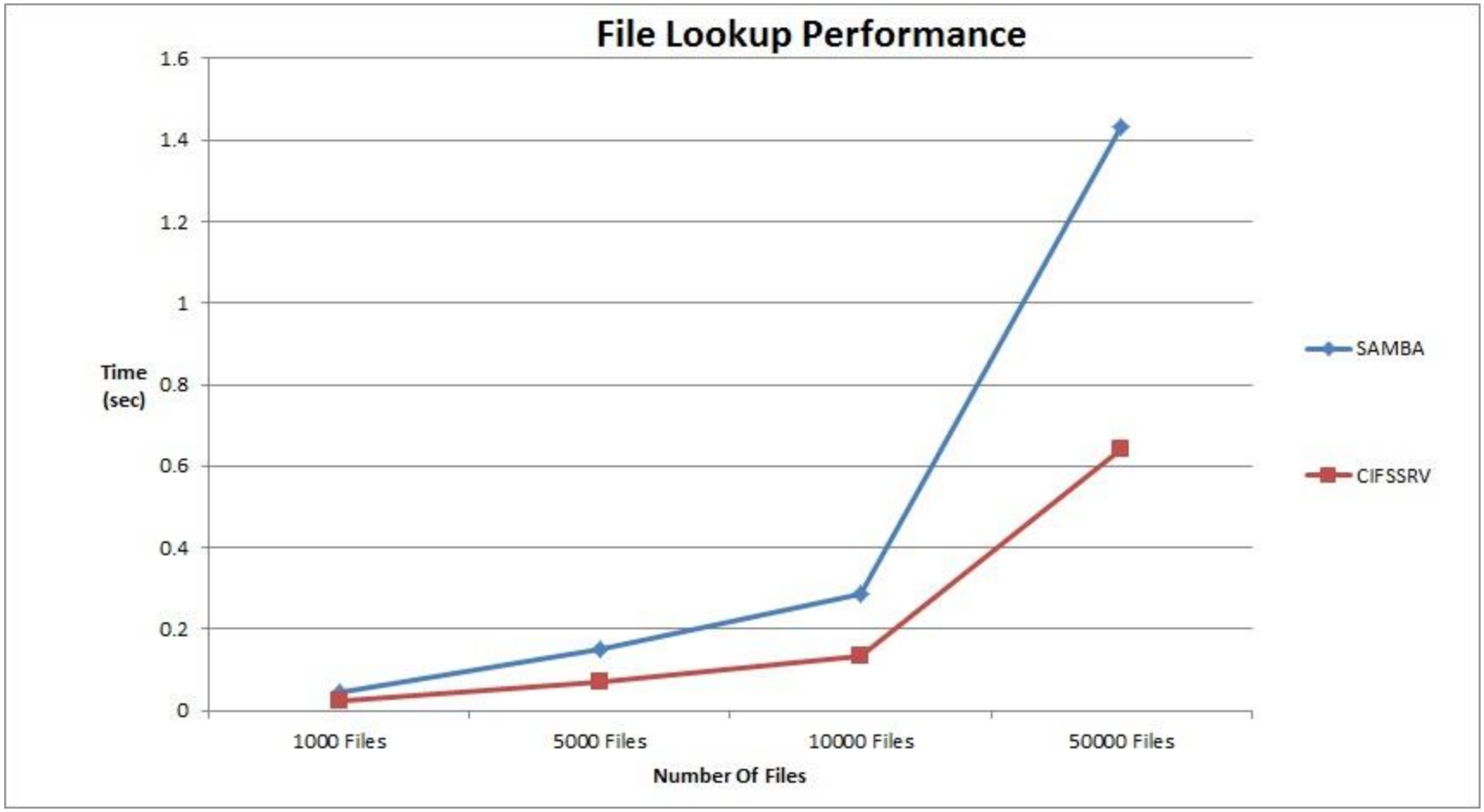






Low Latency, Improved Throughput for each Network I/O

File lookup Performance(ls –l)



Compatibility

SMB CLIENT VERSIONS	CIFSD SUPPORTED
Windows XP (SMB 1.0)	O
Windows Vista (SMB 2 .0)	O
Windows 7 (SMB 2.1)	O
Windows 8 (SMB 3.0)	O
Windows 10 (SMB 3.1.1)	O
Ubuntu File Explorer	O
Linux CIFS Client 4.2.3	O

WIKI & GITHUB

- Wiki :
    1. https://en.wikipedia.org/wiki/CIFSD
    2. https://en.wikipedia.org/wiki/List_of_products_that_support_SMB
    3. https://en.wikipedia.org/wiki/Comparison_of_operating_system_kernels
- cifsd : https://github.com/namjaejeon/cifsd/
- cifsd tools : https://github.com/namjaejeon/cifsd-tools

Future Works

- Stable version release
- SMB Multi-channel, SMB Direct(RDMA) Support
- Compatibility with MacOS, other clients(Smart p hone app)
- Support for Encryption, durable handle v2, direc tory lease feature
- SCSI over SMB3 worked with Azure and HyperV
- Cluster support

Thank you!

namjae.jeon@protocolfreedom.org