



### How Compliant is the Linux client?



Steve French Azure Storage – Microsoft Samba Team

SMB 3.1.1

### Legal Statement

This work represents the views of the author(s) and does not necessarily reflect the views of Microsoft

Linux is a registered trademark of Linus Torvalds.

Other company, product, and service names may be trademarks or service marks of others.

### Outline

- Linux is a lot more than POSIX ...
- What works today?
  - Without POSIX Extensions
  - With POSIX extensions
- Some xfstesting details
- What to work on next?
- How to handle Linux continuing to extend APIs and test it ...

# Linux > POSIX

 Currently huge number of syscalls! (try "git grep SYSCALL\_DEFINE" well over 850 and 500+ are even documented "man syscalls" FS layer has 220). Verified today

#### VS

• Only about 100 POSIX API calls

|     | /stat.c:SYSCALL  | _DEFINE2(fstat64, unsigned long, fd, struct stat64user *         |
|-----|------------------|--|
|     | /stat.c:SYSCALL  | _DEFINE4(fstatat64, int, dfd, const charuser *, filename         |
|     | /stat.c:SYSCALL  | _DEFINE5(statx,  |
|     | /stat.c:COMPAT_  | <pre>SYSCALL_DEFINE2(newstat, const charuser *, filename,</pre>  |
|     | /stat.c:COMPAT_  | <pre>SYSCALL_DEFINE2(newlstat, const charuser *, filename,</pre> |
|     | /stat.c:COMPAT   | SYSCALL DEFINE4(newfstatat, unsigned int, dfd,                   |
|     | /stat.c:COMPAT   | SYSCALL DEFINE2(newfstat, unsigned int, fd,                      |
|     | /statfs.c:SYSCA  | L DEFINE2(statfs, const char user *, pathname, struct st         |
|     | statfs.c: SYSCAL | L DEFINE3(statfs64. const char user *. pathname. size t.         |
|     | statfs.c: SYSCAL | L DEFINE2(fstatfs, unsigned int, fd, struct statfs user          |
|     | statfs.c: SYSCAL | L DEFINE3(fstatfs64. unsigned int. fd. size t. sz. struct        |
|     | statfs.c: SYSCAL | L DEETNE2(ustat, unsigned, dev. struct ustat user *, ubu         |
|     | /statfs.c:COMPA  | SYSCALL DEETNE2(statfs, const char user *, pathname, st          |
|     | /statfs.c:COMPA  | SYSCALL DEFINE2(fstatfs, unsigned int, fd, struct compat         |
|     | statfs c'COMPA   | [  |
|     | Stati Stereon A  |  |
|     | /statfs.c:COMPA  | SYSCALL DEFINE3(fstatfs64, unsigned int, fd, compat size         |
|     | /statfs.c:COMPA  | SYSCALL DEFINE2(ustat, unsigned, dev. struct compat ustat        |
|     | SVDC.C:SYSCALL   |  |
|     | SVDC.C:SYSCALL   | DEFINE1(syncfs, int, fd)   |
|     | SVDC.C:SYSCALL   | DEFINE1(fsync, unsigned int, fd)                                 |
|     | SVDC.C:SYSCALL   | DEFINE1(fdatasync, unsigned int, fd)                             |
|     | SVDC.C:SVSCALL   | DEFINE4(sync file range, int, fd, loff t, offset, loff t,        |
| - 1 |                  | DEFINE4(sync_file_range) int. fd. unsigned int. flags.           |
| - 1 | timerfd.c.SVSC   | ALL DEFINE?(timerfd create, int, clockid, int, flags)            |
|     | timerfd c SVSC   | NI DEFINEA(timerfd settime int ufd int flags                     |
|     | timerfd c SVSC   | NI DEETNE?(timerfd aettime int ufd struct kernel itim            |
|     | timerfd c SVSC   | NI DEETNEA(timerfd settime32 int ufd int flags                   |
| - 1 | timerfd c SVSC   | NI DEETNE2(timerfd gettime32 int ufd                             |
| - 1 | /userfaultfd c   | SVSCALL DEFINE1(userfaultfd int flags)                           |
|     | utimes c:SVSCA   | I DEETNE4(utimensat int dfd const char user * filena             |
| - 1 | utimes c:SVSCA   | L DEFINE3(futimesat int dfd const char user * filena             |
| - 1 | utimes c:SVSCA   | I DEFINE?(utimes char user * filename                            |
|     | utimes c:SVSCA   | I DEFINE2(utime char user * filename struct utimbuf              |
|     | utimes c:SVSCA   | L DEFINE2(utime32 const char user * filename                     |
|     | utimes c:SVSCA   | L DEFINEA(utimensat time32 unsigned int dfd const char           |
|     | flage)           | cc_bchine4(acchensac_cchesz, ansigned inc, ara, const chan       |
|     | /utimes cosvscAl | I DEETNE3(futimesat time32 unsigned int dfd                      |
|     | utimos c:SVSCA   | L DEFINE2(utimes time32) const chas uses * filoname st           |
|     | vatte c'SVSCAL   | DEETNES(setvattr const char user * pathname                      |
|     | vatte cisvscall  | DEETNES(leetvattr const char user * pathname,                    |
|     | vatte c:SVSCAL   | DEFINES((setvattr int fd const char user * name                  |
|     | vatte c:SVSCAL   | DEETNEA(detvattr const char user * pathname                      |
|     |                  | USETINE4(getXatti, const charuser ", pathname,                   |
|     |                  | DEFINE4(lgetXatti, const charuser ", patimame,                   |
|     | Vatte e SVSCAL   | DEFINE2(listwattsconst_chasuses_tastheamachas                    |
|     | Vatte e SVSCAL   | UPETIMES(LISIXALLI, CONSt CharUSer ", pathname, Char             |
|     |                  | UEFINES(ILLSIXALLE, CONST CHAFUSEE *, pathname, Char             |
|     |                  | DEFINES(ILLSLXALLF, LNL, TO, CNAFUSEF *, LLSL, SLZE_L,           |
|     |                  | DEFINE2(Temovexattr, const charuser ^, pathname,                 |
|     |                  |  |
|     | xattr.c:SySCAL   | UEFINEZ(TREMOVEXATOR, INT, TO, CONST CHArUSER *, NAME)           |
|     | rrench@smrrench  | -INLINKPAG-PSZ:~/CLTS-2.65 GIT grep SYSCALL_DEFINE   WC          |
|     | 850 5070         | 69194  |

Linux filesystems are not easy! Responsible for more than 200 of 850 syscalls. +4 since last year

| Syscall name  | Kernel Version<br>introduced |
|---------------|------------------------------|
| epoll_pwait2  | 5.11                         |
| mount_setattr | 5.12                         |
| faccessat2    | 5.8                          |
| close_range   | 5.9                          |

#### goals: Fast! Easy! Transparent!

- Repeating an older slide about goals of SMB3.1.1:
  - Fastest, most secure general purpose way to access file data, whether cloud or on premises or virtualized
  - Implement all reasonable Linux/POSIX features so apps don't know they run on SMB3 mounts (vs. local)
  - As Linux evolves, and needs new features, quickly add to Linux kernel client and Samba and ksmbd



#### Why Not Other Protocols?

- SMB3.1.1 is easily extensible
- SMB3.1.1 works tightly with a set of protocols which can do more than any other file system protocol
- SMB3.1.1 has the best, most exhaustive set of testcases (not just smbtorture ...)
- SMB3.1.1 and related protocols have more documentation (and documentation that has been tested and verified)
- SMB3.1.1 is proven across multiple client types, OS, architectures (and POSIX extensions have been a moving target, done before ...)
- (And don't forget ... SAMBA rocks! And cifs.ko is one of most active FS)

#### What works today without POSIX Extensions

- Normal file and directory operations (open, read, write, fsync, close) to all servers, and hardlinks and even client handled symlinks ("mfsymlinks"), case preserving file name behavior, mapping almost all problematic characters in filenames ("\" is the one exception)
- To most servers:
  - Sparse file operations: setsparse, query allocated ranges, punch hole
  - copy\_range and clone\_range (clone range is less commonly supported)
  - Special file handling via reparse points (or xattrs ala "sfu")
  - Xattrs
- Emulation of mode bits via various alternatives (cifsacl, modefromsid)

#### What can be emulated today without POSIX Extensions

- Fcollapse and finsert
- Most delete and rename scenarios (some exceptions is where the rename fails with access denied with rename onto an existing file)
- Most byte range (easier with OFD rather than "posix" BRLs) and whole file lock scenarios
- Most of the special mode bits

#### What is problematic without POSIX Extensions

- Rename over an open file
- Files with pending delete showing up in the namespace (e.g. readdir)
- Case sensitive file names
- Locking scenarios that require advisory locks
- Populating a few fields in the statfs response (e.g. total and free inodes)
- Ownership can be preserved with "idsfromsid" on create, but only partially if using "multiuser, cifsacl" (especially important if accessing from multiple clients). When using cifsacl, user ownership is ok, but group ownership (gid) of a new file or directory will be the primary group specified by the user, which is not always the correct gid.

### **Quick Overview of POSIX Extensions Status**

Linux kernel client:

- 5.1 kernel or later can be used but 5.8 or later recommended. Enable with mount option "posix." All major features work on client.
  - Readdir, create, mkdir, statfs, queryinfo (stat): complete
  - Support for new reparse tags for special files mostly complete (needs more testing)
- Samba (experimental tree available, enable with smb.conf parm)
  - Server
    - All major features work (thanks to JRA). Merge delayed due to time consuming conflicts with other large charges. Special file handling (Sockets, FIFOs, char device handling) needs to be updated
  - Client tools (smbclient)
    - Major features work. Additional options could be added to cmd set (Thanks to Volker)
- SMB3 Kernel server (cifsd's ksmbd.ko)
  - Partially implemented: it supports the POSIX negotiate context and partially parses POSIX open context
- 3<sup>rd</sup> party prototypes
- Wireshark patches available (network analysis)

#### **POSIX Extensions Easy to Understand**

- A simple negotiate context, an open context, a new file info level and a new fsinfo level
- Everything else relies on existing SMB3.1.1 features



 $\vee$   $\diamond$   $\otimes$ 

#### ▲ ■ Ø ● □ □ X ▲ A A < > > K > □ □ □ X ■ I I I

File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help

#### smb2

| smb2     |                         |             |   |   |   |   | Expression +              |
|----------|-------------------------|-------------|---|---|---|---|---------------------------|
| Jrce     | Destination             | Protoc      | Lenatl Info   |   |   | -NetBIOS Session Service                                      |                           |
| .0.0.1   | 127.0.0.1               | SMB2        | 302 Negotiate Protocol Request  |   |   | ~SMB2 (Server Message Block Protocol version 2)               |                           |
| .0.0.1   | 127.0.0.1               | SMB2        | 366 Negotiate Protocol Response   |   |   | >-SMB2 Header   |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 190 Session Setup Request, NTLMSSP NEGOTIATE                                |   |   | -Negotiate Protocol Request (0x00)                            |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 400 Session Setup Response, Error: STATUS MORE PROCESSING REQUIRED, NTLMSSP | c |   | >-StructureSize: 0x0024                                       |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 460 Session Setup Request, NTLMSSP_AUTH, User: \smfrench                    |   |   | — Dialect count: 4  |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 142 Session Setup Response  |   |   | >-Security mode: 0x01, Signing enabled                        |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 176 Tree Connect Request Tree: \\localhost\IPC\$                            |   |   | - Reserved: 0000  |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 150 Tree Connect Response   |   |   | >-Capabilities: 0x00000077, DFS, LEASING, LARGE MT            | U, PERSISTENT HA          |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 176 Tree Connect Request Tree: \\localhost\test                             |   |   | -Client Guid: 032f6ffc-4993-c44d-8b01-425c8694946             | 9                         |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 150 Tree Connect Response   |   |   | — NegotiateContextOffset: 0x0070                              |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 446 Create Request File: ;GetInfo Request FILE_INFO/SMB2_FILE_ALL_INFO      |   |   | — NegotiateContextCount: 4                                    |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 534 Create Response File: ;GetInfo Response                                 |   |   | - Reserved: 0000  |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 191 Ioctl Request FSCTL_QUERY_NETWORK_INTERFACE_INFO                        |   |   | — Dialect: 0x0210   |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 143 Ioctl Response, Error: STATUS_INVALID_DEVICE_REQUEST                    |   |   | — Dialect: 0x0300   |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 175 GetInfo Request FS_INFO/FileFsAttributeInformation File:                |   |   | — Dialect: 0x0302   |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 162 GetInfo Response  |   |   | — Dialect: 0x0311   |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 175 GetInfo Request FS_INFO/FileFsDeviceInformation File:                   |   |   | >-Negotiate Context: SMB2_PREAUTH_INTEGRITY_CAPABI            | LITIES                    |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 150 GetInfo Response  |   |   | >-Negotiate Context: SMB2_ENCRYPTION_CAPABILITIES             |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 175 GetInfo Request FS_INFO/FileFsVolumeInformation File:                   |   |   | >-Negotiate Context: Unknown Type: (0x5)                      |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 168 GetInfo Response  |   |   | Negotiate Context: SMB2_POSIX_EXTENSIONS_CAPABIL              | ITIES                     |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 175 GetInfo Request FS_INFO/FileFsSectorSizeInformation File:               |   |   | — Type: SMB2_POSIX_EXTENSIONS_CAPABILITIES (0x03)             | 100)                      |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 170 GetInfo Response  |   |   | — DataLength: 16  |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 158 Close Request File:   |   |   | -Reserved: 0000000  |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 194 Close Response  |   |   | — POSIX Reserved: 0x5025ad93                                  |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 224 Ioctl Request FSCTL_DFS_GET_REFERRALS, File: \localhost\test            |   |   |   | -                         |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 143 Ioctl Response, Error: STATUS_NOT_FOUND                                 |   |   | aaba 11 03 00 00 00 00 01 00 26 00 00 00 00 00 01 00          | · · · · · · · · & · · · · |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 262 Create Request File:  |   |   | 2000 20 00 01 00 b4 02 72 b9 3c ed 58 84 13 7e b1 3f          | · · · · · r · ~ · X · ·   |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 354 Create Response File:   |   |   | 00d0 4d 7a 75 80 ee 25 a5 e0 81 20 cc 10 57 7e 31 65          | Mzu··%····W               |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 158 Close Request File:   |   |   | 00f0 02 00 01 00 00 00 05 00 12 00 00 00 00 00 00 60 00 02 00 |                           |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 194 Close Response  |   |   | 0100 6f 00 63 00 61 00 6c 00 68 00 6f 00 73 00 74 00          | o·c·a·l· h·o·s            |
| 1.0.0.1  | 127.0.0.1               | SMB2        | 262 Create Request File:  |   |   | 0120 25 50 9c b4 11 e7 b4 23 83 de 96 8b cd 7c                | %P#                       |
|          |                         |             |   |   |   |   |                           |
| 🔵 🗹 Text | t item (text), 24 bytes |             |   |   |   | Packets: 91 · Displayed: 52 (57.1%)                           | Profile: Defaul           |
| 9        | 🧕 wireshark po 🧿        | זרבי תל אבי | בא 📵 keepassdb.k 🛐 wireshark-po 🧧 pres.odp - Li 💻 Quassel IRC 🜏 Telegran    | m | W | Virtual Machi  mount-mkdir 🔳 🔘 🗌 🥑 🛜 🖻 🕂 4                    | • 12:04 PM ≡              |

```
NetBIOS Session Service
-SMB2 (Server Message Block Protocol version 2)
  >-SMB2 Header
  -Negotiate Protocol Request (0x00)
     >-StructureSize: 0x0024
       Dialect count: 4
      -Security mode: 0x01, Signing enabled
       Reserved: 0000
      Capabilities: 0x00000077, DFS, LEASING, LARGE MTU, PERSISTENT HA
       Client Guid: 032f6ffc-4993-c44d-8b01-425c86949469
       NegotiateContextOffset: 0x0070
       NegotiateContextCount: 4
       Reserved: 0000
       Dialect: 0x0210
       Dialect: 0x0300
       Dialect: 0x0302
       Dialect: 0x0311
       Negotiate Context: SMB2 PREAUTH INTEGRITY CAPABILITIES
       Negotiate Context: SMB2_ENCRYPTION_CAPABILITIES
       Negotiate Context: Unknown Type: (0x5)
     Negotiate Context: SMB2_POSIX_EXTENSIONS_CAPABILITIES
        -Type: SMB2_POSIX_EXTENSIONS_CAPABILITIES (0x0100)
         DataLength: 16
         Reserved: 00000000
         POSIX Reserved: 0x5025ad93
```

#### 

File Edit View Go Capture Analyze Statistics Telephony Wireless Tools Help

#### smb2

| smb2    |                         |             |   |         | ×  | Expression +                       |
|---------|-------------------------|-------------|---|---------|--|------------------------------------|
| irce    | Destination             | Protoc      | Lenati Info   | >-Net   | tBIOS Session Service  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 170 GetInfo Response  | V- SME  | 32 (Server Message Block Protocol version 2)   |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 158 Close Request File:   | >-      | SMB2 Header  |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 194 Close Response  |         | Create Request (0x05)  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 224 Ioctl Request FSCTL DFS GET REFERRALS. File: \localhost\test              |         | >-StructureSize: 0x0039  |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 143 Ioctl Response. Error: STATUS NOT FOUND                                   |         | -Oplock: No oplock (0x00)  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 262 Create Request File:  |         | - Impersonation level: Impersonation (2)   |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 354 Create Response File:   |         | -Create Flags: 0x000000000000000   |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 158 Close Request File:   |         | -Reserved: 000000000000000   |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 194 Close Response  |         | >-Access Mask: 0x00000100  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 262 Create Request File:  |         | >-File Attributes: 0x0000000   |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 354 Create Response File:   |         | >-Share Access: 0x00000007, Read, Write, Delete  | e .                                |
| .0.0.1  | 127.0.0.1               | SMB2        | 158 Close Request File:   |         | — Disposition: Create (if file exists fail, els  | se create it) (2)                  |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 194 Close Response  |         | >-Create Options: 0x00000001   |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 446 Create Request File: ;GetInfo Request FILE_INFO/SMB2_FILE_ALL_INFO        |         | ∽-Filename: 0760   |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 534 Create Response File: ;GetInfo Response                                   |         |  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 174 GetInfo Request FILE_INFO/SMB2_FILE_ALL_INFO File:                        |         | Blob Length: 8   |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 244 GetInfo Response  |         | -Blob Offset: 0x0000088  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 158 Close Request File:   |         | — Blob Length: 40  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 194 Close Response  |         | <pre>~-ExtraInfo SMB2_POSIX_CREATE_CONTEXT</pre>   |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 462 Create Request File: .Trash;GetInfo Request FILE_INFO/SMB2_FILE_ALL_INFO; |         | Chain Element: SMB2_POSIX_CREATE_CONTEXT "   | 5025ad93-b49c-e711-l               |
| .0.0.1  | 127.0.0.1               | SMB2        | 310 Create Response, Error: STATUS_OBJECT_NAME_NOT_FOUND;GetInfo Response, Er |         | — Chain Offset: 0x0000000  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 470 Create Request File: .Trash-1000;GetInfo Request FILE_INFO/SMB2_FILE_ALL_ |         | Tag: 5025ad93-b49c-e711-b423-83de968bcd  | 7c                                 |
| .0.0.1  | 127.0.0.1               | SMB2        | 310 Create Response, Error: STATUS_OBJECT_NAME_NOT_FOUND;GetInfo Response, Er |         | -Blob Offset: 0x00000010   |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 462 Create Request File: 0760;GetInfo Request FILE_INF0/SMB2_FILE_ALL_INF0;Cl |         |  |                                    |
| .0.0.1  | 127.0.0.1               | SMB2        | 310 Create Response, Error: STATUS_OBJECT_NAME_NOT_FOUND;GetInfo Response, Er |         | -Blob Offset: 0x00000020   |                                    |
| 7.0.0.1 | 127.0.0.1               | SMB2        | 246 Create Request File: 0760   |         | —Blob Length: 4  |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 310 Create Response File: 0760  |         | v-Data: POSIX Create Context request   |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 158 Close Request File: 0760  | _       | - POSIX perms: 0740  | I                                  |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 194 Close Response  |         |  |                                    |
| 1.0.0.1 | 127.0.0.1               | SMB2        | 462 Create Request File: 0760;GetInfo Request FILE_INF0/SMB2_FILE_ALL_INF0;Cl | 0000    | 00 00 00 00 00 00 00 00 00 00 00 00 08 00 45   | 00                                 |
| .0.0.1  | 127.0.0.1               | SMB2        | 678 Create Response File: 0760;GetInfo Response;Close Response                | 0010    | 00 e8 c8 34 40 00 40 06 73 d9 7f 00 00 01 7f<br>00 01 a0 cc 01 bd b5 6f c3 fc 26 6c fe 77 80 | 00 ···40·0·s····<br>18 ·····o··&l· |
| 🔵 🎽 m   | ount-mkdir-posix.pcapng | 3           |   |         | Packets: 91 · Displayed: 52 (57.1%)  | Profile: Defau                     |
| 3       | 🌢 wireshark po 📀        | הרבי תל אבי | ده 🚯 keepassdb.k 🚺 wireshark-po 🧧 pres.odp - Li 🗮 Quassel IRC 💽 Telegram      | 🐙 Virtu | Jal Machi 🥖 mount-mkdir 🔳 🔘 🔵 🛜 🛃 🛚  | I ▲ 12:05 PM =                     |

```
>-NetBIOS Session Service
-SMB2 (Server Message Block Protocol version 2)
  >-SMB2 Header
  -Create Request (0x05)
     >-StructureSize: 0x0039
       Oplock: No oplock (0x00)

    Impersonation level: Impersonation (2)

      Create Flags: 0x0000000000000000
       Reserved: 0000000000000000
     >-Access Mask: 0x00000100
     >-File Attributes: 0x00000000
     >-Share Access: 0x00000007, Read, Write, Delete
       Disposition: Create (if file exists fail, else create it) (2)
     >-Create Options: 0x00000001
     ∽-Filename: 0760
         Blob Offset: 0x00000078
         Blob Length: 8
       Blob Offset: 0x00000088
       Blob Length: 40
     ✓-ExtraInfo SMB2 POSIX CREATE CONTEXT
       Chain Element: SMB2_POSIX_CREATE_CONTEXT "5025ad93-b49c-e711-
            Chain Offset: 0x00000000
          V-Tag: 5025ad93-b49c-e711-b423-83de968bcd7c
              Blob Offset: 0x00000010
              Blob Length: 16
            Blob Offset: 0x00000020
            Blob Length: 4
          -Data: POSIX Create Context request
             - POSIX perms: 0740
```

### Some key problems with or without POSIX Extensions

- O\_TMPFILE support
- Mapping of POSIX ACLs and RichACL on the wire
- SELinux integration
- Case sensitive xattrs
- Better integration of Quota and Snapshot API with current Linux local fs tools (currently can be viewed with cifs-utils like "smbinfo")

## What Next?

- Examine the xfstest skips (and failures) in much detail and add small incremental changes
  - "xfstests" is the standard Linux fs functional test suite and no one file system can pass all tests due to various fs optional features.
  - Some can be emulated some need new flags
- Where that is not possible, consider adding new POSIX extensions version (simply adding additional uuid to the POSIX negotiate context)

## What Next?

- What about minor extensions to reduce roundtrips and provide better/safer emulation?
  - Fcollapse and finsert are two examples
  - NTFS fsctls like FSCTL REARRANGE\_FILE and SHUFFLE\_FILE could help if available over SMB3
  - What about exposing Windows's FILE\_FLAG\_POSIX\_SEMANTICS
  - More compounding can help too
  - What about adding rename swap?

### **Examples from xfstest investigations**

- Add support for renameat2 and rename exchange
- POSIX ACLs (can be emulated and there is pushback on implementing primitive POSIX ACLs)
- Support for additional chattr flags ("immutable" and "noatime" updates e.g.)
- fallocate –collapse-range
- Dedupe support
- Defragmentation support (may require VFS changes)

## **Examples from xfstest investigations**

- Richacl support (tests 362 through 370) ??
- O\_TMPFILE support (emulatable, but VFS changes would help)
- FITRIM support (may be emulatable)
- Quota support (may be emulatable already)
- Support for NFS export (nfs server on smb3 mounts)
- Case sensitive xattrs (EAs)
- SELinux support

### **Examples from xfstest investigations**

- Support for online 'label manipulation' (see e.g. xfstest generic/492)
- Support for casefolding ("chattr +F")
- Would native (rather than emulated) BSD flock (whole file lock) support help?

## More details (with example xfstest #)

- atime options irrelevant (test 003)
- O\_TMPFILE (generic/004)
- Defragmentation (018)
- Renameat2 (025)
- POSIX ACLs (026)
- FITRIM (038)
- Metadata journaling (049)
- Freezing fsctl (068) https://lwn.net/Articles/287435/



### More details (continued)

- Chattr +ia (079) ("immutable", "append only")
- Chattr +A (277) ("no atime updates")
- Linux disk quotas (082)
- Security (093) and trusted (097) xattr namespaces
- preallocated extent not marked with FIEMAP\_EXTENT\_UNWRITTEN (094)
- Dedupe (121)
- Advisory locks (131)

### More details (continued)

- suid/sgid bits are cleared after direct write (test 355)
- Richacls (362)
- Encryption support (395)
- Timestamp bounds unknown (402)
- chattr +d ("nodump") (424)
- Information about fiemap of attribute fork (425)
- NFS export (open by inode #) (426)
- Backslash in name ("Key urk moo does not exist for FAKESLASH test??" in test 453)

### More details (continued)

- Conflicting xattrs (test 454)
- XATTR\_REPLACE (test 486)
- xfs\_io label (492)
- Lsattr -d (508)
- Xattrs with slashes in name (523)
- Casefolding support (556)
- Dupremove utility (559), actton utility (596)
- Fsverity (571)



 Remember we can prototype to ksmbd as well now ... and experiment ...



### Thank you for your time

• A very exciting time for ...

