Shrinking Data for Fun & Profit

Dietmar Hauser
Head of Programming @ Sproing

Why shrink data?

Money!

Reduce bandwidth costs

Initial download / updates

Continuous connections

Expand reach

Decreased loading times

Smaller app size
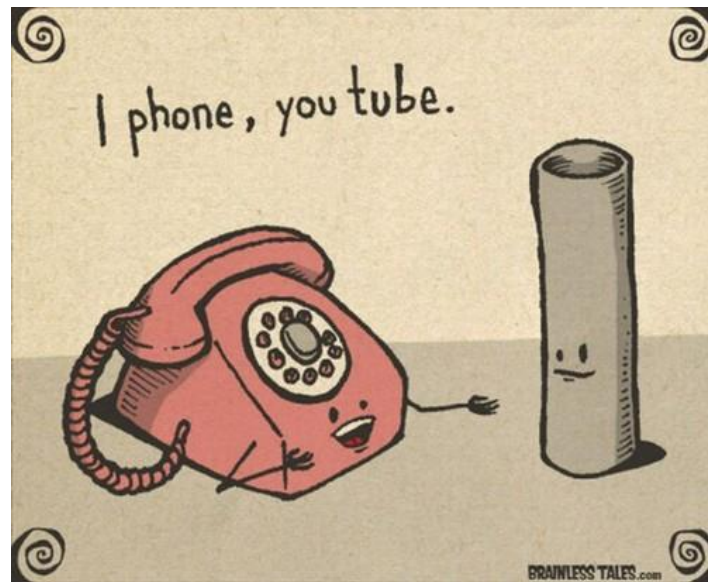
#profit

Why is it fun?

A history of neglect
+
More use cases every day
+
A recent „Gold rush"

Why is now a good time?
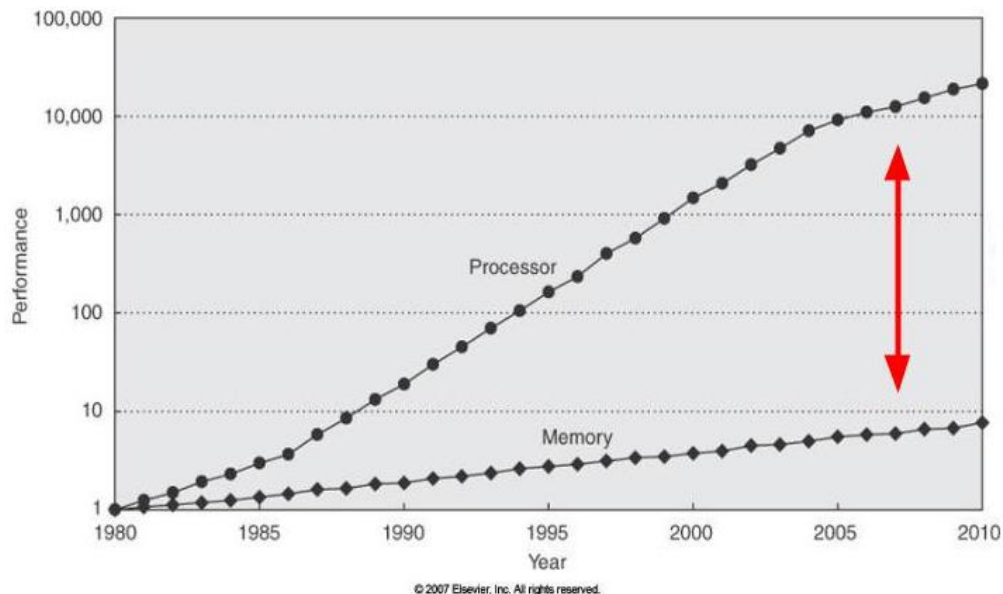
CPU / Memory gap

Fidelity /

   Bandwidth gap

  Hi res displays

     versus

  bandwidth & storage

What is data compression?

Wikipedia:

„[…] encoding information
using fewer bits
than the original representation"

Two flavours of compression

Lossless

    All information is retained

Lossy
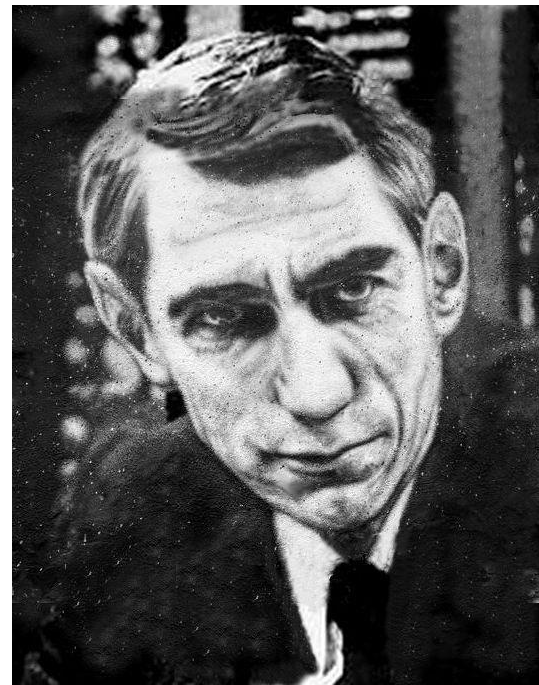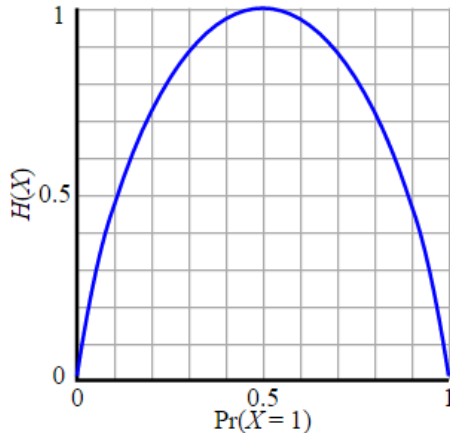
    An approximation is retained

History & Concepts

Information Theory, ~1948

Claude Shannon

Entropy

Shannon limit

History & Concepts

Prefix code, ~1952

    Variable length code

    Translated with a dictionary

    Constructed with Huffman tree

    Fast and efficient

    Still used today

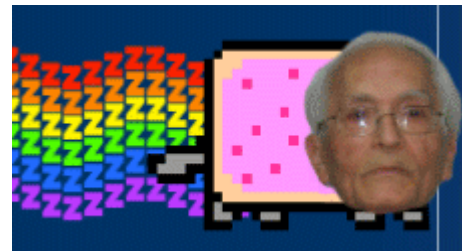| Char	Freq	Code
space	7	111
a	4	010
e	4	000
f	3	1101
h	2	1010
i	2	1000
m	2	0111
n	2	0010
s	2	1011
t	2	0110
l	1	11001
o	1	00110
p	1	10011
r	1	11000
u	1	00111
x	1	10010

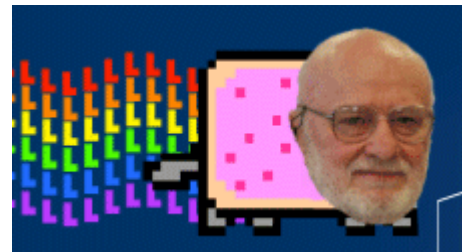# History & Concepts

Lempel-Ziv, 1977

    Base for the LZ-family

    Refers back to already processed data

    „Sliding Window"

    Implicit dictionary creation

History & Concepts

Deflate, 1991

    LZ77 + Huffman

    Used everywhere!

    http://zlib.net

    25 years old!

Compression In Practice 1

Reducing

Network

Traffic

Reducing Network Traffic

HTTP 1.1 has compression built in

    Likely already available to you

    Only GZIP widely supported

     Google is pushing BROTLI

    Hardware support available

    Just turn it on!

Reducing Network Traffic

A closer look at the data

HTML, JSON, XML,… compress well

Human readable → low entropy

Different from data in memory

Conversion wastes CPU / memory

Reducing Network Traffic

Data treatment options

Omit whitespace and comments

Separate static from dynamic data

Transfer static data once (or never)

i.e. replace Strings with IDs

Reducing Network Traffic

Use binary data formats

    i.e. MsgPack, ProtoBuffers, Binary XML,…

    Ditch HTTP, TCP/UDP have less overhead

    If HTTP, consider WebSocket support

    Beware: Base64 re-adds ~25%

Reducing Network Traffic

Faster compression options
  Free: LZ4, Density
  Commercial: LZO, Selkie, LZB16
  Much (!) faster than GZIP
  Lower to equal compression ratio

Reducing Network Traffic
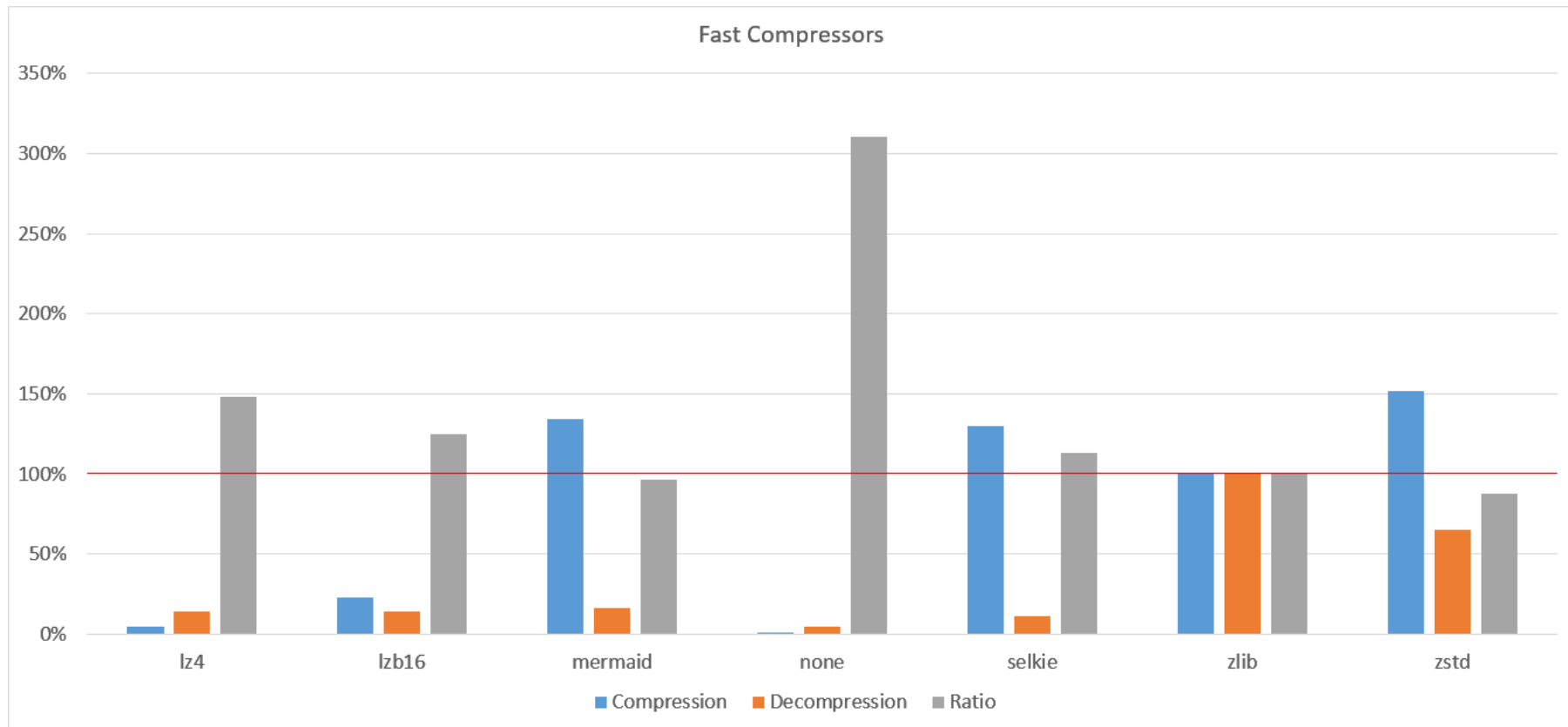
Stronger compression options

    Free: ZStd, BROTLI

    Commercial: Mermaid

    Faster decompression speed

    Slower to equal compression speed

    Equal to higher compression ratio

Fast Compressors

Reducing Network Traffic

General hints

Make compression configurable

If on HTTP, turn HTTP compression off

Encrypt after compression

Beware compressor memory overhead

Make use of streaming, when possible
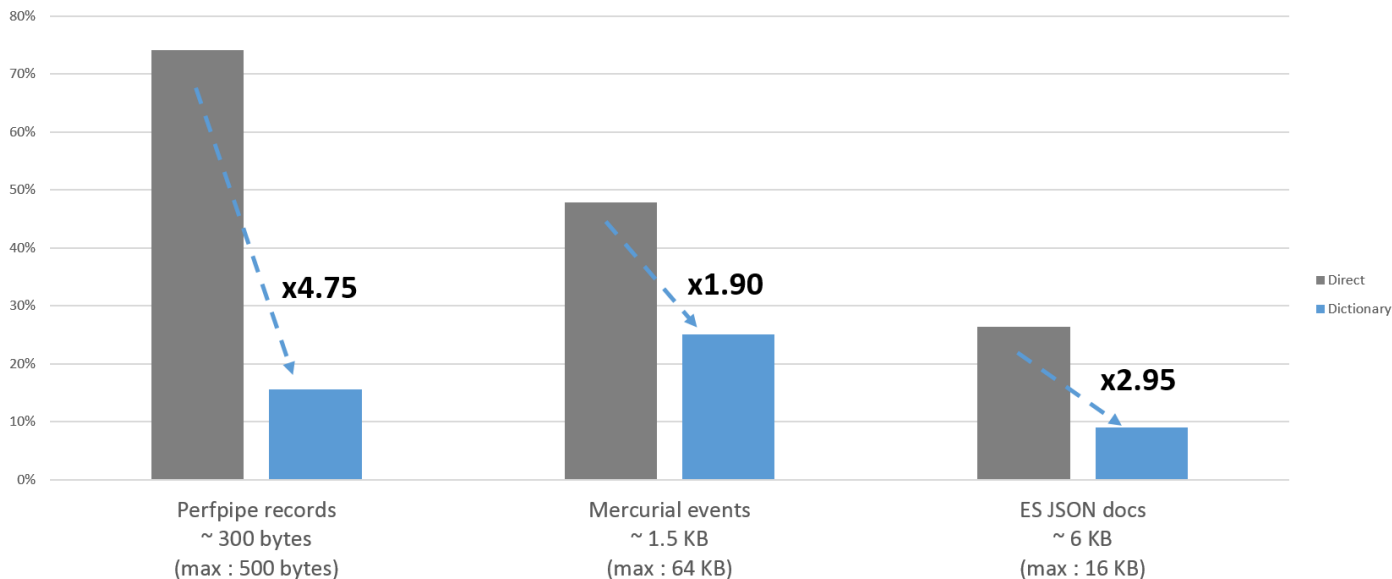
Reducing Network Traffic

Teh Future

HTTP/2 will be a binary protocol

Shared dictionaries

SDCH or home made (i.e. using ZStd)

BROTLI has a generic dictionary built in

Compressing small data : with Dictionary

Source: http://zstd.net

Compression In Practice 2

Shrinking

    Download Size

Shrinking Download Size

Game asset downloads of all kinds

　　HTTP is usually a must (CDN)

　　　HTTP overhead insignificant

　　　HTTP compression not optimal

　　Data is rarely changed

　　→ Use strongest compression available

Shrinking Download Size

Compression Options
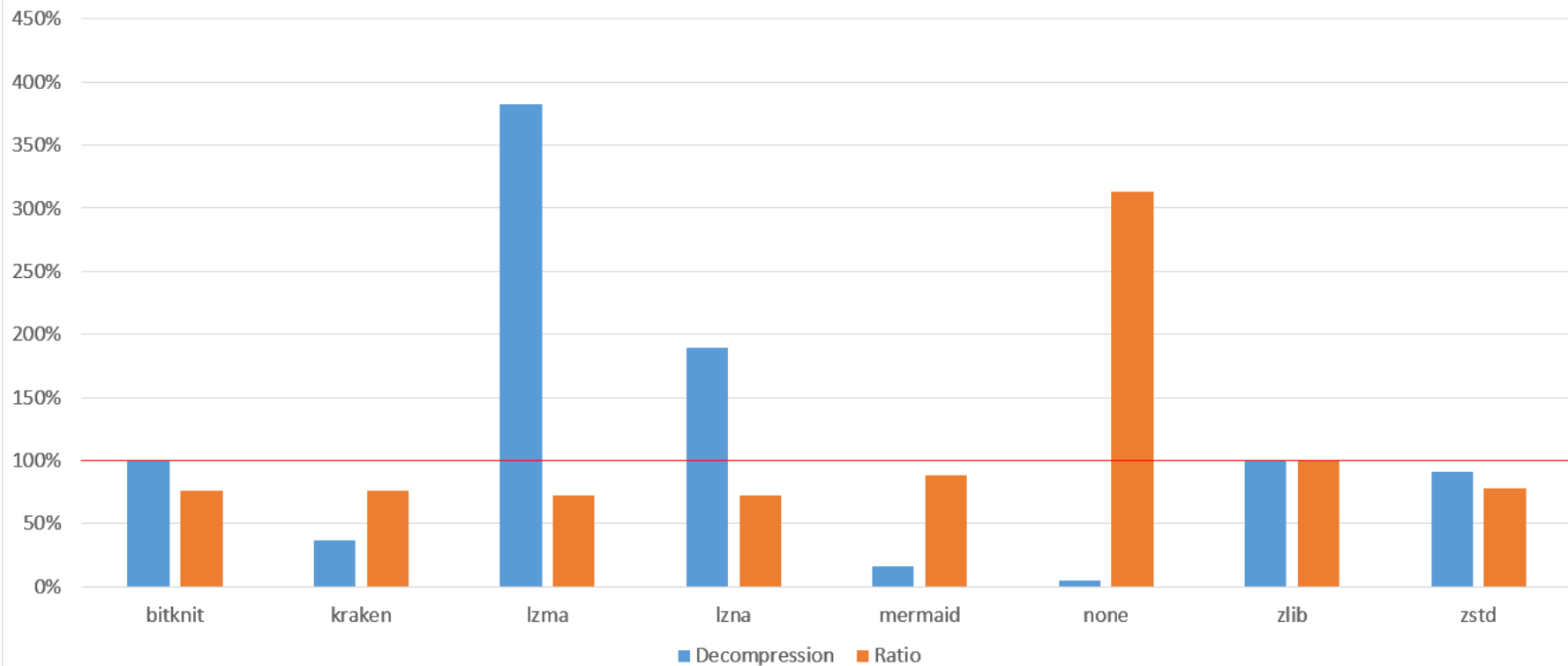   Free: LZMA, XZ, LZHAM
    Commercial: LZNA, Kraken, BitKnit
    Slow to very slow compression
    Very high ratios
    Slow to fast decompression

Strong Compressors

Shrinking Download Size

General Hints

Consider keeping files compressed locally

HTTP request delays and limits

Few big files > many small files

Use parallel downloads, if possible

Don't forget about decompression time

Creating Small App Packages

Why is it different?

Platform owners enforce package format

.apk, .ipa, .appx, …

Actually just .zip files

Built in compression far from optimal

→ Compress before packaging

Creating Small App Packages

Textures

   Best compression: JPEG (or similar)
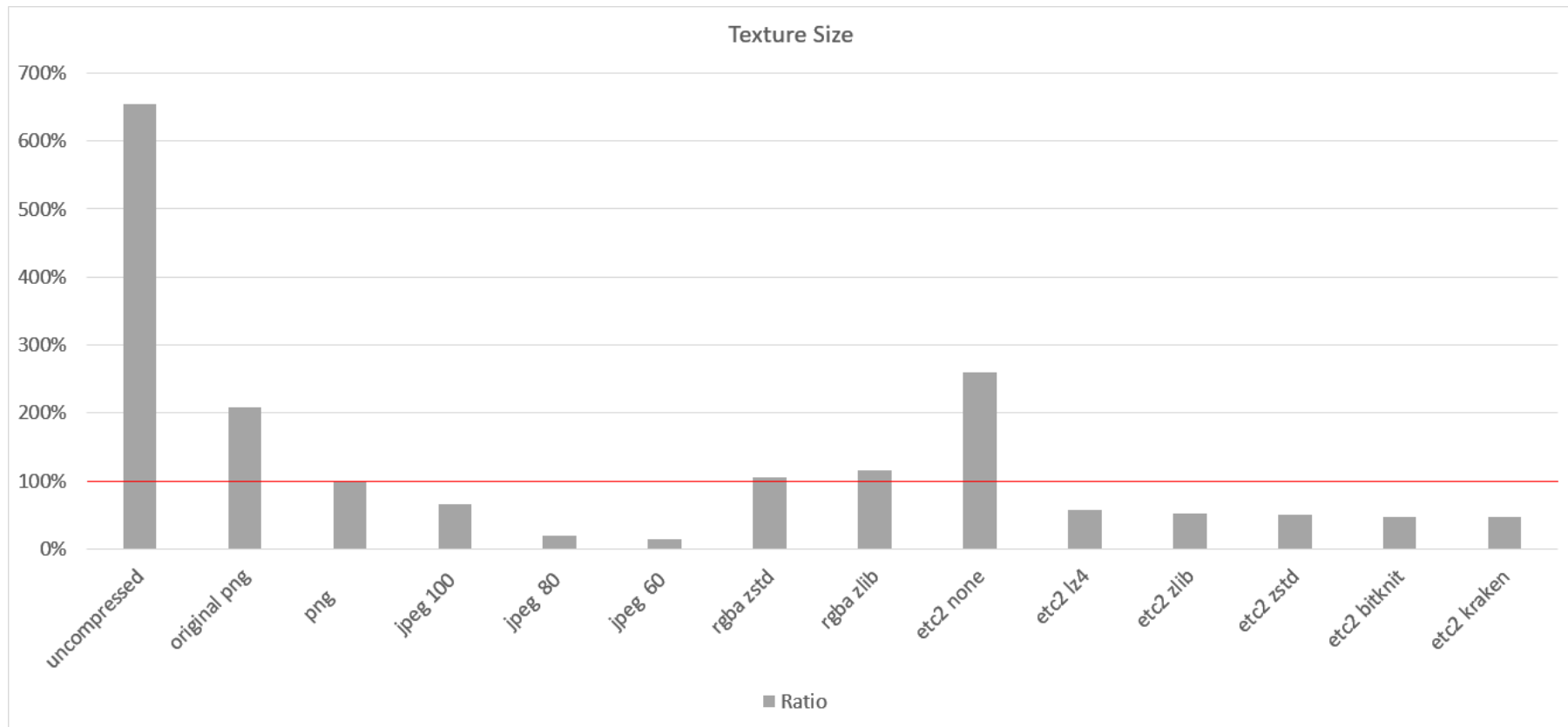
   Most pitfalls: PNG

     Don't use Photoshop output for final images!

   Use compressed texture formats if possible

     Don't forget to apply regular compression

   Consider custom image format

Texture Size

Creating Small App Packages

Textures – The Future

RDO – Rate-distortion optimization

https://github.com/BinomialLLC/crunch

Transcoding between compressed formats

New compressed GPU formats

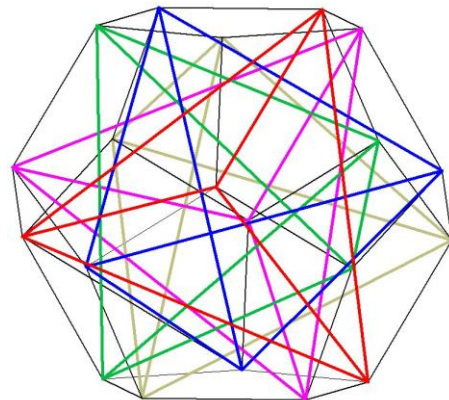# Creating Small App Packages

Geometry & Animation

Highly format dependent

Strip unneeded data

Tangents, Binormals, Extra Uvs,…

Lossy animation compression

Compress using a generic algorithm

Creating Small App Packages

Sound and Music

Use lossy compression

MP3, Ogg/Vorbis, BINKA, …

Depends on audio platform

Check back with provider

Consider mono for music

Ogg Vorbis

Creating Small App Packages

Config, Settings, Loca,…

Use generic algorithm

BROTLI is aimed at text

Consider binary formats

Convert at packaging time

LIVING
LA VIDA LOCA

Creating Small App Packages

Further complications

    Certain files have fixed formats

      App icons, splash screens, …

    Exe is encrypted / signed

      Consider interpreted code

    Lobby platform owners?

Conclusions

Take care of your data from day 1

There is more than Deflate / Zlib

Smaller data makes people happy!

Resources

Yann Collet
  Blog: http://fastcompression.blogspot.com/
  LZ4: http://cyan4973.github.io/lz4/
  ZStd: http://www.zstd.net/
Oodle
  Official: http://www.radgametools.com/oodle.htm
  Charles Bloom: http://cbloomrants.blogspot.com/
  Fabian Giesen: https://fgiesen.wordpress.com/

Resources

BROTLI

  Standard: https://www.ietf.org/rfc/rfc7932.txt

  Source: https://github.com/google/brotli

Misc

  Rich Geldreich (LZHAM): http://richg42.blogspot.com/

  Binomial: http://www.binomial.info/

  LZO: http://www.oberhumer.com/

  7z / LZMA / XZ: http://www.7-zip.org/

  Density: https://github.com/centaurean/density

Contact Information

dietmar.hauser@sproing.com

https://github.com/Rattenhirn

@Rattenhirn

https://www.sproing.com

https://fb.me/sproing

@SproingGames

sproing

one jump ahead