

Delta encoding

Not to be confused with **Elias delta coding**.
Not to be confused with **delta modulation**.

Delta encoding is a way of storing or transmitting data in the form of *differences* (deltas) between sequential data rather than complete files; more generally this is known as **data differencing**. Delta encoding is sometimes called **delta compression**, particularly where archival histories of changes are required (e.g., in **revision control software**).

The differences are recorded in discrete files called “deltas” or “diffs”. In situations where differences are small – for example, the change of a few words in a large document or the change of a few records in a large table – delta encoding greatly reduces data redundancy. Collections of unique deltas are substantially more space-efficient than their non-encoded equivalents.

From a logical point of view the difference between two data values is the information required to obtain one value from the other – see **relative entropy**. The difference between identical values (under some **equivalence**) is often called 0 or the **neutral element**.

1 Simple example

Perhaps the simplest example is storing values of bytes as differences (deltas) between sequential values, rather than the values themselves. So, instead of 2, 4, 6, 9, 7, we would store 2, 2, 2, 3, -2. This reduces the **variance** (range) of the values when neighbor samples are correlated, enabling a lower bit usage for the same data. **IFF 8SVX** sound format applies this encoding to raw sound data before applying compression to it. Unfortunately, not even all 8-bit sound **samples** compress better when delta encoded, and the usability of delta encoding is even smaller for 16-bit and better samples. Therefore, compression algorithms often choose to delta encode only when the compression is better than without. However, in video compression delta frames can considerably reduce frame size, and are used in virtually every video compression **codec**.

2 Definition

A delta can be defined in 2 ways, *symmetric delta* and *directed delta*. A *symmetric delta* can be expressed as:

$$\Delta(v_1, v_2) = (v_1 \setminus v_2) \cup (v_2 \setminus v_1),$$

where v_1 and v_2 represent two successive versions.

A *directed delta*, also called a change, is a sequence of (elementary) change operations which, when applied to one version v_1 , yields another version v_2 (note the correspondence to **transaction logs** in databases).

2.1 Variants

A variation of delta encoding which encodes differences between the **prefixes** or **suffixes** of **strings** is called **incremental encoding**. It is particularly effective for sorted lists with small differences between strings, such as a list of **words** from a **dictionary**.

3 Implementation issues

The nature of the data to be encoded influences the effectiveness of a particular compression algorithm.

Delta encoding performs best when data has small or constant variation; for an unsorted data set, there may be little to no compression possible with this method.

In delta encoded transmission over a network where only a single copy of the file is available at each end of the communication channel, special **error control codes** are used to detect which parts of the file have changed since its previous version. For example, **rsync** uses a rolling checksum algorithm based on Mark Adler’s **adler-32** checksum.

4 Sample C code

The following C code performs a simple form of delta encoding and decoding:

```
void delta_encode(char *buffer, int length) { char last = 0; for (int i = 0; i < length; i++) { char current = buffer[i]; buffer[i] = current - last; last = current; } }
void delta_decode(char *buffer, int length) { char last = 0; for (int i = 0; i < length; i++) { char delta = buffer[i]; buffer[i] = delta + last; last = buffer[i]; } }
```

5 Examples

5.1 Delta encoding in HTTP

Another instance of use of delta encoding is [RFC 3229](#), “Delta encoding in HTTP”, which proposes that [HTTP](#) servers should be able to send updated Web pages in the form of differences between versions (deltas), which should decrease Internet traffic, as most pages change slowly over time, rather than being completely rewritten repeatedly:

This document describes how delta encoding can be supported as a compatible extension to HTTP/1.1.

Many HTTP (Hypertext Transport Protocol) requests cause the retrieval of slightly modified instances of resources for which the client already has a cache entry. Research has shown that such modifying updates are frequent, and that the modifications are typically much smaller than the actual entity. In such cases, HTTP would make more efficient use of network bandwidth if it could transfer a minimal description of the changes, rather than the entire new instance of the resource.

5.2 Online backup

Main article: [Online backup services](#)

Many of the [online backup services](#) adopt this methodology, often known simply as *deltas*, in order to give their users previous versions of the same file from previous backups. This reduces associated costs, not only in the amount of data that has to be stored as differing versions (as the whole of each changed version of a file has to be offered for users to access), but also those costs in the uploading (and sometimes the downloading) of each file that has been updated (by just the smaller delta having to be used, rather than the whole file).

5.3 Git

Main article: [Git \(software\)](#)

The Git source code control system employs delta compression in an auxiliary “[git repack](#)” operation. Objects in the repository that have not yet been delta-compressed (“loose objects”) are compared against a heuristically chosen subset of all other objects, and the common data and differences are concatenated into a “pack file” which is then compressed using conventional methods. In common use cases, where source or data files are changed incrementally between commits, this can result in signifi-

cant space savings. The repack operation is typically performed as part of the “[git gc](#)” process, which is triggered automatically when the numbers of loose objects or pack files exceed configured thresholds.

5.4 VCDIFF

Main article: [VCDIFF](#)

One general format for delta encoding is VCDIFF, described in [RFC 3284](#). Free software implementations include [Xdelta](#) and [open-vcdiff](#).

5.5 GDIFF

Generic Diff Format (GDIFF) is another delta encoding. It is submitted to [W3C](#) in 1997.^[1] In many cases, VCDIFF has better compression rate than GDIFF.

5.6 Diff

Main article: [Diff](#)

Diff is a file comparison program, which is mainly used for text files.

5.7 bsdiff

[Bsdiff](#) is a binary diff program using suffix sorting.

6 See also

- [Data differencing](#)
- [Interleaved deltas](#)
- [Source Code Control System](#)
- [String-to-string correction problem](#)
- [Xdelta](#): open-source delta encoder

7 References

- [1] [Generic Diff Format Specification](#)

8 External links

- [RFC 3229 – Delta Encoding in HTTP](#)

9 Text and image sources, contributors, and licenses

9.1 Text

- **Delta encoding** *Source:* https://en.wikipedia.org/wiki/Delta_encoding?oldid=679437230 *Contributors:* Bryan Derksen, Stevertigo, Michael Hardy, Nikola Smolenski, Dysprosia, Doradus, Furrykef, Omegatron, Robbot, Fredrik, Vikreykja, Alan Liefiting, Acn, David-Cary, Inking, Ofus, AlistairMcMillan, Eequor, KirbyMeister, Wmahan, Mennucc, Karl-Henner, Jim H, Poccil, Cacycle, Jarsyl, Jmah, Cwolfsheep, Ciaran H, MoraSique, AirBa~enwiki, ShadowLord, Rjwilmsi, FlaBot, VKokielov, Rbonvall, Quuxplusone, YurikBot, Russ-Bot, SFEley, NawlinWiki, Mikeblas, PhS, Ieopo, CapitalSasha, Charlesb, Nbarth, Frap, Wickethewok, BasEI, Teks~enwiki, Chris83, Yukoba~enwiki, Ambulnick, Davidhorman, Peck jon, Magioladitis, Tedickey, Schily, Abednigo, Gwern, STBot, Glrx, Abortz, Swaq, Jimthing, Denisarona, Jotterbot, Baudway, Mifter, Addbot, Tsunanet, AgadaUrbanit, Legobot, Yobot, AnomieBOT, Yancheng.chang, HRoestBot, MoreNet, Oliver H, Christoph hausner, Tartaroz, DRAGON BOOSTER, EmausBot, RockMagnetist, EdoBot, Mikhail Ryazanov, Helpful Pixie Bot, Lagoset and Anonymous: 35

9.2 Images

- **File:Mergefrom.svg** *Source:* <https://upload.wikimedia.org/wikipedia/commons/0/0f/Mergefrom.svg> *License:* Public domain *Contributors:* ? *Original artist:* ?
- **File:Revision_controlled_project_visualization-2010-24-02.svg** *Source:* https://upload.wikimedia.org/wikipedia/commons/a/af/Revision_controlled_project_visualization-2010-24-02.svg *License:* CC-BY-SA-3.0 *Contributors:*
 - *Revision_controlled_project_visualization.svg* *Original artist:* *Revision_controlled_project_visualization.svg*: *Subversion_project_visualization.svg: Traced by User:Stannered, original by en:User:Sami Kerola
- **File:Symbol_template_class.svg** *Source:* https://upload.wikimedia.org/wikipedia/en/5/5c/Symbol_template_class.svg *License:* Public domain *Contributors:* ? *Original artist:* ?

9.3 Content license

- Creative Commons Attribution-Share Alike 3.0