Exact JPEG recompression and forensics using interval arithmetic

Andrew B. Lewis and Markus G. Kuhn
Computer Laboratory Security Group

MM&Sec08 rump session
What is recompression?

Exact recompression is useful because it allows us to

- reduce generative losses,
- characterize tampering subsequent to decompression and
- locate regions of JPEG decompressor output in an uncompressed image.

Exact recompression outputs a set of possible streams rather than a single stream.
In exact recompression, we are given an uncompressed image and work back through the decompression steps keeping track of the set of possible intermediate states, ultimately determining the possible input bit-streams.
We initially consider the colour space conversion and chroma up-sampling operations.
Reversing colour space conversion

During decompression, intermediate values are calculated as a function of those available from the previous decompression step.

Example: YCbCr to RGB colour space conversion, inputs and outputs $I_{x,y}, O_{x,y} \in \{0, \ldots, 255\}^3$ for $x, y$ in the image.

$$O_{x,y} = f(I_{x,y})$$

where $f$ converts a tuple from the stored colour space to a tuple in the output colour space.

In this example, the size of the set of possible outputs is at most $256^3$ elements. To map output tuples back on to inputs, we store an inverted look-up table.
Reversing chroma up-sampling (1)

- Chroma up-sampling cannot be tackled in the same way because the set of possible outputs is huge.
- The up-sampling operation weights contributions from the four closest samples to determine an output value.
Reversing chroma up-sampling (2)

- We represent the computation of each output sample as a function of the inputs and constants, involving addition, multiplication and shifting.
- We store an interval for each unknown as a current estimate, re-arrange the equation and repeatedly update these intervals until we reach a fixed point.
Forensic application (1)

- On images output by the decompressor, the operation converges to a fixed point.
- On images which were output by the decompressor and then tampered in uncompressed form, inconsistencies appear in the equations.
- These are output to an image to reveal the location of tampering.
Forensic application (2)

- Original

- Tampered
Forensic application (3)

- Locations of inconsistencies

- Overlay of inconsistencies with tampered image
Further work

- I am currently extending the recompressor to cover the IDCT step.
- Support other decompressor implementations.
- Can this technique be applied to other types of image/video compression?
- General framework for inverting linear/overdetermined systems of equations involving information loss.
Multimedia forensics bibliography

http://www.cl.cam.ac.uk/~abl26/bibliography/main.html