

Implementing FSCK for Erasure Coded Files in EOS

Mano Ségransan, ID-SD Trainee

Supervisor: Andreas Joachim Peters

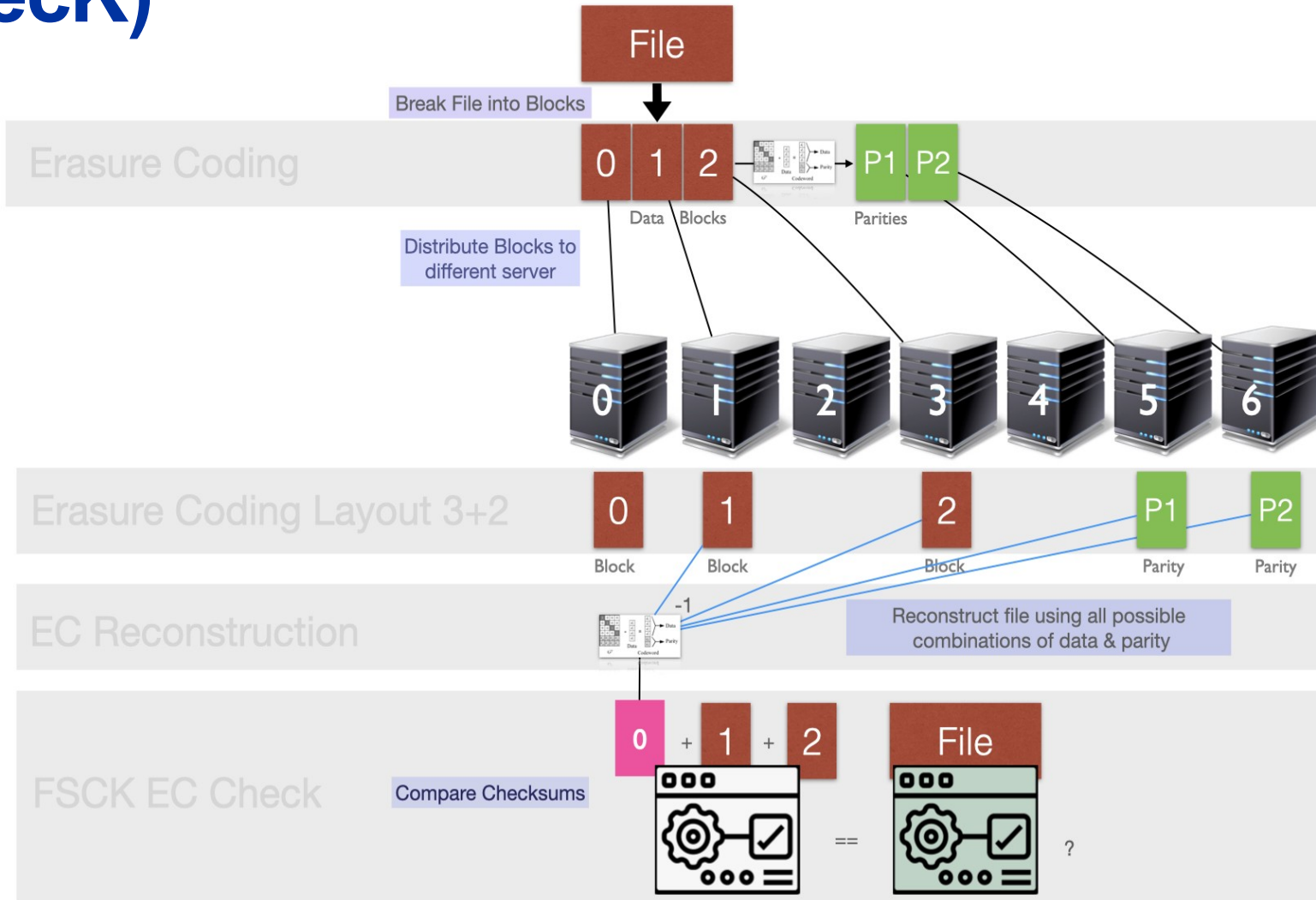
FSCK (File System Check)

- **Context**

- Erasure coded files are stored in blocks, distributed on multiple FSTs
- Each block has a corresponding checksum file

- **Problem**

- FSCK only checks a block against its checksum
- FSCK does not take into account other blocks from the same file
- Over-replicated rain files were not handled



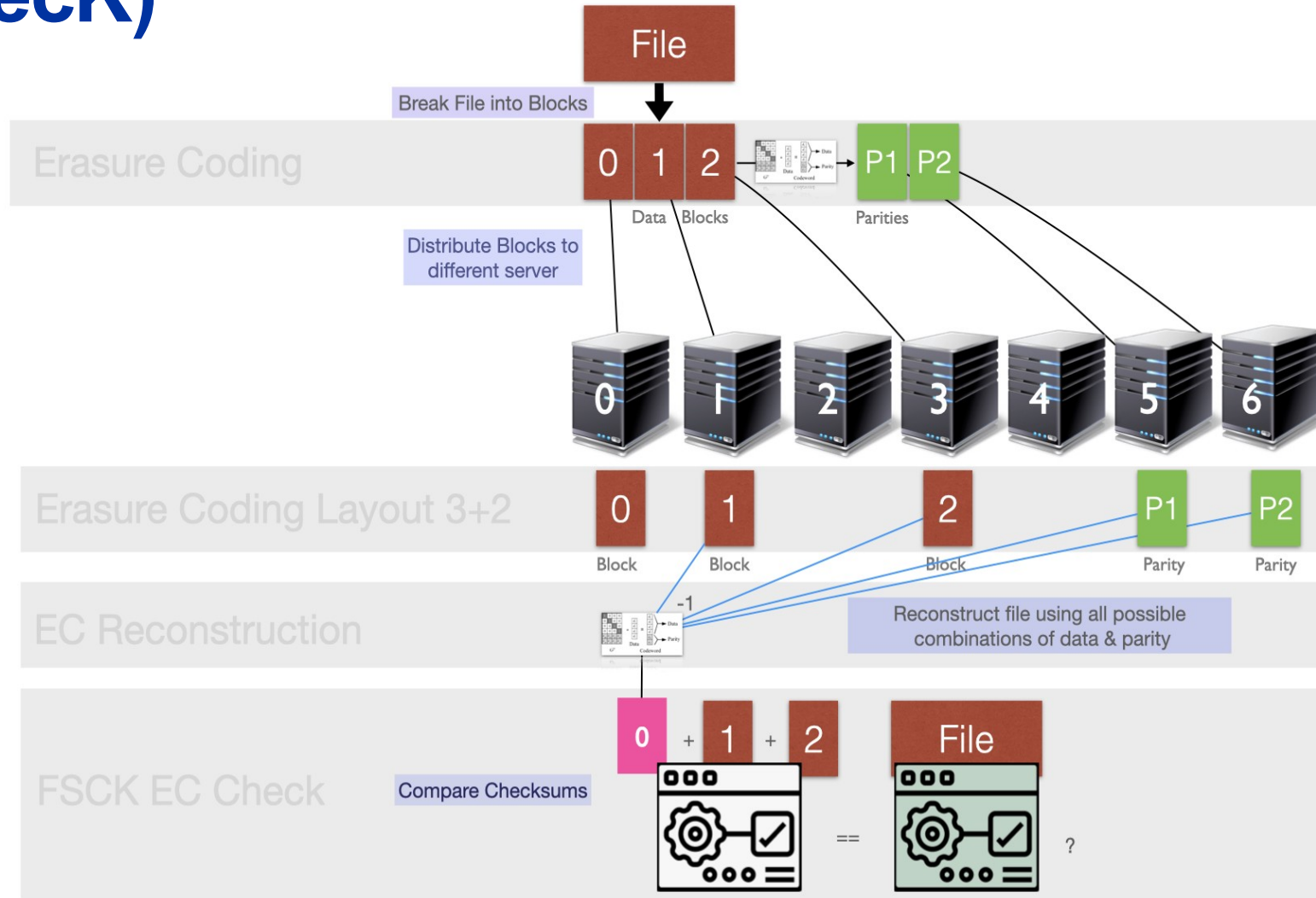
FSCK (File System Check)

- **Issues solved**

- Repair operation was “successful” but stripes were repaired incorrectly
- Checksum of a corrupted data block was recomputed
- General protection against client bugs

- **Solution**

- Created a “raincheck” binary to manually check a file
- Added a monthly FSCK that verifies that each block can be used to reconstruct the original file

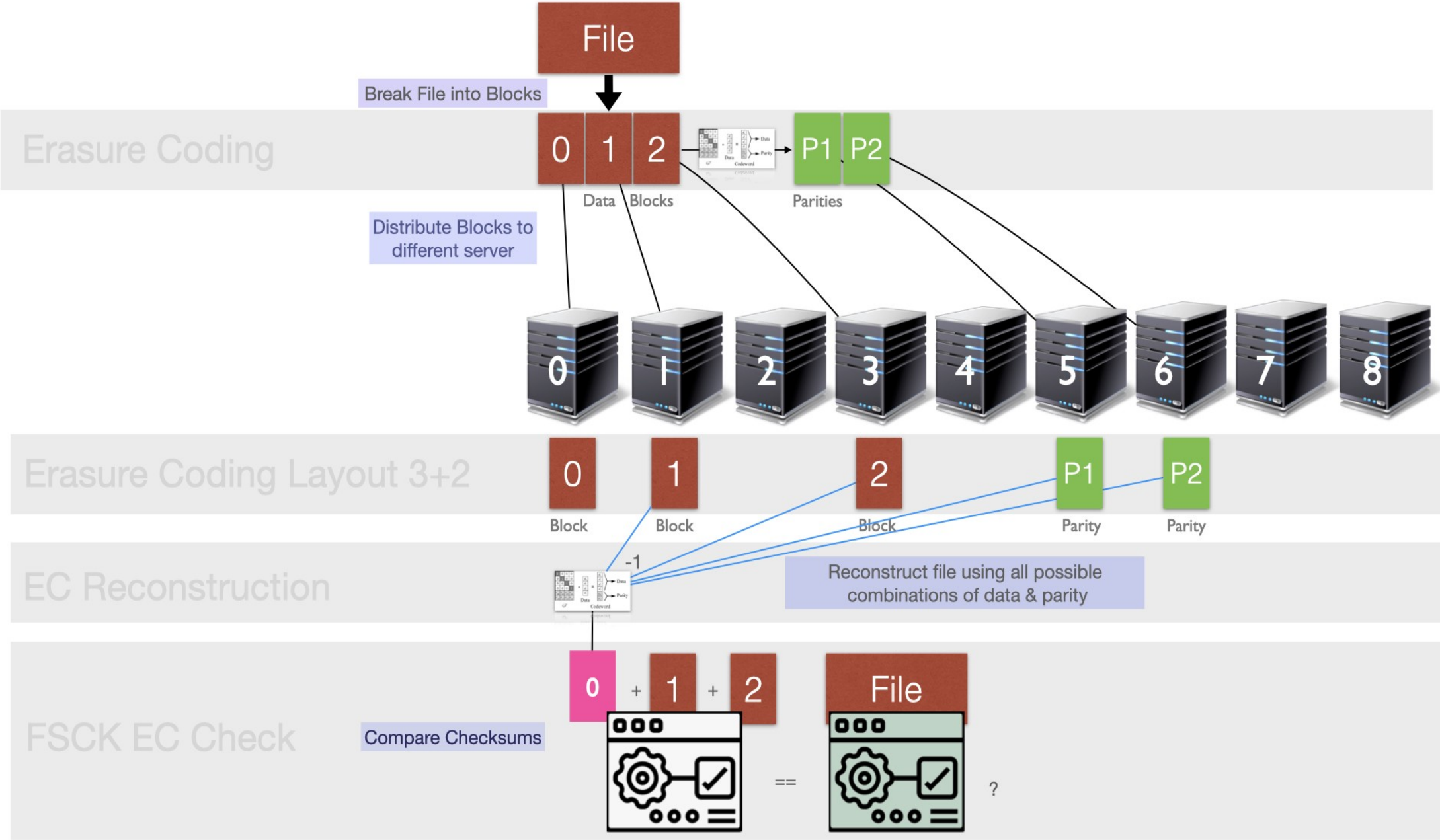


Rain stripe corruption detection algorithm

1. Read the header of each stripe to detect stripes with bad headers, and duplicates
2. Test every stripe combination until we find one that is valid
 - A stripe combination consist the same number of stripes as there is data stripes (Ndata)
 - Read the file using this combination, calculating its checksum
 - Compare the calculated checksum with the original file checksum
3. Test every remaining stripe independently, using the known good stripes
 - Read the file using Ndata – 1 good stripes, and one unknown stripe
 - If the checksum matches, the unkown stripe is valid, otherwise, it is not

Beware: We need to skip combinations with duplicated stripes, as they will always be invalid

4. Generate an FSCK Error for all invalid stripes and duplicated stripes
5. FSCK Repair job will regenerate stripes that were invalid



Conclusion

- This check is expensive, it will only run monthly on files that were modified since the last check
- In most case it will only read the file twice
- Worst case it will need to try every block combination
- Available in EOS 5.2.4



home.cern