Hardware Acceleration for RAID5/6, Deduplication & Security for parallel workloads

Vikas Aggarwal

Storage Developer Conference Bangalore 2017
Team Members

- Venkatesh Bolla
- Achyutha Krishna
- Anil Reddy

Responsibilities: Storage software acceleration on OCTEON TX.
Outline

- RAID 5/6
- Deduplication
- Security
- Open Source Storage Applications
- OCTEON TX Acceleration blocks
- Accelerated Data Flow
- Observations
Offloading CPU intensive operations

- RAID6
- XOR
- Galois Field Multiplication

- Deduplication
  - Fingerprint Generation
  - Fingerprint Lookup

- Security
  - Encryption/Decryption of Data Blocks
## Open Source Storage Software Offload Case Studies

<table>
<thead>
<tr>
<th>RAID</th>
<th>DEDUPLICATION</th>
<th>SECURITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>MD RAID</td>
<td>DM Dedup</td>
<td>ecryptfs</td>
</tr>
<tr>
<td>DM RAID</td>
<td>OpenZFS</td>
<td>DM Crypt</td>
</tr>
<tr>
<td>btrFS</td>
<td>btrFS</td>
<td></td>
</tr>
<tr>
<td></td>
<td>OpenDedup</td>
<td></td>
</tr>
</tbody>
</table>
Linux MD RAID

- Implements the following:
  - RAID 0, 1, 5, 6

- MDRAID can offload following to hardware using async_tx Linux offload infrastructure:
  - memcpy
  - XOR
  - Galois Field Arithmetic

- Current Offload benefits the following RAID variants:
  - RAID 5, 6
Linux DM-dedup

- Implements the following:
  - 4KB block fingerprint
  - Fingerprint to PBN (Physical block number) lookup
  - LBN to PBN (Physical block number) lookup

- DM Dedup (modified) can offload following to hardware:
  - Fingerprint (Digest using MD5, SHA1, SHA2)
  - Fingerprint and LBN lookup.

- Current Offload benefits the following:
  - Lookups.
Linux eCryptfs

- In-kernel standalone implementation.
- Security gets inherited into incremental backups.
- Cryptographic metadata is stored along with encrypted file.
- Supports Linux cryptographic ciphers.
- Utilizes Linux crypto framework
- eCryptfs can offload following to hardware
  - AES CBC
  - DES3 CBC
  - AES XTS
- Offload benefits: Encryption, Decryption.
RAID Architecture

Application

Filesystem

MDRAID Block Device

RAID5/6, async_tx

Data Block
Data Block
Data Block
P Block
Q Block
Linux MDRAID Offload

VM (1..N) \( \rightarrow \) Network Clients (1..N) \( \rightarrow \) READ / WRITE

BTRfs / ext4 \( \rightarrow \) linux mdraid

1..N RAID-6 Volumes

BTRfs / ext4 \( \rightarrow \) linux mdraid

Linux async_tx

DDR \( \rightarrow \) RAD Offload Block

Cavium async_tx driver
Dmdedup Architecture

Application

Filesystem

Dmdedup Block Device

Deduplication Logic

Meta Disk

Data Disk
DmDedup WRITE Offload

Data Chunk

Compute HASH

Hash->PBN

Dup WR: Yes

LBN->PBN

Yes: dmdedup code

No: Uniq WR

LBN->PBN

DDF Offload Block

DISK

dmdedup code

C A V I U M
DmDedup READ Offload

1. Sector
2. Compute LBN
3. LBN->PBN
   - Miss
   - Hit
4. DDF Offload Block
5. DISK
   - END
   - DATA

Code: dmdedup
Storage Security Offload

VM (1..N) ➔ READ / WRITE ➔ eCryptfs ➔ BTRfs/EXT4

Network Clients (1..N) ➔ READ / WRITE ➔ eCryptfs ➔ BTRfs/EXT4

Keyring / LUKS

1..N encrypted Volumes

Linux Krypto / CryptoAPI

Cavium Crypto Offload driver

CPT Offload Block

DDR
## Accelerations Integrated

<table>
<thead>
<tr>
<th>Application 1</th>
<th>Application 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTRFS+DMDEDUP+RAID6</td>
<td>ECRYPTFS+EXT4+RAID6</td>
</tr>
<tr>
<td>DDF</td>
<td>CPT</td>
</tr>
<tr>
<td>RAD</td>
<td>RAD</td>
</tr>
</tbody>
</table>

| D | D | D | P | Q | D | D | D | D | P | Q | D | D | D | D | P | Q |

[CAVIUM Logo]
RAID offload relative CPU savings

![Graph showing RAID offload relative CPU savings with OCTEON TX-raid-offload(cpu savings) and No. disks/raid6 grp on the x and y axes, respectively.]
Dmdedup offload Ingest rate

OCTEON TX throughput

0GB 40GB 80GB 160GB 200GB

x86_64 OCTEON TX OCTEON TX-offload

5xGain
Dmdedup fio WRITE offload latency

5xReduction
Ecryptfs

The graph shows the performance of OCTEON TX and OCTEON TX-offload across different numbers of threads. The y-axis represents the percentage of CPU used, while the x-axis represents the number of threads.

- OCTEON TX shows a steep increase in CPU usage as the number of threads increases.
- OCTEON TX-offload, on the other hand, shows a more gradual increase in CPU usage.

The graph indicates that OCTEON TX-offload is more efficient, particularly with higher numbers of threads, as it maintains a lower CPU usage compared to OCTEON TX.
Dedup + RAID6, eCryptFS+RAID6

OCTEON TX %cpu idle

OCTEON TX throughput

Fio WRITE

50GB, 4Tx5GB

50GB, 8Tx2GB

100GB, 16Tx2GB

OCTEON TX

OCTEON TX-offload
Status

- Upstream the drivers.
- Other Platforms:
  - DPDK+SPDK
  - ODP
Q & A

Vikas Aggarwal
Vikas.Aggarwal@Cavium.com