

# Alternate Means of “Data Compression”

## Part I: Overview

By Thomas O’Hare  
Thomas@RedTile.Com

*Copyright© 2004, Thomas O’Hare – All Rights Reserved.*

### **Introduction**

“Data Compression” in very general terms means reducing the size of data to a fraction of its original size so it can be handled with more efficiency and is easier to interface with. There are numerous ways to physically “compress” data, but there are also alternate ways to decrease the physical size of data. This paper will deal with alternate means of compression, or the limiting of data to a more manageable size.

In any collection of data there are at least two constants; Time and Change. Time is a quantity that is totally dependent on the type of data being collected; using measurements such as hours, minutes or seconds. Change (which we will call “Delta”) is when data moves beyond its present value to a new or different value.

Both of these constants are fully dependent upon the type of data being collected. As an example, monitoring of jet engines in prototyping requires a constant stream of data for a constant amount of time and every change that occurs no matter how small. This is called critical or “mission critical” data collection as jet engines must be 100% proficient. Any variation of data, even for a millisecond, must be noted and logged. The airframe the jet engine is connected to is not as mission critical as an engine. It is a more “static” environment so data collection can be only “samplings” of data that change over time.

The premise of this paper is not to show mathematical algorithms of creating tokens to represent data in a compressed state, but rather to control or throttle the amount of data being collected and then how to effectively use that data “subset”.

Since this paper is dealing mainly with raw Data, Time and Delta, the data can remain in a more “human readable” form as when it was captured or stored.

Both the math compression scenario and the alternate methods described here require a processing or manipulating of data. No matter what kind of compression schema you eventually decide to use there is always overhead associated with it.

### **Time**

Time as a quantity is dependent upon the object which the data is being collected for. In other words, time segments are set by the environment in which the data collection takes place. Obviously dependent on the type of data being collected, the length or quantity of time slices used will then be chosen.

*Copyright© 2004, Thomas O’Hare – All Rights Reserved.*

In this scenario, only changes in a quantity of time produce a capture of data for that particular moment. The use of Time as a data capture method does not require any knowledge of the value or Delta being recorded. Data capture is purely on a preset timed interval.

### **Delta**

Delta in scientific terms means “change”, or when a known quantity changes in value. In this alternate method of data compression the amount of change, or Delta, is also dependent upon the type of object that data is being collected for. In this method, defined limits of known values and changes at or beyond these limits are then recorded. When this change occurs, the apparatus doing the data capture must then adjust to the new value and set or reset new limits dependent on the newest value recorded.

Delta can also be a singular method of data collection. In this pure scenario Time is not a dependent value, only Delta. This means only occurrences of Delta are recorded and these recordings are not influenced in any way by Time.

### **Time and Delta**

The most secure or robust way of capturing data is by combining the Time and Delta methods. In this approach you will always have a second parameter to rely on in the data capture mechanism. You will maintain a better scope of what is going on because you can compare Time and Delta data captures against each other. This combination of Time and Delta always gives you a better point of reference because you can more easily cross reference both Time and Delta values to derive a particular or desired data point.

In effect, by using a Time and Delta approach you can create an “X” and “Y” axis graph point approach. The “X” axis is a time line and the “Y” axis is the plus and minus limits of a current value. By plotting the X and Y values (Time and Delta) you maintain a very reliable method of capturing data for detailed data analysis later on.

The “compression” used in this case is caused by *not* capturing un-needed data. The point is; “why capture data that remains at the same value during a specified period of time?” We already know what that value is and we already know what the time intervals are. We can verify that value, especially with the X axis (Time), as “insurance” in case for any reason there is a problem with monitoring the actual Deltas.

Collecting data using these Time and Delta capture techniques, we have in effect eliminated much of the data “bloat” and reduced the amount of processing needed for later data analyses.

*Change Log Part I:*

Original Release: March 15, 2004.