A formal analysis of recovery in a preservational data grid

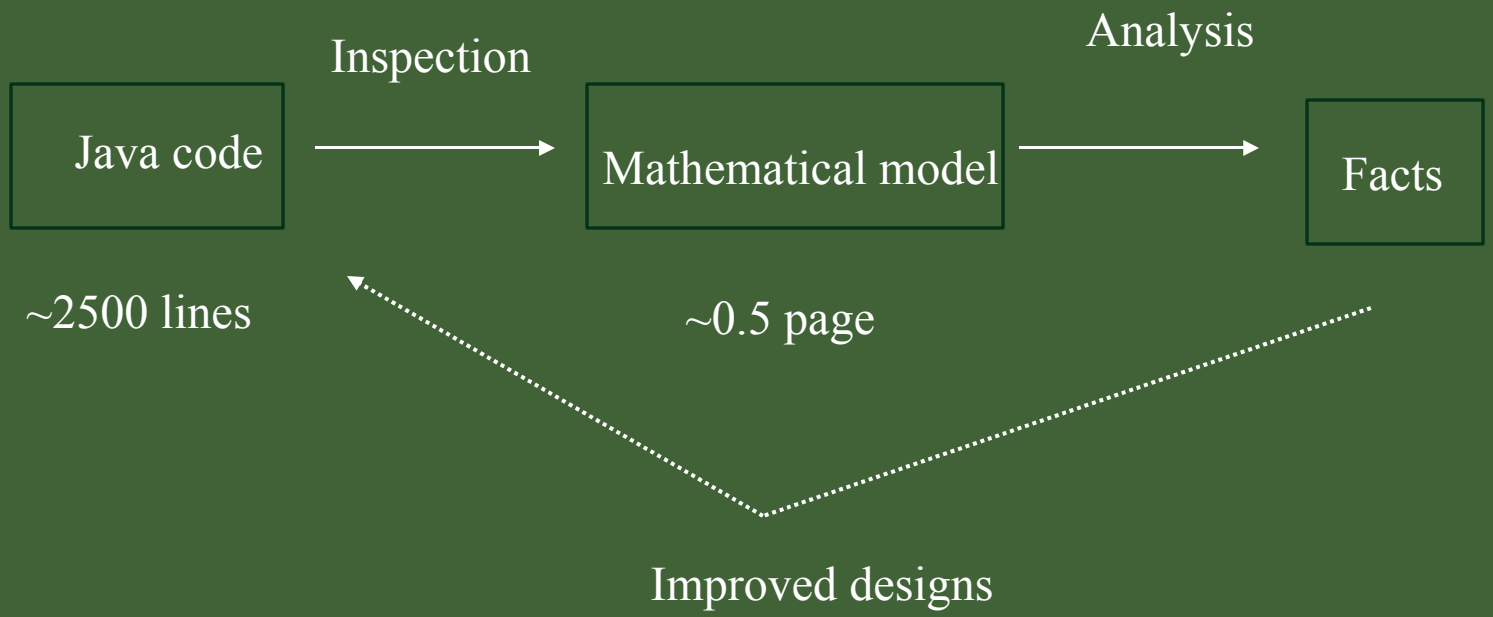
Niels H. Christensen
nhc@kb.dk
Royal Library of Denmark, Dept. of Digital Preservation
&
Netarkivet.dk

The software that has been analysed

- Data grid for long-term archiving of (large) data files.
- 100% Java, developed by Netarkivet.
- Each storage node assigned to one of 2 subgrids.
- Each subgrid receives a copy of every data file.
- The software actively compares copies to detect silent failures, e.g. ”bit rot”.
- The softare also provides operations for recovery after several kinds of failure.

The analysis

Focus: operations for recovery after failures



Example of failure and recovery

Central index

...
file42.arc FCK06DM
...

Subgrid 1

/datadir/file42.arc: 01011001...

Subgrid 2

/datadir/file42.arc: 01011001...

One scenario to be modelled

Example of failure and recovery

Central index

D_{idx}(file42.arc) = {FCK06DM }

Subgrid 1

D_{sg1}(file42.arc) = {FCK06DM }

Subgrid 2

D_{sg2}(file42.arc) = {FCK06DM }

The model

Example of failure and recovery

Central index

D_{idx}(file42.arc) = {FCK06DM }

Subgrid 1

D_{sg1}(file42.arc) = {FCK06DM }

Subgrid 2

D_{sg2}(file42.arc) = {32DR1TV }

Failure!

Example of failure and recovery

Central index

D_{idx}(file42.arc) = {FCK06DM }

Subgrid 1

D_{sg1}(file42.arc) = {FCK06DM }

Subgrid 2

D_{sg2}(file42.arc) = {32DR1TV }

del2(f) can be applied when (for some checksum d):

D_{idx}(f) = {d} and
(D_{sg1}(f) = {d} or |D_{sg1}(f)| != 1) and
D_{sg2}(f) = {d'} where d' != d

Example of failure and recovery

Central index

D_{idx}(file42.arc) = {FCK06DM }

Subgrid 1

D_{sg1}(file42.arc) = {FCK06DM }

Subgrid 2

D_{sg2}(file42.arc) = Ø

Deleted

Example of failure and recovery

Central index

D_{idx}(file42.arc) = {FCK06DM }

Subgrid 1

D_{sg1}(file42.arc) = {FCK06DM }

Subgrid 2

D_{sg2}(file42.arc) = Ø

cp2(f) can be applied when (for some checksum d):

D_{idx}(f) = {d} and
D_{sg1}(f) = {d} and
D_{sg2}(f) = Ø

Example of failure and recovery

Central index

D_{idx}(file42.arc) = {FCK06DM }

Subgrid 1

D_{sg1}(file42.arc) = {FCK06DM }

Subgrid 2

D_{sg2}(file42.arc) = {FCK06DM }

Restored

Conclusions

- More operations and many failure scenarios.
- Long-term preservation – must consider all scenarios.
- The model makes this possible.

- Now we know: what's fixable and what's not.

- Found a bug in the software (scenario where deletion was possible but not the proper action).

- The analysis has led to designs for software improvements.

A formal analysis of recovery in a preservational data grid

- More details in the paper.
- Questions?
- nhc@kb.dk