**Provenance Performance**

Speed

Load – statically present last saved document completely (re)watermarked

This happens too slowly

Karthik. has values across watermarking schemes

See AutoType.xlsx, AutoLoad for values across document size

Problem seems to be with OpenOffice itself

Simple traversal time is too long

See FormatTimer

Could be worse, since Word takes ~~4:35 min.~~ <2 sec. vs. OO 8.5 sec.

Options

Forum says process raw document file

Expensive, fragile, error prone, no timing guarantee

Document still has to be loaded and that is itself slow

See AutoLoad, Load & Save Time

Save and load with watermarks included

Still very slow

See AutoType, Load & Save Time

Need to rethink traitor tracing

Retype document with new watermarks

Insertions are no longer needed, only appends

Compare FormatTimer and AutoLoad

Fragile, error prone, can lose other formatting information

Speed improvement may be minimal

It sped up simple traversal, but was slower on formatting

Don’t watermark the entire document right away

Difficult to track what has and has not been watermarked

Need to break up document into runs anyway

Modify source code to enable bulk setting of styles

Expensive, difficult, no timing guarantee

Find a faster word processor

Very limited free beer choices

Work only with small documents for the time being

Temporary solution

Do not zip files

Capability may not have been implemented in OO

Include fewer watermarks overall

Still need to break up documents into runs

Faster hardware

OK for us, but not for eventual users

Could have some fast server provide the watermarking

Use the AttributeEmbedder instead of StyleEmbedder

There might have been slight improvement

It is difficult to take care of the runs this way.

This saves 30% on longer documents, but needs assistance

Use TextPart, XContentEnumerationAccess,

XMultiPropertySet, XTolerantMultiPropertySet

First two are slower, others won’t help with styles

Type – dynamically add watermarks (and Provenance) as user types

Difficult to automate due to problem with OpenOffice

Cheated by modifying the source code

Would like to stay above ~100 wpm

See AutoType, Watermark Typing Rates

Speed slows significantly as document size increases

Tests are indeed watermarking individual characters plus

inserting a complete watermark elsewhere

Quality

Was waiting for alignment algorithm

See WatermarkBitAnalyzer.xlsx

OCR

Quality (N)

10-100%, whereby 100% is still lossy

OCR works well down to 20%

Resolution (R)

1200 - 100 dpi

No values below 100 dpi. OCR refuses.

Performance deteriorates under 175 dpi at this font size

Watermark bit extraction

Quality

Need to improve algorithm for FgLum, esp. in1out0

Resolution

1200 required overnight run of ~10 hours

Lower resolutions work quickly

Performance deteriorates more slowly than OCR

Substitutions are generally OK

Need to look at specific problem characters like punctuation

Need to write more watermark extractors

Complete watermark

Will depend on how errors cluster