Automatic Format Generation Techniques For Network Data Acquisition Systems

Benjamin Kupferschmidt
Technical Manager - TTCWare
Teletronics Technology Corporation

Eric Pesciotta
Director of Systems Software
Teletronics Technology Corporation

ABSTRACT

Configuring a modern, high-performance data acquisition system is typically a very time-consuming and complex process. Any enhancement to the data acquisition setup software that can reduce the amount of time needed to configure the system is extremely useful. Automatic format generation is one of the most useful enhancements to a data acquisition setup application. By using Automatic Format Generation, an instrumentation engineer can significantly reduce the amount of time that is spent configuring the system while simultaneously gaining much greater flexibility in creating sampling formats.

This paper discusses several techniques that can be used to generate sampling formats automatically while making highly efficient use of the system's bandwidth. This allows the user to obtain most of the benefits of a hand-tuned, manually created format without spending excessive time creating it. One of the primary techniques that this paper discusses is an enhancement to the commonly used power-of-two rule, for selecting sampling rates. This allows the system to create formats that use a wider variety of rates. The system is also able to handle groups of related measurements that must follow each other sequentially in the sampling format.

This paper will also cover a packet based formatting scheme that organizes measurements based on common sampling rates. Each packet contains a set of measurements that are sampled at a particular rate. A key benefit of using an automatic format generation system with this format is the optimization of sampling rates that are used to achieve the best possible match for each measurement's desired sampling rate.

KEYWORDS

Automatic Format Generation, Networked Data Acquisition, IENA

INTRODUCTION

Creating a PCM Format is one of the most complicated and error prone aspects of configuring a high-performance data acquisition system. Data acquisition systems that are designed to collect data from a wide variety of sources typically require a sampling format that specifies the exact sequence of measurements that the system collects. By running this format at a particular bit rate, a desired sampling rate can be achieved for each measurement. The deterministic nature of PCM sampling formats greatly simplifies the analysis of the data that is collected by the data acquisition system.

Proper sampling format creation depends on the instrumentation engineer's ability to satisfy a large set of requirements for each measurement. The instrumentation engineer receives a set of requirements that specify the rate of the measurement and any other special restrictions such as periodicity. The format must be laid out so that each measurement is sampled at or slightly above the rate that is required. The engineer must satisfy all of the system's rules and other requirements such as limited transmission bandwidth. Some of the requirements may even be contradictory which will require the engineer to make tradeoffs in the design of the format. In order to do this, the instrumentation engineer must determine which requirements are more important and must be satisfied completely. The other less important requirements can be partially satisfied.

Automatic format generation is a technology that simplifies the process of creating a PCM sampling format. This allows instrumentation engineers to configure data acquisition systems more rapidly. Instead of manually handling all of the measurement and system requirements, the engineer only has to specify the desired sampling rate and restrictions for each measurement. The format creation software can then use this information to generate a sampling format automatically.

The principle benefit of automatic format generation is analogous to the benefit of a programming language's compiler. A compiler allows a programmer to focus on the logic of their program rather than the implementation details of the computer's assembly language. Similarly, automatic format generation allows the instrumentation engineer to focus on collecting measurement data without worrying about lower level implementation details. While there are situations where engineers can write better assembly code or sampling formats then a computer, the time savings from using automatic format generation more than makes up for the slight decrease in efficiency.

This paper discusses a number of issues involved in creating a highly efficient automatic format generator. It suggests several solutions to these common issues. In addition, this paper shows how automatic format generation techniques can be applied to non-PCM data acquisition systems. A common example of a non-PCM system where automatic format generation can be applied is one that groups data by sample rate and outputs Ethernet packets. Each packet consists of a set of measurements that are sampled at a particular rate. Since each measurement in a packet is sampled at the same rate, the packet is much easier to analyze than a conventional PCM Format that contains measurements sampled at different rates.

GETTING STARTED WITH AUTOMATIC FORMAT GENERATION

In order to discuss some of the more complex issues involved in the creation of an automatic format generator, we must describe the basic operation of a format generator. The most common method that is used for automatic format generation is called the power-of-two rule. This rule restricts the set of available sampling rates for measurements to the set of rates that are powers of two. For example, some of the supported rates are 1, 2, 4, 8 and 16 samples per second. Each of these rates can be represented by the formula 2^n where n is any positive integer. The key advantage of applying this rule is that it makes it much easier to lay out the measurements in the PCM format. If a measurement needs to appear in a format one time in order to achieve one sample per second, then the measurement needs to appear twice for two samples per second and four times for four samples per second. If we setup the format so that it contains a power-of-two number of total words then it is very simple to place the individual measurements in the format to achieve the power-of-two sampling rates.

There are several key steps to the process of automatically generating a format. The first step is to gather all of the information about the measurements that need to be sampled. For each measurement, the automatic format generator needs to know the user's desired sampling rate. The system must also know whether the measurement needs to be sampled periodically. The instrumentation engineer must also tell the system about any special restrictions and limitations that apply to the measurement. These restrictions may exist because of the design of the data acquisition card that actually collects the data. The system must also know the length of each data word that is acquired so that it can create extended or residual read parameters automatically.

Once the instrumentation engineer specifies all of the sampling rates, the software can determine the best sampling rates to use in the actual format based on the power-of-two rule. In the next step, the software adds the measurements to the sampling format at the proper intervals. Finally, the software sets the format's bitrate so that all of the measurements are sampled at the proper rates.

PROBLEMS AND LIMITATIONS OF THE POWER-OF-2 RULE

The basic implementation of the power-of-two rule that is described above has several serious limitations. These problems can be solved by making the automatic format generation system slightly more sophisticated. Historically, formats have been created manually because instrumentation engineers want to achieve a precise set of sampling rates while also following the system's rules and satisfying other requirements. These limitations have to be addressed because they prevent the automatic format generation system from achieving the same performance that an instrumentation engineer can achieve with a manually created format.

One of the most serious limitations that a simple automatic format generator suffers from is the need to use power of two rates. This can become a problem for higher rate measurements because the gap between the allowable sampling rates can grow quite large. For example, if I have a parameter that needs to be sampled at 9000 samples per second, it is very inefficient to have to sample the parameter at 16384 samples per second as required by the power-of-two rule.

This problem also becomes readily apparent for common sets of sampling rates that are not based on powers-of-two. For example, many projects have measurements that need to be sampled at 10, 100, 1000 and 10,000 samples per second. This set of rates is not easily supported without significant oversampling of either the high or low rate measurements.

In order to achieve power-of-two sampling rates, some measurements will typically need to be over-sampled. Deciding which measurements can be oversampled and which measurements must be sampled at exactly the specified rate is one of the more complex aspects of automatic format generation. For some types of data, oversampling is harmless and the extra samples do not affect the analysis of the data. Other types of data have stricter requirements and there can be problems if the data is oversampled. Another complication is that most data cannot be under-sampled without losing critical information. Some types of data are stored internally in FIFOs and can compensate for over-sampling by inserting fill words. However, FIFO based data cannot be under-sampled without losing data.

Although some oversampling is inevitable with an automatically generated format, a key goal is minimizing the amount of oversampling. Oversampled data results in wasted transmission bandwidth and/or recorder disk space. It is also important that the sampling format have a minimal amount of unused data words. The goal is to create as small a sampling format as possible while still achieving the desired sampling rates for all measurements.

Finally, the automatic format generator must take in to account the fact that some measurements need to be sampled periodically in order to facilitate time based data analysis. These measurements must be placed into the sampling format at the proper intervals so that they are sampled periodically. Other measurements like video or bus data do not have this restriction.

ENHANCING THE POWER-OF-2 RULE

There are several ways to enhance the power-of-two rule so that it produces better sampling formats. By improving the sampling formats, the amount of oversampling can be reduced and the format will be much more efficient. The first technique for improving the power-of-two rule is to add scaling. Essentially this involves multiplying the power-of-two sampling rates by a constant scaling factor. This retains all of the benefits of having power-of-two sampling rates while also significantly reducing the amount of oversampling that is required. By scaling the sampling rates, it is possible to select a set of sampling rates that are much closer to the desired sampling rates that are specified by the instrumentation engineers.

In order to scale the sampling rates, it is necessary to determine the best scaling factor. The best way to do this is to analyze the set of desired sampling rates and select one of the rates to be the base for the power-of-two rule. This base rate is sampled at exactly the desired rate while all of the other sampling rates are raised so that they have a power-of-two relationship with the base rate. An easy way to determine the base rate is to use a brute force approach. This method involves taking each unique desired sampling rate and calculating the amount of oversampling that would be required if that rate was the base rate. By considering the entire set of possible base sampling rates, it is straightforward to select the best rate to use. The calculation can also consider the fact that some rates are much more common than other rates. Therefore, if most measurements have similar desired samplings rates and only a few have significantly higher or lower desired sampling rates then the system will choose a base rate that satisfies the majority of the desired rates. The few outliers will end up being oversampled more than the average measurements but this is acceptable because the average measurement is not oversampled very much.

The best way to see the scaled power-of-two approach in action is to consider a simple example. Suppose that we want to create a sampling format for eight measurements with the following desired sampling rates: 3, 6, 10, 12, 15, 30, 30, and 60 samples per second. First, we will consider the standard power-of-two rule. The first seven power-of-two rates are 1, 2, 4, 8, 16, 32, and 64 samples per second. We can use oversampling to map the desired rates on to the set of power-of-two rates. If we then use these sampling rates to create a format, the format will oversample the measurements by approximately 23%.

In order to apply the scaled power-of-two rule to our example, we need to determine the best base sampling rate. This can be done by computing the oversampling percentage for each of the seven distinct rates. The result of this calculation is that 30 samples per second is the best base rate because it is used twice and the rates 15 samples per second and 60 samples per second will be satisfied exactly. The only oversampling will occur on the four measurements that are sampled at lower rates. By using the scaled power-of-two rule, the oversampling percentage can be reduced to 15.6%.

In order to improve the efficiency of the automatic format generator further, we need to devise a mechanism that allows a larger set of possible sampling rates. This will allow us to satisfy the desired sampling rates for a greater percentage of the measurements. One approach that achieves this goal is the factorized power-of-two rule. This rule takes advantage of the fact that some integers have a much larger set of factors than others do. If we use one of these integers as our base sampling rate then we can reduce the amount of oversampling significantly. Crucially, this increase in the number of allowable sampling rates does not interfere with our ability to sample all of the measurements periodically.

The key to making the factorized power-of-two rule work is the fact that the entire sampling format does not need to use the same set of sampling rates. Each subcom in the format can use its own set of sampling rates that are based on the power-of-two rule. As long as the sampling rates for each subcom are all factors of the same base rate, it will still be possible to lay out the format in a manner that samples all of the measurements periodically.

	Word 1	Word 2	Word 3	Word 4	Word 5	Word 6	Word 7
Frame 1	SFID (0000)	**60 SPS**	30 SPS	20 SPS	12 SPS	2 SPS	
Frame 2	SFID (0001)	**60 SPS**	15 SPS	10 SPS	6 SPS	1 SPS	
Frame 3	SFID (0002)	**60 SPS**	30 SPS		3 SPS		
Frame 4	SFID (0003)	**60 SPS**		20 SPS			
Frame 5	SFID (0004)	**60 SPS**	30 SPS				
Frame 6	SFID (0005)	**60 SPS**	15 SPS		12 SPS		
Frame 7	SFID (0006)	**60 SPS**	30 SPS	20 SPS			
Frame 8	SFID (0007)	**60 SPS**		10 SPS			
Frame 9	SFID (0008)	**60 SPS**	30 SPS				
Frame 10	SFID (0009)	**60 SPS**	15 SPS	20 SPS			
Frame 11	SFID (000A)	**60 SPS**	30 SPS		12 SPS		
Frame 12	SFID (000B)	**60 SPS**			6 SPS		
Frame 13	SFID (000C)	**60 SPS**	30 SPS	20 SPS			

Figure 1: Using 60 samples per second as the shared base sampling rate

Figure 1 shows the result of using the base rate of 60 samples per second for the factorized power-of-two rule. In this example, the sampling format has 60 minor frames per major frame. Word 2 contains a main frame word that is sampled in every minor frame resulting in 60 samples per second. The measurements 15 SPS and 30 SPS are sampled in word three. In word four, the measurements 20 SPS and 10 SPS are sampled. All of these sampling rates are divisors of the base rate of 60 samples per second. This means that by varying the frame increments that are used in each subcom, it is possible to sample measurements at all of these sampling rates periodically. If we apply the factorized power-of-two rule to the example that we discussed above then it is possible to achieve a perfect match for the desired sampling rates.

Desired Sampling Rates	Power-of-two Rule	Scaled Power-of-two Rule	Factorized Power-of-two Rule
3 SPS	4 SPS	3.75 SPS	3 SPS (Exact)
6 SPS	8 SPS	7.5 SPS	4 SPS (Exact)
10 SPS	16 SPS	15 SPS	6 SPS (Exact)
12 SPS	16 SPS	15 SPS	10 SPS (Exact)
15 SPS	16 SPS	15 SPS	12 SPS (Exact)
30 SPS	32 SPS	30 SPS (Exact)	30 SPS (Exact)
30 SPS	32 SPS	30 SPS (Exact)	30 SPS (Exact)
60 SPS	64 SPS	60 SPS (Exact)	60 SPS (Exact)
Oversampling Percentage	23.3%	15.6%	0%

Figure 2: Sampling Rate Example

The factorized power-of-two rule can also be used to solve the problem of automatically generating sampling formats for sets of sampling rates that are not related by powers-of-two. For example, a set of measurements that are sampled at the rates 1, 10, 100, 1000 and 10000 samples per second cannot be arranged by the power-of-two rule without a significant amount of oversampling. If we apply the scaled power-of-two rule, we can reduce but not eliminate the oversampling. However, the factorized power-of-two rule makes it possible to sample measurements at any of these five rates in the same format.

AUTOMATICALLY PLACING MEASUREMENTS IN THE FORMAT

The basic approach to laying out the sampling format that is used in automatic format generation is to place the highest rate measurements first and then to fill in the gaps with the lower rate measurements. This method ensures that there is enough room in the format for the highest rate measurements. If required, this also allows the highest rate measurements to be placed into the format periodically. Generally, all measurements that need to be sampled periodically should be laid out first and then non-periodic measurements should be added.

This method of laying out sampling formats can be extended for groups of related measurements. This is needed to accommodate measurements that are longer then the common word length. Typically, extended read measurements must be placed immediately after the original measurement in order to read the residual bits. The automatic format generator can handle these measurements by adjusting the order that the measurements are added to the format. Measurements that require more than one word per sample need to be added before measurements that only need one sample. This feature can also be useful for sampling bus data. All of the data words from a bus message should be sampled sequentially in order to guarantee time coherency and make the message easier to interpret during data analysis.

In addition to sampling rate and periodicity requirements, the automatic format generator must take other system restrictions and requirements into account. An example of these restrictions is the fact that some data acquisition cards require a minimum amount of time between consecutive samples. While this time interval is typically quite small it is still an issue that the must be checked when laying out measurements in a sampling format.

One issue that applies mainly to network-based data acquisition systems is minor frame latency. The automatic format generator needs to consider this when generating sampling formats for network-based systems. A network based PCM system encapsulates each minor frame of a sampling format as a network packet. Typically, users will want to ensure that the data in packets is no more than 50 to 100 ms old before it is transmitted over the network. To achieve this goal, the total time needed for capturing a minor frame must be kept to a minimum. The automatic format generator can accommodate this restriction by emitting more minor frames and placing less data in each minor frame.

NON-PCM APPLICATIONS OF AUTOMATIC FORMAT GENERATION

Automatic Format Generation has applications beyond the realm of standard PCM based data acquisition systems. Some network-based systems have non-PCM based sampling schemes. One example of a non-PCM based sampling system is the Instrumentation d'Essais des Nouveaux Avions (IENA) system used by Airbus on the A380 program. This sampling system generates a set of network packets containing data acquisition samples. Each packet contains data for one of several keys. A key is defined as containing a repeating pattern of one or more measurements. The key's data consists of the repeating sequence of data acquisition samples for the measurements that are assigned to the key. Typically, the user will define the sampling rate for each key. Each measurement that is assigned to a key will be sampled at the same rate. Since all of the measurements are collected at the same rate, they are implicitly collected simultaneously. This greatly simplifies data analysis because multiple measurements that are collecting data on a particular event can be easily sampled in the same IENA packet.

A typical IENA system supports 16 distinct keys. Each key runs at its own sampling rate. However, the system requires that the sampling rate for all of the keys be a divisor of a common base rate. The IENA system can use automatic format generation to help the instrumentation engineer to determine the optimal set of sampling rates to use for the 16 keys. By using automatic format generation, the engineer could simply enter the desired sampling rate for each measurement and then the software will select the best set of rates to use for the IENA keys.

CONCLUSION

Automatic format generation is an extremely useful and versatile technology. It can greatly simplify the setup of a data acquisition system. The power-of-two rule provides a basis for generating sampling formats that approximate a desired set of sampling rates. By applying the scaled and factorized power-of-two rules, software can be used to generate efficient sampling formats. These sampling formats are able to sample most measurements at or slightly above the user's desired sampling rates. This allows the user to save the time that would be required to manually create a similar sampling format. In addition, to its use for standard PCM based sampling formats, automatic format generation can be used for other sampling systems. In these systems, the components of the automatic format generation that are used to determine the optimal set of sampling rates can be reused.

REFERENCES

1. Kupferschmidt Benjamin, and Berdugo Albert, "Designing An Automatic Format Generator For A Network Data Acquisition System", Proceedings of the 42nd Annual International Telemetering Conference (ITC 2006), USA, October 23-26, 2006.
2. Caturla, J.P., "IENA Packet Format", Proceedings of the European Test and Telemetry Conference 2003, Toulouse France, June 2003.
3. Jones, Charles H. And Gardner, Lee S., "Automated Generation of Telemetry Formats," Proceedings of the 32nd Annual International Telemetering Conference (ITC1996), USA October 1996.