ONLINE DISK REENCRYPTION
WITH LUKS2

ONDŘEJ KOZINA <OKOZINA@REDHAT.COM>
Devconf 2019 - Brno
WHY?

- Different data lifetime and algorithm lifetime
**WHY?**

- Different data lifetime and algorithm lifetime
- Cut-off access to data with volume key backup (LUKS header backup)
WHY?

- Different data lifetime and algorithm lifetime
- Cut-off access to data with volume key backup (LUKS header backup)
  - LUKS passphrase change **does not** affect volume key (data encryption key)
WHY?

- Different data lifetime and algorithm lifetime
- Cut-off access to data with volume key backup (LUKS header backup)
  - LUKS passphrase change **does not** affect volume key (data encryption key)
- Volume key change may be enforced by policy
WHY?

- Different data lifetime and algorithm lifetime
- Cut-off access to data with volume key backup (LUKS header backup)
  - LUKS passphrase change does not affect volume key (data encryption key)
- Volume key change may be enforced by policy
- Offline cryptsetup-reencrypt misses few features
WHY?

- Different data lifetime and algorithm lifetime
- Cut-off access to data with volume key backup (LUKS header backup)
  - LUKS passphrase change does not affect volume key (data encryption key)
- Volume key change may be enforced by policy
- Offline cryptsetup-reencrypt misses few features
  - not online
WHY?

- Different data lifetime and algorithm lifetime
- Cut-off access to data with volume key backup (LUKS header backup)
  - LUKS passphrase change **does not** affect volume key (data encryption key)
- Volume key change may be enforced by policy
- Offline cryptsetup-reencrypt misses few features
  - not online
  - not robust enough in case of failure
ONLINE DISK REENCRIPTION WITH LUKS2

ONDŘEJ KOZINA <OKOZINA@REDHAT.COM>
Devconf 2019 - Brno
RESILIENT...?

- Reencrytion writes device in segments (reencrytion zone)
- Crash may produce torn write
- To detect and correct torn write we require:
  - Storage to keep additional information
  - Metadata format capable properly track reencrytion progress
- LUKS2 format is flexible enough for it
BETTER TO BE SAFE...

- Checksums
  - 1 checksum per underlying physical sector size of underlying device
  - Single extra checksum of new ciphertext
  - Stored in keyslots binary area
BETTER TO BE SAFE...

- Checksums
  - 1 checksum per underlying physical sector size of underlying device
  - Single extra checksum of new ciphertext
  - Stored in keyslots binary area
- Journal
**BETTER TO BE SAFE...**

- Checksums
  - 1 checksum per underlying physical sector size of underlying device
  - Single extra checksum of new ciphertext
  - Stored in keyslots binary area
- Journal
- Data shift
BETTER TO BE SAFE...

- Checksums
  - 1 checksum per underlying physical sector size of underlying device
  - Single extra checksum of new ciphertext
  - Stored in keyslots binary area
- Journal
- Data shift
- Noop
  - No syncs
  - No commit points
  - Only gracefully interrupted reencryption is safe
IT CRASHED!

- Checksums
  - Compare content of reencryption zone to stored checksums
  - Reencrypt only sectors with matching checksums
  - Verify new ciphertext segment matches final checksum
IT CRASHED!

- Checksums
  - Compare content of reencryption zone to stored checksums
  - Reencrypt only sectors with matching checksums
  - Verify new ciphertext segment matches final checksum
- Journal
  - Replay
IT CRASHED!

- Checksums
  - Compare content of reencryption zone to stored checksums
  - Reencrypt only sectors with matching checksums
  - Verify new ciphertext segment matches final checksum
- Journal
  - Replay
- Data shift
  - Repeat
IT CRASHED!

- Checksums
  - Compare content of reencryption zone to stored checksums
  - Reencrypt only sectors with matching checksums
  - Verify new ciphertext segment matches final checksum
- Journal
  - Replay
- Data shift
  - Repeat
- Noop
  - :(
ONLINE REENCRYPTION

- Independent layer
- Filesystem/applications access the data via device-mapper stack
- Reencryption process controls access to reencryption zone
mounted filesystem

dm-linear

private overlay device

dm-crypt old key

LUKS2 hdr

old ciphertext
FOR EACH REENCRYPTION ZONE (CHECKSUMS):

1. Read old ciphertext
2. Calculate and store checkums in LUKS2 metadata (commit point)
3. Write new ciphertext
4. Update metadata (commit point)
FOR EACH REENCRIPTION ZONE (CHECKSUMS):

1. Read old ciphertext
2. Calculate and store checkums in LUKS2 metadata (commit point)
3. Write new ciphertext
4. Update metadata (commit point)
FOR EACH REENCRYPTION ZONE (CHECKSUMS):

1. Read old ciphertext
2. Calculate and store checksums in LUKS2 metadata (commit point)
3. Write new ciphertext
4. Update metadata (commit point)
FOR EACH REENCRYPTION ZONE (CHECKSUMS):

1. Read old ciphertext
2. Calculate and store checksums in LUKS2 metadata (commit point)
3. Write new ciphertext
4. Update metadata (commit point)
PERFORMANCE BASELINE:

- sequential **read**
- sequential **write**
- random reads and writes, **randrw** (70% reads, 30% writes)
- fio utility (iodepth=16, 4 threads, block size 32KiB, direct io)

**ROTATIONAL HDD (5400 RPMS, 512 GB)**

<table>
<thead>
<tr>
<th>storage</th>
<th>read</th>
<th>write</th>
<th>randrw</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw</td>
<td>125 MiB/s</td>
<td>94 MiB/s</td>
<td>read 8,6 MiB/s, write 3,6 MiB/s</td>
</tr>
<tr>
<td>dm-crypt</td>
<td>125 MiB/s</td>
<td>65 MiB/s</td>
<td>read 9,3 MiB/s, write 4,0 MiB/s</td>
</tr>
</tbody>
</table>

**NVME (512 GB)**

<table>
<thead>
<tr>
<th>storage</th>
<th>read</th>
<th>write</th>
<th>randrw</th>
</tr>
</thead>
<tbody>
<tr>
<td>raw</td>
<td>2,2 GiB/s</td>
<td>875 MiB/s</td>
<td>862 MiB/s, 369 MiB/s</td>
</tr>
<tr>
<td>dm-crypt</td>
<td>2,2 GiB/s</td>
<td>875 MiB/s</td>
<td>833 MiB/s, 374 MiB/s</td>
</tr>
</tbody>
</table>
## REENCRYPTION PERFORMANCE:

### NVME

<table>
<thead>
<tr>
<th>Reencryption zone size</th>
<th>Resilience mode</th>
<th>Idle (MiB/s)</th>
<th>ETA (min)</th>
<th>Load (MiB/s)</th>
<th>ETA (min)</th>
</tr>
</thead>
<tbody>
<tr>
<td>95 MiB (3.7 MiB)</td>
<td>checksums</td>
<td>206</td>
<td>43</td>
<td>156</td>
<td>56</td>
</tr>
<tr>
<td>198 MiB (7.7 MiBs)</td>
<td>checksums</td>
<td>212</td>
<td>42</td>
<td>175</td>
<td>50</td>
</tr>
<tr>
<td>3.7 MiB</td>
<td>journal</td>
<td>116</td>
<td>1h49</td>
<td>45</td>
<td>3h15</td>
</tr>
<tr>
<td>7.7 MiB</td>
<td>journal</td>
<td>159</td>
<td>55</td>
<td>67</td>
<td>2h11</td>
</tr>
<tr>
<td>20 MiB</td>
<td>noop</td>
<td>400</td>
<td>22</td>
<td>240</td>
<td>37</td>
</tr>
<tr>
<td>95 MiB (detached hdr)</td>
<td>checksums</td>
<td>206</td>
<td>43</td>
<td>156</td>
<td>56</td>
</tr>
<tr>
<td>198 MiB (detached hdr)</td>
<td>checksums</td>
<td>212</td>
<td>42</td>
<td>175</td>
<td>50</td>
</tr>
<tr>
<td>8 MiB</td>
<td>data shift</td>
<td>235</td>
<td>38</td>
<td>97</td>
<td>1h31</td>
</tr>
</tbody>
</table>

ETA: estimated time to reencrypt 512 GiB device
## REENCRYPTION PERFORMANCE:

### ROTATIONAL HDD

<table>
<thead>
<tr>
<th>reencrytion zone size</th>
<th>resilience mode</th>
<th>idle</th>
<th>ETA</th>
<th>load</th>
<th>ETA</th>
</tr>
</thead>
<tbody>
<tr>
<td>95 MiB (3.7 MiB)</td>
<td>checksums</td>
<td>37 MiB/s</td>
<td>4h</td>
<td>17 MiB/s</td>
<td>8h34m</td>
</tr>
<tr>
<td>198 MiB (7.7 MiB)</td>
<td>checksums</td>
<td>40 MiB/s</td>
<td>3h40m</td>
<td>29 MiB/s</td>
<td>5h2m</td>
</tr>
<tr>
<td>3.7 MiB</td>
<td>journal</td>
<td>11 MiB/s</td>
<td>13h15m</td>
<td>4 MiB/s</td>
<td>1d12h</td>
</tr>
<tr>
<td>7.7 MiB</td>
<td>journal</td>
<td>16 MiB/s</td>
<td>9h6m</td>
<td>6 MiB/s</td>
<td>1d</td>
</tr>
<tr>
<td>20 MiB</td>
<td>noop</td>
<td>50 MiB/s</td>
<td>2h55m</td>
<td>25 MiB/s</td>
<td>5h50m</td>
</tr>
<tr>
<td>95 MiB (detached hdr)</td>
<td>checksums</td>
<td>45 MiB/s</td>
<td>3h15m</td>
<td>33 MiB/s</td>
<td>4h25m</td>
</tr>
<tr>
<td>198 MiB (detached hdr)</td>
<td>checksums</td>
<td>45 MiB/s</td>
<td>3h15m</td>
<td>35 MiB/s</td>
<td>4h10m</td>
</tr>
<tr>
<td>8 MiB</td>
<td>data shift</td>
<td>15 MiB/s</td>
<td>9h43m</td>
<td>5 MiB/s</td>
<td>1d5h</td>
</tr>
</tbody>
</table>

*ETA: estimated time to reencrypt 512 GiB device*
SUMMARY

- Reencryption
- Encryption (short offline period)
- Decryption
  - currently detached LUKS2 header only
- Can interrupted, paused, resumed with different parameters
- Recovery performed in `crypt_activate_by_*` calls (cryptsetup open)
DEMO
THANK YOU!

Q&A

- https://gitlab.com/cryptsetup/cryptsetup
- contacts: Ondřej Kozina <okozina@redhat.com>