



# SMB3.1.1 POSIX Protocol Extensions: Summary and Current Implementation Status

Steve French

Azure Storage – Microsoft

Samba Team

and

Jeremy Allison

Google/Samba Team



POSIX  
SMB  
3.1.1

# Legal Statement

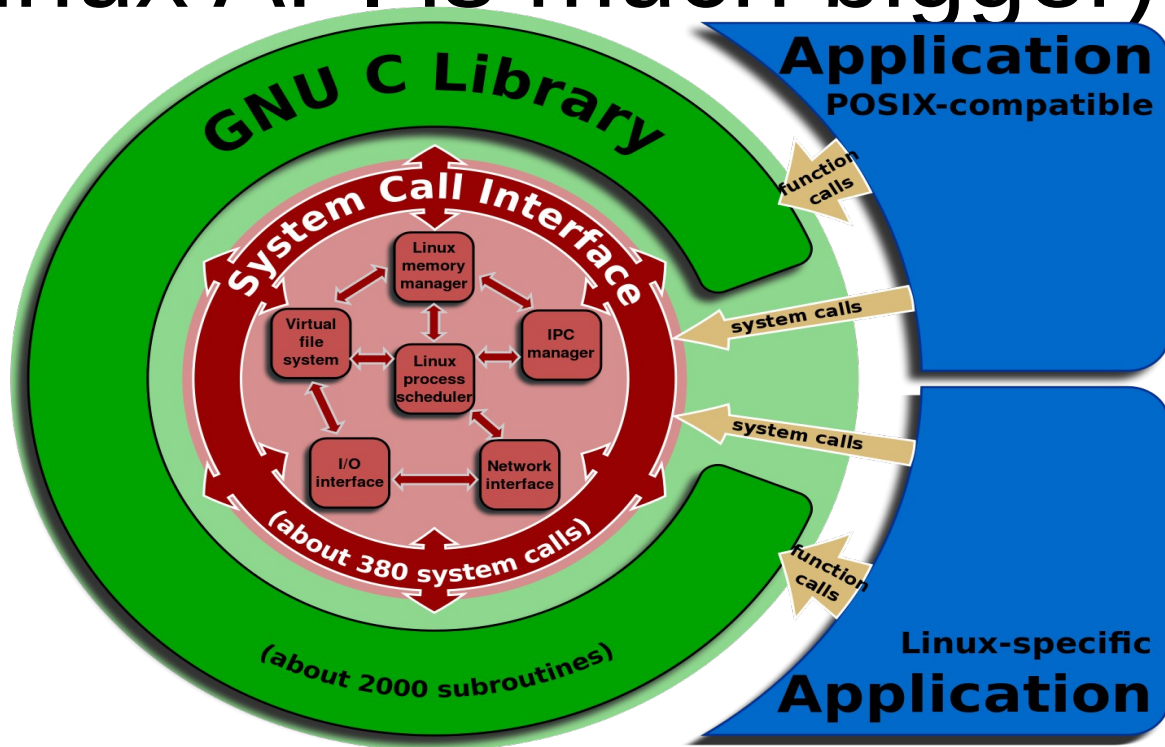
- This work represents the views of the author(s) and does not necessarily reflect the views of Microsoft or Google
- Linux is a registered trademark of Linus Torvalds.
- Other company, product, and service names may be trademarks or service marks of others.

# Outline

- What is POSIX?
- Why do these extensions matter?
- Demo
- What if we don't have them?
  - What works?
  - Some history: CIFS Extensions
  - Alternatives
- Some details
- How to handle Linux continuing to extend APIs, to improve?

# POSIX != Linux

(Linux API is much bigger)



# Linux is BIG

- Currently 293 Linux syscalls!
- vs
- About 100 POSIX API calls



# Example of how Linux is evolving

- ❑ Fallocate alone now has 7 flags
  - Insert range
  - Unshare range
  - Zero range
  - Keep size
  - ...
- ❑ Rename (renameat2) now has 3 flags
  - noreplace, whiteout and exchange
- ❑ And these are just two syscalls ...



# Motivations for Extensions



- Linux Apps work!
  - Case sensitivity e.g. is required for the kernel to build on Linux
  - (And Linux and other posix-like operating systems want posix behavior for files whether on premise or in cloud)
- Improve common situations where customers have Linux and Windows and Mac clients accessing the same data
- Deprecation of CIFS – make sure extensions work with most secure, most optimal SMB3.1.1 dialect



# Why Samba?

Since 1992 ...

Top Server on Linux

- Proven
- Broadly Implemented
- Extensible
- Secure
- Well Tested
- Implements richest File Protocol
- Enormous Client Base (Mac, Windows, now better on Linux! And more)
- 3.5 Million LOC!





# Samba IO Lab

September 17<sup>th</sup>-21<sup>st</sup> 2018

@ Reactor in Redmond, Washington



## A year ago ... and now ... kernel (including SMB3 client cifs.ko) improving

- A year ago Linux 4.16 “Fearless Coyote”



Three days ago: 5.2-rc3 “Golden Lions”



To experiment with POSIX Extensions use 5.1 kernel or backport patch below

- commit 0d481325a9e5e3a31bf83bfcd3690a7a7152ece1

Author: Steve French <stfrench@microsoft.com>

Date: Sun Feb 24 17:56:33 2019 -0600

smb3: Update POSIX negotiate context with POSIX ctxt GUID

# Quality Much Improved – Top Priority

- ❑ More xfstests pass (up to 127 now and growing), vast majority of the rest are skipped due to missing features or being inappropriate for network file systems
- ❑ Crediting (flow control) hugely improved (thanks to Pavel Shilovsky and others)
- ❑ Many potential issues pointed out by static analysis addressed
- ❑ **The “Buildbot”** reduced regressions and is VERY exciting recent addition for CIT (thanks to Ronnie, Aurelien and Paulo)
- ❑ POSIX Extensions (jra’s tree) now a buildbot target for automated regression tests. Will be expanding test list run vs. it soon ...

# POSIX: What could you try today?

- For obvious reasons these experimental changes not enabled by default:
  - With current mainline Linux (4.18 or later) must mount with “vers=3.1.1”  
AND also specify new mount option “posix” and turn off remapping of reserved characters (ie append “nomapposix”)
  - Limited protocol features (posix open context request) can be tried but this small change VERY useful, enough to experiment with and test various apps
- JRA has a tree on samba.org ([git.samba.org/jra/samba/.git](https://git.samba.org/jra/samba/.git) in branch “master-smb2”) with prototype server code
- Other vendors testing experimental distinct implementations of POSIX extensions as well (e.g. at the test event in January)

# Example

- On the client:
  - “mount -t smb3 //<address>/<share> /mnt -o username=<user>,password=<pass>, vers=3.1.1,posix,mfsymlinks,nomapposix,noperm
- On the server add to smb.conf
  - “mangled names = no”
  - “directory mask = 07777”
  - “create mask = 07777”
  - Consider removing “obey pam restrictions”

# POSIX Path Names Work

File Edit View Search Terminal Help

```
root@smf-Thinkpad-P51:~# ls /posix-extensions-mount/
d0754 filecolon: 'fileexclamation!' hello
'file!' 'fileasterisk*' 'filequestion?' newfile
root@smf-Thinkpad-P51:~# ls /scratch
d0754 filecolon: 'fileexclamation!' hello
'file!' 'fileasterisk*' 'filequestion?' newfile
root@smf-Thinkpad-P51:~# uname -a
Linux smf-Thinkpad-P51 5.0.0-rc4+ #67 SMP Sun Jan 27 20:49:32
019 x86_64 x86_64 x86_64 GNU/Linux
root@smf-Thinkpad-P51:~# ~/posix/bin/smbd -V
Version 4.10.0pre1-DEVELOPERBUILD
root@smf-Thinkpad-P51:~#
```

# Note the new mount option “posix” (vs “nounix”)

```
root@Ubuntu-17-Virtual-Machine:~/cifs-2.6# cat /proc/mounts | grep cifs
//localhost/test-no-posix /mnt1 cifs rw,relatime,vers=3.1.1,cache=strict,username=testuser,domain=,uid=0,noforceuid,gid=0,noforcegid,addr=127.0.0.1,file_mode=0755,dir_mode=0755,soft,nounix,serverino,mapposix,rsize=1048576,wsize=1048576,echo_interval=60,actimeo=1 0 0
//localhost/test /mnt cifs rw,relatime,vers=3.1.1,cache=strict,username=testuser,domain=,uid=0,noforceuid,gid=0,noforcegid,addr=127.0.0.1,file_mode=0755,dir_mode=0755,soft,posix,posixpaths,serverino,mapposix,rsize=1048576,wsize=1048576,echo_interval=60,actimeo=1 0 0
root@Ubuntu-17-Virtual-Machine:~/cifs-2.6# cat /proc/fs/cifs/DebugData
Display Internal CIFS Data Structures for Debugging
-----
CIFS Version 2.12
Features: dfs fscache lanman posix spnego xattr acl
Active VFS Requests: 0
Servers:
Number of credits: 16 Dialect 0x311 posix
1) Name: 127.0.0.1 Uses: 2 Capability: 0x300047 Session Status: 1      TCP status: 1
   Local Users To Server: 1 SecMode: 0x1 Req On Wire: 0
   Shares:
   0) IPC: \\127.0.0.1\IPC$ Mounts: 1 DevInfo: 0x0 Attributes: 0x0
      PathComponentMax: 0 Status: 1 type: 0
      Share Capabilities: None      Share Flags: 0x0
      tid: 0x4f5511db Maximal Access: 0x1f00a9

   1) \\localhost\test Mounts: 1 DevInfo: 0x20 Attributes: 0x1006f
      PathComponentMax: 255 Status: 1 type: DISK
      Share Capabilities: None Aligned, Partition Aligned,      Share Flags: 0x0
      tid: 0x8579c31d Optimal sector size: 0x200      Maximal Access: 0x1f01ff

   2) \\localhost\test-no-posix Mounts: 1 DevInfo: 0x20 Attributes: 0x1006f
      PathComponentMax: 255 Status: 1 type: DISK
      Share Capabilities: None Aligned, Partition Aligned,      Share Flags: 0x0
      tid: 0x1813a493 Optimal sector size: 0x200      Maximal Access: 0x1f01ff

MIDs:
```



# Mode bits on create and case sensitivity work!

```
root@Ubuntu-17-Virtual-Machine:/mnt# ~/create-4-files-with-mode-test
root@Ubuntu-17-Virtual-Machine:/mnt# cd /mnt1
root@Ubuntu-17-Virtual-Machine:/mnt1# ~/create-4-files-with-mode-test
root@Ubuntu-17-Virtual-Machine:/mnt1# ls /test /test-no-posix -la
/test:
total 12
drwxrwxrwx  3 root    root    4096 May 31 16:55 █
drwxr-xr-x 32 root    root    4096 May 31 16:46 ..
-rwx----- 1 testuser testuser  0 May 31 16:55 0700
-rwxrwx---  1 testuser testuser  0 May 31 16:55 0770
-rwxrwxr-x  1 testuser testuser  0 May 31 16:55 0775
drwxr-xr-x  2 sfrench sfrench 4096 Mar 24 10:34 tmp

/test-no-posix:
total 8
drwxrwxrwx  2 root    root    4096 May 31 16:55 █
drwxr-xr-x 32 root    root    4096 May 31 16:46 ..
-rwxrw-r--  1 testuser testuser  0 May 31 16:55 0700
-rwxrw-r--  1 testuser testuser  0 May 31 16:55 0770
-rwxrw-r--  1 testuser testuser  0 May 31 16:55 0775
root@Ubuntu-17-Virtual-Machine:/mnt1# mkdir UPPER
root@Ubuntu-17-Virtual-Machine:/mnt1# touch upper
root@Ubuntu-17-Virtual-Machine:/mnt1# cd /mnt
root@Ubuntu-17-Virtual-Machine:/mnt# mkdir UPPER
root@Ubuntu-17-Virtual-Machine:/mnt# touch upper
root@Ubuntu-17-Virtual-Machine:/mnt# ls /test /test-no-posix
/test:
0700 0770 0775 tmp upper UPPER

/test-no-posix:
0700 0770 0775 UPPER
```

# Mode bits on mkdir works!

```
root@smf-Thinkpad-P51:~/cifs-2.6# mount -t smb3 //127.0.0.1/scratch /mnt -o username=testuser,
,vers=3.11,posix
root@smf-Thinkpad-P51:~/cifs-2.6# umask 0000
root@smf-Thinkpad-P51:~/cifs-2.6# mkdir /mnt/0774 -m 0774
root@smf-Thinkpad-P51:~/cifs-2.6# mkdir /mnt/0770 -m 0770
root@smf-Thinkpad-P51:~/cifs-2.6# mkdir /mnt/0444 -m 0444
root@smf-Thinkpad-P51:~/cifs-2.6# ls /scratch -la
total 20
drwxrwxrwx  5 root      root      4096 Jun 15 20:42 
drwxr-xr-x 35 root      root      4096 Jun 15 20:39 ..
dr--r--r--  2 testuser testuser 4096 Jun 15 20:42 0444
drwxrwx---  2 testuser testuser 4096 Jun 15 20:42 0770
drwxrwxr--  2 testuser testuser 4096 Jun 15 20:42 0774
root@smf-Thinkpad-P51:~/cifs-2.6# cat /proc/version
Linux version 4.17.0+ (sfrench@smf-Thinkpad-P51) (gcc version 7.3.0 (Ubuntu 7.3.0-16ubuntu3))
```

# Rename works with POSIX extensions!

```
root@Ubuntu-17-Virtual-Machine: ~  
File Edit View Search Terminal Help  
root@Ubuntu-17-Virtual-Machine:~# ls /mnt-rename-test -la  
total 2852  
drwxr-xr-x  2 root root    0 May 31 18:19 .  
drwxr-xr-x 34 root root 4096 May 31 18:13 ..  
-rwxr-xr-x  1 root root    0 May 31 18:18 emptyfile  
-rwxr-xr-x  1 root root    0 May 31 18:19 emptyfile-posix  
-rwxr-xr-x  1 root root  16 May 31 18:17 targetfile  
-rwxr-xr-x  1 root root  16 May 31 18:19 targetfile-posix  
root@Ubuntu-17-Virtual-Machine:~# mount | grep rename  
//localhost/rename-test on /mnt-rename-test type cifs (rw,relatime,vers=3.1.1,cache=strict,username=testuser,domain=,uid=0,noforceuid,gid=0,noforcegid,addr=127.0.0.1,file_mode=0755,dir_mode=0755,soft,posix,posixpaths,serverino,mapposix,rsize=1048576,wsz=1048576,echo_interval=60,actimeo=1)  
root@Ubuntu-17-Virtual-Machine:~# mv /mnt-rename-test/emptyfile /mnt-rename-test/targetfile  
mv: cannot move '/mnt-rename-test/emptyfile' to '/mnt-rename-test/targetfile': Permission denied
```

```
root@Ubuntu-17-Virtual-Machine: ~  
File Edit View Search Terminal Help  
root@Ubuntu-17-Virtual-Machine:~# mount | grep rename  
//localhost/rename-test on /mnt-rename-test type cifs (rw,relatime,vers=3.1.1,cache=strict,username=testuser,domain=,uid=0,noforceuid,gid=0,noforcegid,addr=127.0.0.1,file_mode=0755,dir_mode=0755,soft,posix,posixpaths,serverino,mapposix,rsize=1048576,wsz=1048576,echo_interval=60,actimeo=1)  
root@Ubuntu-17-Virtual-Machine:~# mv /mnt-rename-test/emptyfile-posix /mnt-rename-test/targetfile-posix  
root@Ubuntu-17-Virtual-Machine:~#
```

## Statfs ("stat -f") without POSIX extensions:

```
root@smf-Thinkpad-P51:~/cifs-2.6# cat /proc/mounts | grep cifs
//localhost/scratch /mnt cifs rw,relatime,vers=3.0,cache=strict,username=testuser,domain=
,uid=0,noforceuid,gid=0,noforcegid,addr=127.0.0.1,file_mode=0755,dir_mode=0755,soft,nou
x,serverino,mfsymlinks,noperm,rsize=1048576,wsiz=1048576,echo_interval=60,actimeo=1 0
root@smf-Thinkpad-P51:~/cifs-2.6# stat -f /mnt
  File: "/mnt"
    ID: 0          Namelen: 4096      Type: smb2
Block size: 1024      Fundamental block size: 1024
Blocks: Total: 234804176  Free: 28323720   Available: 28323720
Inodes: Total: 0        Free: 0
root@smf-Thinkpad-P51:~/cifs-2.6# stat -f /scratch
  File: "/scratch"
    ID: e94471edc7140504 Namelen: 255      Type: ext2/ext3
Block size: 4096      Fundamental block size: 4096
Blocks: Total: 58701044  Free: 10080212   Available: 7080929
Inodes: Total: 14983168  Free: 13901548
```

## Statfs (“stat -f”) with POSIX extensions – works!

```
root@smf-Thinkpad-P51:~/cifs-2.6# cat /proc/mounts | grep smb3
//127.0.0.1/scratch /mnt1 smb3 rw,relatime,vers=3.1.1,cache=strict,username=testuser,doma
in=,uid=0,noforceuid,gid=0,noforcegid,addr=127.0.0.1,file_mode=0755,dir_mode=0755,soft,p
six,posixpaths,serverino,mapposix,noperm,rsize=1048576,wsiz=1048576,echo_interval=60,act
imeo=1 0 0
root@smf-Thinkpad-P51:~/cifs-2.6# stat -f /mnt1
File: "/mnt1"
  ID: 0          Namelen: 4096      Type: smb2
Block size: 4096      Fundamental block size: 4096
Blocks: Total: 58701044  Free: 10080249   Available: 7080966
Inodes: Total: 14983168  Free: 13901538
root@smf-Thinkpad-P51:~/cifs-2.6# stat -f /scratch
File: "/scratch"
  ID: e94471edc7140504 Namelen: 255      Type: ext2/ext3
Block size: 4096      Fundamental block size: 4096
Blocks: Total: 58701044  Free: 10080127   Available: 7080844
Inodes: Total: 14983168  Free: 13901536
```

# Details – Negotiate Req (w/POSIX)

- ❑ Just changed
- ❑ Now 16 byte GUID
- ❑ Allows versioning

No.	Time	Source	Destination	Protocol	Length	Info
4	0.000299443	127.0.0.1	127.0.0.1	SMB2	270	Negotiate Protocol Request
6	0.007362607	127.0.0.1	127.0.0.1	SMB2	366	Negotiate Protocol Response
8	0.007474638	127.0.0.1	127.0.0.1	SMB2	190	Session Setup Request, NTLMSSP_NEGOTIATE
9	0.009171825	127.0.0.1	127.0.0.1	SMB2	360	Session Setup Response, Error: STATUS_MORE_PRO
10	0.009274815	127.0.0.1	127.0.0.1	SMB2	430	Session Setup Request, NTLMSSP_AUTH, User: \tes
11	0.020064831	127.0.0.1	127.0.0.1	SMB2	142	Session Setup Response
12	0.020179968	127.0.0.1	127.0.0.1	SMB2	176	Tree Connect Request Tree: \\localhost\IPC\$
13	0.022185931	127.0.0.1	127.0.0.1	SMB2	150	Tree Connect Response
14	0.022247781	127.0.0.1	127.0.0.1	SMB2	182	Tree Connect Request Tree: \\localhost\scratch
15	0.023237975	127.0.0.1	127.0.0.1	SMB2	150	Tree Connect Response

```

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 HANDLES, DIRECTORY LEASING, ENCRYPTION
      └─ Client Guid: 9c1a6c92-2039-cf41-a995-d5c896dc2fb9
      └─ NegotiateContextOffset: 0x0070
      └─ NegotiateContextCount: 3
      └─ Reserved: 0000
      └─ Dialect: 0x0210
      └─ Dialect: 0x0300
      └─ Dialect: 0x0302
      └─ Dialect: 0x0311
    └─ Negotiate Context: SMB2_PREAUTH_INTEGRITY_CAPABILITIES
      └─ Type: SMB2_PREAUTH_INTEGRITY_CAPABILITIES (0x0001)
      └─ DataLength: 38
      └─ Reserved: 00000000
      └─ HashAlgorithmCount: 1
      └─ SaltLength: 32
      └─ HashAlgorithm: SHA-512 (0x0001)
      └─ Salt: 842d3b5175cda14bc15291acf5d8a823fbfdd19673351560...
    └─ Negotiate Context: SMB2_ENCRYPTION_CAPABILITIES
      └─ Type: SMB2_ENCRYPTION_CAPABILITIES (0x0002)
      └─ DataLength: 4
      └─ Reserved: 00000000
      └─ CipherCount: 1
      └─ CipherId: AES-128-CCM (0x0001)
    └─ Negotiate Context: Unknown Type: (0x100)
      └─ Type: Unknown (0x0100)
      └─ DataLength: 16
      └─ Reserved: 00000000
      └─ Unknown: 93ad25509cb411e7b42383de968bcd7c
  
```

```

0000 00 00 00 00 00 00 00 00 00 00 00 00 08 00 45 00 .....E
0010 01 00 43 51 40 00 40 06 f8 a4 7f 00 00 01 7f 00 ..CQ@
0020 00 01 b9 04 01 bd 68 36 9b 78 e7 7e d9 43 80 18 .....h6-x--C
0030 02 00 fe f4 00 00 01 01 08 0a e3 27 ee ea e3 27 .....SMB
0040 ee e9 00 00 00 c8 fe 53 4d 42 40 00 00 00 00 00 .....
0050 00 00 00 00 02 00 00 00 00 00 00 00 00 00 00 .....
0060 00 00 00 00 00 00 e0 0e 00 00 00 00 00 00 00 .....
0070 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 .....
0080 00 00 00 00 00 00 24 00 04 00 01 00 00 00 77 00 .....S.....w
0090 00 00 92 6c 1a 9c 39 20 41 cf a9 95 d5 c8 96 dc .....1-9 A
00a0 2f b9 70 00 00 00 03 00 00 00 10 02 00 03 02 03 /.p
00b0 11 03 00 00 00 01 00 26 00 00 00 00 00 01 00 .....&
00c0 20 00 01 00 84 2d 3b 51 75 cd a1 4b c1 52 91 ac .....;Q u-K-R
00d0 f5 d8 a8 23 fb fd d1 96 73 35 15 60 80 60 b2 99 .....#...s5
00e0 f1 5b 04 5b 00 00 02 00 04 00 00 00 00 00 01 00 .....[.....
00f0 01 00 00 00 00 00 00 01 10 00 00 00 00 00 93 ad .....#
0100 25 50 9c b4 11 e7 b4 23 83 de 96 8b cd 7c .....XP...#
  
```

# Details continued – Create (POSIX) req

The image shows a Wireshark network traffic analysis window. The filter is set to 'smb2'. The packet list pane shows several SMB2 packets, with packet 1 highlighted in red. The packet details pane shows the structure of the SMB2 Create Request, including chain elements and a data field. The packet bytes pane shows the raw data in hexadecimal and ASCII.

No.	Time	Source	Destination	Protocol	Length	Info
1	0.000000000	127.0.0.1	127.0.0.1	SMB2	382	Create Request File: newposixf
2	0.005646911	127.0.0.1	127.0.0.1	SMB2	390	Create Response File: newposix1
4	0.005807273	127.0.0.1	127.0.0.1	SMB2	174	GetInfo Request FILE_INFO/SMB2_
5	0.005988789	127.0.0.1	127.0.0.1	SMB2	150	GetInfo Response
6	0.006088712	127.0.0.1	127.0.0.1	SMB2	262	Create Request File: newposixf

Chain Element: SMB2\_CREATE\_REQUEST\_LEASE "RqLS"  
Chain Element: SMB2\_CREATE\_DURABLE\_HANDLE\_REQUEST "DHnQ"  
Chain Element: <invalid> "5025ad93-b49c-e711-b423-83de968bcd7c"  
Chain Offset: 0x00000000  
Tag: 5025ad93-b49c-e711-b423-83de968bcd7c  
Offset: 0x00000010  
Length: 16  
Data  
Offset: 0x00000020  
Length: 4

```
0d0 69 00 6c 00 65 00 00 00 00 00 00 00 50 00  i.l.e... ..P.
0e0 00 00 10 00 04 00 00 00 18 00 34 00 00 00 52 71  .....4...Rq
0f0 4c 73 00 00 00 00 32 91 57 d8 3b 99 47 1c a3 7d  Ls...2. W;.G..}
100 a7 81 cc 01 a1 0e 07 00 00 00 00 00 00 00 00 00  .....
110 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00 00  .....
120 00 00 00 00 00 00 00 00 00 00 00 00 28 00  .....(
130 00 00 10 00 04 00 00 00 18 00 10 00 00 00 44 48  .....DH
140 6e 51 00 00 00 00 00 00 00 00 00 00 00 00 00 00  nQ.....
150 00 00 00 00 00 00 00 00 00 00 10 00 10 00 00 00  .....
160 20 00 04 00 00 00 93 ad 25 50 9c b4 11 e7 b4 23  .....%P...#
170 83 de 96 8b cd 7c a4 81 00 00 00 00 00 00 00 00  .....|.....
```

# Details continued – create response

The image shows a Wireshark packet capture window. The filter is set to 'smb2'. The packet list shows several SMB2 packets, with packet 2 highlighted as the 'Create Response File: newposixfi'. The details pane for this packet shows the following structure:

- Length: 52
- LEASE\_V2
- Chain Element: <invalid> "5025ad93-b49c-e711-b423-83de968bcd7c"
- Chain Offset: 0x00000000
- Tag: 5025ad93-b49c-e711-b423-83de968bcd7c
- Offset: 0x00000010
- Length: 16
- Data
- Offset: 0x00000020
- Length: 56

The data field is displayed in hexadecimal and ASCII. The ASCII portion shows characters like '4...Rq', '2. W.;.G..', and '8...'. The status bar at the bottom indicates 'Tag of chain entry (smb2 tag)', 'Packets: 18', 'Displayed: 12 (66)', and 'Profile: Default'.



# Summary of What works

- Without Extensions
- With Extensions

# Example (w/o POSIX extensions) (thank you to Aurelien at SuSE!)

- ❑ Mode bits (in special ACE)
  - Mode bits on left
  - File name on right
- ❑ NB mkfifo not finished
- ❑ Ownership works too

444 mnt/smb/0444

540 mnt/smb/0540

777 mnt/smb/0777

1777 mnt/smb/1777

2777 mnt/smb/2777

3777 mnt/smb/3777

1777 mnt/smb/d01777

2777 mnt/smb/d02777

444 mnt/smb/d0444

4777 mnt/smb/d04777

540 mnt/smb/d0540

777 mnt/smb/d0777

644 mnt/smb/emptyfile

stat: cannot stat 'mnt/smb/fifo': Operation not supported

644 mnt/smb/file-as-sfrench

# CIFS Unix/POSIX Extensions

- What was wrong with what we had?
  - Remember CIFS Deprecation?
  - And not just due to WannaCry ...
    - SMB3 is really good ...
- Apple SMB2/SMB3 create context does handle case sensitivity, but not all POSIX compatibility issues

# Client Perspective

- What about the Linux Kernel?
  - New API changes added about once a year to the VFS (minor global changes added more often, but not all could affect what we need to send on the wire in perfect world ...)
  - Need to quickly update protocol when not possible to do over SMB3
  - Need better interaction with key communities (containers, databases and many others) about what they would like to see

# The challenges of Create/ Rename/Delete

# The challenges of POSIX inode metadata

- What do we need to be able to return?
- What about mode bits and ACLs?

# The Challenges of POSIX locking

# The Challenges of POSIX FS info



Remember the Server  
Perspective?

Details

C U V

# SMB3 POSIX Extensions

- Negotiate Protocol
  - SMB3.1.1 (or later required)
    - POSIX Negotiate Context 0x100
    - Version is included in the context by including the GUID of the supported POSIX open context(s) – currently only one supported
  - If POSIX open contexts not supported, negotiate context must be ignored
  - If POSIX open contexts supported for some files then negotiate context is returned, but server must fail opens with POSIX contexts for files where POSIX is not supported (rather than ignoring the POSIX context)
- Tree Connect – in future dialects tree connect contexts may allow more granularity in allowing servers to tell clients which shares they can't use POSIX opens on
- Case sensitivity yes/no can be exposed via existing QFS Info call

# POSIX Extension

## Requirements

- If server returns a POSIX create context on an open:
  - It supports case sensitive names on this path
  - It supports POSIX unlink/rename semantics on this file
  - It supports advisory (POSIX) locking on this file.
    - Actually they are “OFD” not “POSIX” locks (see e.g. <https://gavv.github.io/blog/file-locks/#emulating-open-file-description-locks>)
  - PATH names are not remapped (no SFU remap needed for \* and \ and > and < and : ...). UCS2 converted directly to UTF-8 and server supports POSIX pathnames

# We Leverage Existing SMB3 features

- Hardlinks use Windows setinfo call (long ago implemented)
- Symlinks, mkfifo, mknod use “nfs reparse point” (MS-FSCC 2.1.2.6)
- ACE with special SID (with mode at end) ala “NFS ACL” mapping can be used to set mode (SID: S-1-5-88-3) see <http://people.redhat.com/steved/Bakeathon-2010/SDC2010-NFS-Windows.jbiseda.20100921.pdf>
- Other linux extensions, e.g. fallocate are mapped to existing SMB3 operations where possible

# SMB311 POSIX Extensions

- Create/Open
  - New POSIX create context
    - If POSIX supported then context must be returned on all opens for which POSIX create context was sent (or open should be failed)
    - It is allowed to have POSIX and non-POSIX opens on the same file
    - It is allowed to have some files in a server which are POSIX and some which are not

# Format of the POSIX owner and mid information in the ACL

<NTSecurityAuthority>-<SECURITY\_NFS\_ID\_BASE\_RID>-  
<NfsSidType>-<NfsSidValue>

- Owner SID for UID: “S-1-5-88-1-<uid>”
- Group SID for GID: “S-1-5-88-2-<gid>”
- Mode SID: “S-1-5-88-3-<mode>”
- Everyone: “S-1-5-88-4”

# Proposed POSIX Infolevels

- Query/SetInfo and Query\_DIR (and also FSInfo)
  - Level 0x64 SMB2\_FIND\_POSIX\_INFORMATION
  - Payload variable (Max = 216 bytes)
    - Timestamps
    - File size
    - Dos attributes
    - U64 Inode number
    - U32 device id
    - U32 zero
    - Struct posix\_create\_context\_response



# Wireshark

- See Aurelien's dissector improvements
  - <https://github.com/aaptel/wireshark/commits/smb3unix>
  - And Pike sample test code
    - <https://github.com/aaptel/pike/tree/smb3unix>

# POSIX Extensions – Where do we go from here?

- Continue debugging test implementations (cifs.ko and JRAs Samba POSIX test branch). Current focus (eg at this test event) on enhancing smb3 client to handle mode bits in ACL, and handle special files and server side symlinks via reparse points
- Continue to add xfstests to the 'jrposix' test group in the buildbot (to regression test the client against Samba server with POSIX extensions)
- Examine EVERY xfstest skip and every xfstest fail for potential match to features in (or to add to) SMB3.1.1 POSIX Extensions
  - e.g. enabling special files (fifos, blkdevs etc.) to be stored as reparse points enables five tests.
- Continue extending the wireshark dissectors (see Aurelien)
- Continue testing here, and in two weeks in Redmond and in September at SNIA SDC
- Continue updating the wiki with details: <https://wiki.samba.org/index.php/SMB3-Linux>
- Questions/comments welcome: samba-technical and linux-cifs lists

Thank you for your time

- This is a very exciting time for ...



**S**  
**+**  
**M**  
**B**  
**3**