Enhanced Data Logging Capabilities in Vsystem

J.P. Girard, R. T. Westervelt, E. A. Zharkov and P. N. Clout
Vista Control System Inc.
1137 18th St.
Los Alamos, NM 87544 USA

Abstract

The Vsystem logging engine and related utilities have been greatly enhanced to support a range of advanced features and also to support a wide variety of data archiving needs. The new data storage format of the Vsystem Vlog is described. How the new storage format provides support for new features such as circular logging, flexible data timestamps, and SQL access to the data without sacrificing archiving performance will be discussed. The new Vlog storage format supports many different ways to analyze the archived data. The flexible analysis tools built onto this structure that allows for structured queries of the data will be described in the context of several typical applications using Vlog. These applications include "black box" recording for post failure analysis, very high resolution recording, training systems and alarm recording.

The new Vlog system is integrated with the rest of Vsystem and ported to a wide variety of platforms, including Windows NT. Implementing Vlog on these platforms illustrates some of the innovative techniques used in developing Vlog. The integration between Vlog and Valarm, the alarm monitoring and recording portion of Vsystem, shows the flexibility of the new Vlog architecture. The functionality of Valarm greatly increased with the new Vlog architecture.

1 Vlog -- Initial version

Control systems function most effectively with access to information about past states. This is generally true in the area of large experimental physics and accelerator systems. These systems tend to be unique and often are incompletely understood state of the art systems. Information about the past state of the system can be very useful in optimizing the performance characteristics of the accelerator and the control system.

The Vsystem logger was designed to provide a tool to meet these requirements. When the Vsystem logger was initially designed the primary consideration was to provide a data archiving tool that would achieve the highest possible data logging throughput on the target platforms along with flexible logging modes. Secondary goals were to provide a flexible interface to control the data archiving in diverse systems. Vlog works in a client server environment, with control of the logging from any client. Vlog could log data on a periodic basis, on an external event, and on change notification. In addition, logging could be control via a database channel to serve as a gate. In addition a set of tools was provided to analyze the data being archived. These tools included Vtrend; a graphical tool for plotting the archived data, Playback; an application for putting the archived data back into the Vsystem database for simulation and analysis, as well as other tools for extracting the data in the archive into other formats. [1]

The initial version of the Vsystem logger met its design goals but had a number of problems that limited its usability. While the design goal of high data throughput had been achieved, subsequent access to the log file was slow and this made analysis difficult and very time consuming. Shared access to logs was not supported and log files had to be closed to be examined. Also, the logging mechanism did not provide any automatic control over the log once it was started. The control of the log was through user action or had to be scripted. This made the process of starting and stopping logging to coincide with external events or system states cumbersome and not as precise as possible. Also, the event logging method used did not store sufficient information to uniquely determine the channel state at the start of the log period. Thus later analysis of the archive could be incomplete. These deficiencies and increasing requirements forced the redesign of Vlog.

2 Vlog -- Version 2

As the use of the Vsystem logger increased, customer requirements drove the need for substantial increases in the functionality of the Vlog system. The functional requirements were driven by the requirements in a large diverse set of industries. Vlog and associated applications were totally redesigned to meet these requirements, while still satisfying the requirements of the first version of Vlog. The new requirements can be classified into features enhancing usability, performance, and data quality enhancements. The major new features included in the redesign are detailed below with a description of the justification or customer need driving the requirement.

2.1 Circular logging

A very common use of data archiving systems is to monitor the system for unusual conditions and to later analyze the process leading to this unusual state. This is similar to the black box flight recorder type of log. This type of logging system typically requires large amount of disk space since the occurrence of anomalous conditions is typically not predictable. Also, much of the data that is logged is not of any use and is discarded at sometime in the future when it is determined that there is no need for the data. To address this requirement, the Vlog data format was enhanced to support circular logging. The storage requirements for the archive are predetermined and the storage is reused when the buffer fills. In essence, this

provides the effect of an infinite data log in a finite amount of storage space. A byproduct of circular loggnig is that shared access to the log file becomes a very necessary feature. The Vsystem implementation of circular logging, reduce the overall management needs for this type of data archive by automatically deleting "old" data under conditions specified when the archive is created.

2.2 Shared archives

For large systems, data may be continually logged and online analysis must be possible. Sharing the Vsystem archive allows the data to be analyzed while new data is added to the archive. To support this feature, support for defining the synchronization latency between the Vlog doing the archiving and the application analyzing the data was designed as part of the interface. This defines the interval at which the Vlog flushes data into the archive and makes it available to the other applications sharing the log. While shared logs do exhibit a performance penalty, the usability gains more than offset this penalty.

2.3 Archive of channel fields

Vsystem channels are complex entities with many sub-components called fields. The value of these fields is often as important as the actual value of the channel. For example, one field in a channel is the units associated with the value. As part of the Vlog redesign, support was added to log fields as well as just the channel value. This feature provides more information to the analysis tools to better interpret the data. This is a selectable component of the archive so that the size of the archive is not increased by logging unnecessary data.

2.4 Automatic archive closure

In some applications, the process being logged fixes the start and stop parameters for the archive. One example is in the steel industry, where data archives are tied to each piece of steel through the mill. In this application, the archive should be closed when the piece of steel has left the mill. Under normal operations, the time a piece of steel is in the mill being processed is fixed. The capabilities of Vlog were enhanced to support specifying either a maximum size or an end time for the log file. These criteria are used to automatically stop the Vlog and to close the log.

2.5 System restart support

Under many circumstances, when a computer system is restarted data should be appended to an existing archive. When a computer system crashes unexpectedly or when a system is restarted manually, the data archive might already exist on disk. Previously, the data archive would be overwritten losing data already present or another archive of the same name would be created, requiring more work for the operations staff to manually manage the "extra" archives. The new restart allows logging to append data to an existing archive.

2.6 External data timestamps

Often the system time resolution on the archive computer is not accurate enough for the purposes of the archive. The timestamp for the data needs to be generated at the data source site and passed with the data. This provides the most accurate mechanism for timestamping the data. A side effect of this feature is that special care is taken in Vlog for circular logs since it is important to store accurate timestamps at the start of an interval.

2.7 SQL interface

An SQL (Structured Query Language) Interface was provided for accessing the data from the Vlog archive. This interface provides all of the analysis tools (Vtrend, Playback, Vlog/table, etc) with a consistent interface to the archive. Also, the power of SQL provides increased analysis capabilities. This interface was chosen as a common and well-understood interface so that users could apply knowledge from other domains to the task of extracting data from a Vsystem archive. With the expanded set of information stored in a Vsystem archive and the possibility of much larger logs, an efficient method for partitioning the data set is vital to the usefulness of Vlog.

2.8 Merge/Union post processing

When using Vlog archives for data analysis, it is very important to be able to extract defined portions of the data into a smaller data sample to speed analysis. The merge/union capability gives the user great control over the data in the Vsystem logs allowing them very fine control over the creation of tailored data sets for off-line analysis. The merge post processing function also provides more optimized access to the selected data than was available in the original set of archives.

2.9 Vlog write API

The applications programming interface used for writing the Vlog archives in Vlog was standardized and documented. This lets other portions of Vsystem and user applications to use the standard data format. With this standard data format, the many powerful tools for manipulating and analyzing the data in a Vlog archive can be applied to other types of data sets. The availability of this interface greatly increased the capabilities of Valarm; the Vsystem tool for monitoring alarm conditions. The Vlog Write API generates an archive in event format. It also helped increase the speed of building an archive of alarm events in Valarm.

2.10 Vlog GUI

A graphical user interface was designed and implemented with the new version of Vlog. The graphical user interface makes the many options and parameters available with the increased Vlog capability easily accessible to the user. In addition, the graphical tools provide a standard interface to all of the Vlog analysis tools that are available. Figure 1 shows a sample of the Vlog User Interface screen.
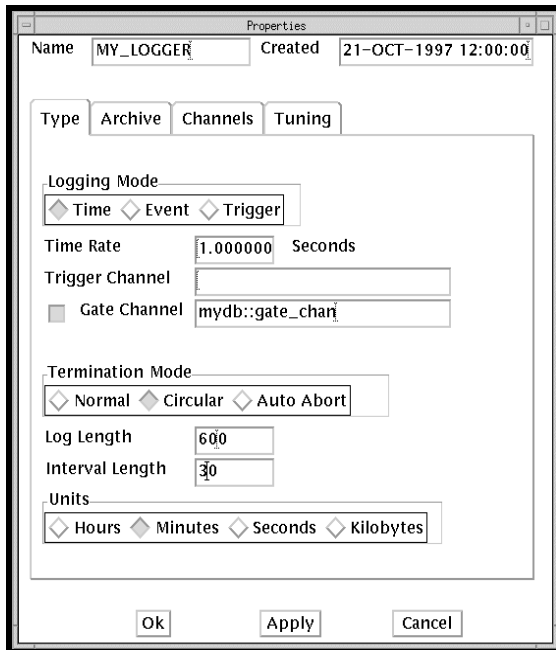
Figure 1. Vlog Graphical User Interface

3 Vlog archive format

The objectives of supporting a high data rate, providing support for fast efficient data retrieval, and circular shared logging made substantial changes in the archive format necessary. The previous format for Vlog archives used a single file. The information about channels contained in the log file was interspersed with the data in the log file and this was a primary cause of the slow data access after the archive was created. The redesign of the Vlog archive format maintained support for all of the features of the previous version of the archive files, while providing a general structure that can be expanded to meet diverse data archiving needs.

The new format uses a separate file for the header and a data file for each interval of the circular log. The header file for the circular log holds the information about the channels being logged. Indices are one of the key features that increase data retrieval times. One important feature the indices provide is the ability to quickly index deep into the data based on a timestamp. Since retrieving data based on a time period is one of the most common methods, this optimization provides a substantial speedup for most applications accessing the data. Another optimization is to bypass the SQL interface internally when records for a single channel are being retreived. There are two other indices that support increased performance; add channel block indices and mark record indices. Timestamps are generally stored as an interval from the last full timestamp rather than using the full 8 bytes required by the timestamp.

There are two file formats for the Vlog data archives; sequential and circular. Having different formats allows internal optimization for the different characteristics of the log, while maintaining the same external interface for the applications reading the log. Vlog supports a sequential format that is very useful when logging directly to tape.

4 Client examples

A few client examples showing the many uses of the new version of Vlog will illustrate the impact that the redesign has had on Vsystem users. These examples are representative examples drawn from a few of our customers of Vlog and are by no means a complete representation of the applications of Vlog. These customers are primarily drawn from our industrial customers, since they have been quickest to accept and incorporate the new Vlog.

4.1 Fast logging with timestamp at the source

One customer needed to have 72 hours of plant data online for post "failure" analysis should such an event occur. The novel features of this application are that there were several hundred data points that had to be recorded at relatively high rates of tens of hertz and precision of 100 Hz.. The front end technology was somewhat limited in its' capabilities, but was more than adequate for the required purposes. A VAXELN system talking to RTP IO provided the data source and the accurate timestamp of the data. The VAXELN system provided a timestamp for the data accurate to 0.01 seconds. More accurate tmestamps could be achieved by using a better time source in the front end. Because of the architecture of Vsystem, only the VAXELN system was required to exhibit realtime response and the other portions of the system only had be able to support the average throughput of the system. By timestamping the data at the source, the accuracy of the timestamp is maintained even if the system loading should interrupt the data logging for a significant interval. The buffering capability of Vaccess and Vlog smoothed out the response on the OpenVMS Alpha host. The host system was a Digital Equipment Corporation Alpha 3000/300x running OpenVMS. This system, even with "old" hardware by todays standards was capable of logging 5000 datapoints per second. Current systems with better front end systems can achieve rates of greater than an order of magnitude better.

In the event of a plant failure, the online log files were analyzed with Vtrend to determine the cause of the failure. The ability of Vtrend to examine several channels simultaneously makes it easy to distinguish causal relationships in the plant data. Since only this one computer is used for both analysis and data acquisition, the changes made in Vlog for improving data access are vital for these applications. Without the high speed data retreival, analysis was too slow to be useful.

4.2 Operator training systems

A significant investment is made in the training of operators in any environment. One innovative use of the enhanced Vsystem logger is to use the archive files as part of the training process. The actions of the operator in training are recorded using Vlog and are used later by the

instructor to review the work of the operator and as an instructional aid to illustrate what went wrong during the training session. The novel use of the logged data is to allow the student operator to retry certain operations tasks in the training course. The data to be replayed for the training session using Vsystem playback can be selected exactly using the advanced query capabilities of Vlog.

4.3 Coil logs

A steel customer is using the merge capability to build an archive of the data related to a single piece of steel (coil). As the coil is fabricated, it passes through several stands, each of which has many channels associated with it. There are coils in different parts of the mills at the same time. A single shared, circular log is created which holds the data from all the rolling stands. Using the merge capability allows the customer to choose several separate channels and times from the big single log, creating a single much smaller log with only the data related to that coil.

A separate methodology used is to start a new logger on each stand as the piece of steel enters. The auto-abort mode is used to ensure that this logger does not continue in case the stopping script does not run properly (the coil is a reject). After a coil is finished, all the resulting logs from each stand are merged together into a single 'coil' log.

4.4 Ultra fast logging

A rocket motor test system is proposed to use Vlog to satisfy a requirement to archive 10 channels at 25000 Hz (250,000 samples/second). This is beyond the current event log throughput of most workstations as is producing accurate 200 Hz intervals. Therefore the software will timestamp the data in a real-time operating system, and then write the archive directly using the Vlog write API.

4.5 Long term logging

Many operations need to have data available and being logged for long periods of time. One customer needs data archived for periods of longer than one year spanning times when the machine may have crashed or be brought down intentionally. The restart capability in Vlog easily supports this feature. Each time the system is restarted, the existing log file is detected and the data is appended to this existing file.

5 Valarm enhanced by vlog

The enhanced capabilities of the redesigned Vlog were used to add significant capabilities to Valarm, the alarm monitoring portion of Vsystem. Valarm was enhanced to support logging of alarm data and an interface was added to Valarm to allow the user to query from the historical alarm data. Valarm uses the Vlog write API to generate its historical files and uses the standard Vlog interfaces via SQL to access the alarm archive files. Using the same data storage method means that the alarm data can be analyzed in conjunction with the normal online data archives to determine conditions related to the alarm condition. The flexibility in logging data fields provides substantial power in storing all of the important information associated with an alarm condition, such as the alarm limits of the channel in the alarm condition, when the channel went into alarm, as well as when the operator acknowledged the alarm condition.

The archive manipulation functions of the Vlog provides examination of data in the alarm historical log using familiar tools; such as Vtrend and Playback. The enhancements and the advanced query capabilities built into Vlog provide Valarm with the ability to filter the historical alarm data by very complex criteria. For example, it is possible to select historical alarm events that correspond to a channel having a value above a certain level. In addition, the shared capabilities of the Vlog data format allow the Valarm historian to periodically be updated with newly logged data, while analysis is underway. The query/merge capabilities of Vlog can be used on the Valarm file just like any other Vlog archive.

6 Summary

By redesigning Vlog; the data archiving component of Vsystem, Vista Control Systems has been able to vastly increase the functionality and usefulness of Vlog and Vsystem for our users. The new and advanced features of Vlog have lead to its use in several novel ways that were not anticipated in the initial design of the Vlog enhancement. The expansion of the data archiving portion of Vsystem has lead to a dramatic increase in the use of Vsystem as a data analysis tool kit as well as a data acquisition and control toolkit.

References

1. P. Clout, "The Status of Vsystem", NIM A352 (1994) 442-446