SMB3 Offload Update

Zambezi Project

Chris Hertel
Samba Team
SambaXP 2021
Introductions

Me:

- Long-haul Samba Team Member (22 years)
- jCIFS project co-founder
- Book: Implementing CIFS
- Lead author of the SMB1 specs: [MS-CIFS], [MS-SMB]

A ruminant mammal (Geekus geekus) with humped shoulders, spindly legs, rollerblades, and broadly palmated antlers.
The opinions expressed are my own and not necessarily those of:

☆ my spouse,
☆ my dog,
☆ my spirit familiar,
☆ my colleagues,
☆ the monster under the bed,
☆ or the basement bugs.
Zambezi
Project Overview
Zambezi SMB3 Offload

- TOE Cards: TCP Offload Engines
- iSCSI NICs: Networked SCSI
- SMB3 Offload: API Undefined
Zambezi SMB3 Offload

- SMB2 and SMB3 — no SMB1
- Encryption / Compression
- Syntax Layer
  - Message Packing / Parsing
- Host-provided State
  - Authentication and Keys
  - Feature Support
- Fast I/O Processing
The host handles the Semantics:

- **Local Filesystem Interface**
  - E.g.: POSIX Layer
  - Sync'd Access (Local, NFS, Object...)

- **Metadata Management**
  - ACLs, Identity, Attributes

- **File System Behaviors**
  - Locking, Change Notify, etc.
Zambezi SMB3 Offload

The Semantic Layer

...runs above the offload engine.

- The SMBOE API needs to be useful, stable, and open.
- SMB2/3 implementations should all be able to use the same API.
- It should be possible to run different SMBOE implementations in parallel.
Building a Rational, Well-Documented API

- Stackable for adding new dialects, features, and capabilities.
- Device Driver / Library / Toolkit / All?
- It needs a community.
So...

SmartNICS
SmartNICs

Network Interface Cards

with powerful processing capabilities

BRAINS!
SmartNICs

Network Interface Cards with powerful processing capabilities

Programmable:
➢ FPGA-based,
➢ Multi-core ASICs, or
➢ A Combo of the above
### SmartNICs

**Network Interface Cards** with powerful processing capabilities are A generic networking offload device:

- Multi-interface
- Multi-physical layer
- Multi-core processor
- PCIe Interconnect

Ethernet, FC, NVMe-oF, RDMA...
Some SmartNICs can run Linux.
DPU: Data Processing Unit
...the BRAINS of the SmartNIC.

DPUs can also be placed on motherboards.
DPUs are typically RISC-based:  
- ARM  
- MIPS  
- RISC-V?  

DPUs have *lots* of I/O capacity.
For convenience, I arranged the SMB3 commands into six debatably semi-logical categories:

<table>
<thead>
<tr>
<th>Managing Connections</th>
<th>Share Access</th>
<th>Open/Close, Lock/Unlock</th>
</tr>
</thead>
<tbody>
<tr>
<td>● NEGOTIATE (0x0000)</td>
<td>● TREE_CONNECT (0x0003)</td>
<td>● CREATE (0x0005)</td>
</tr>
<tr>
<td>● SESSION_SETUP (0x0001)</td>
<td>● TREE_DISCONNECT (0x0004)</td>
<td>● CLOSE (0x0006)</td>
</tr>
<tr>
<td>● LOGOFF (0x0002)</td>
<td></td>
<td>● LOCK (0x000A)</td>
</tr>
<tr>
<td>● ECHO (0x000D)</td>
<td></td>
<td>● OPLOCK_BREAK (0x0012)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Fundamental I/O</th>
<th>Metadata Query and Set</th>
<th>Odds and Ends</th>
</tr>
</thead>
<tbody>
<tr>
<td>● READ (0x0008)</td>
<td>● QUERY_DIRECTORY (0x000E)</td>
<td>● CANCEL (0x000C)</td>
</tr>
<tr>
<td>● WRITE (0x0009)</td>
<td>● CHANGE_NOTIFY (0x000F)</td>
<td>● SMB2 Error Response</td>
</tr>
<tr>
<td>● FLUSH (0x0007)</td>
<td>● IOCTL (0x000B)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>● QUERY_INFO (0x0010)</td>
<td></td>
</tr>
<tr>
<td></td>
<td>● SET_INFO (0x0011)</td>
<td></td>
</tr>
</tbody>
</table>
Several messages have the same format:

typedef struct
{
    uint16_t StructureSize;
    uint8_t  Reserved[2];
} smb2_BaseMsg;

- LOGOFF Request/Response
- TREE_DISCONNECT Request/Response
- ECHO Request/Response
- CANCEL Request
- LOCK Response
- FLUSH Response
Consider Create

- In SMB2/3, Create is a Monster.
  - Lots of flags fields
  - Access & Share Modes
  - Leases/OpLocks
  - Contexts

It's practically a sub-protocol. How should it be presented?
Zambezi Update

Short Term Goals:
★ Proxy Dæmon
★ Transform Headers
★ Unit Tests
★ Finish Messages (including Create)
★ Study other code
Zambezi Update

Where else might this code be useful?

- Software Defined Storage Switches
- Proxy and Cache Servers
- WAN Accelerators
- Hypervisors
- Remote Access Portals
Zambezi Update

- LGPLv3 & AGPLv3
- Published modules are not well tested, but are fairly complete.
- ...and excessively well documented.

https://gitlab.com/ubiqx/zambezi
In the '67 Doctor Dolittle movie...

The **Giant Luna Moth** is attracted by the light of the moon, and flies there. Once on the moon the Moth is attracted by the light of the earth, and so flies back again.

The cycle repeats.
Distractions

- SmartNIC and DPU Docs and Videos
  - I enjoy learning about all this new stuff
- Embedded Tinkering Platforms (RPi, etc.)
  - **Ultibo** for Raspberry Pi:
    - Development toolkit based on FreePascal
    - Built-in support for NTFS

Block Storage

- AoE is small and simple, so...
Summery [sic]
Project Zambezi

Goals:

★ Git 'er done.
★ Work with the SNIA
  ○ Standardize the API
  ○ Fork a reference implementation under an additional license
★ Partner with others to implement on SmartNICs
★ Find new & interesting use cases