THE DESIGN OF VERY HIGH AVAILABILITY ACCELERATOR CONTROL SYSTEMS

S. Hunt, PSI, Villigen, Switzerland

Abstract

As expectations of control system reliability and availability increase over time, both new and existing control systems have delivered a steady improvement in these areas. Traditional accelerator control systems have however, not required the very high reliabilities and availability figures needed in other industries such as avionics, robotics, and some sectors of the process control industry. This situation is changing. New accelerators for medical and industrial purposes sometimes need to operate outside the traditional environment of a laboratory where service, support, and expertise are readily available. Also large future accelerators may require very much higher reliability of sub-systems if they are going to meet their operational goals, particularly when considering the potential delays in replacing components for a very large machine. A reduction (or elimination!) of development time of a system requires much better up-front requirements analysis and more robust design. This can only be achieved through following a strict development methodology, well thought through and properly implemented. One negative aspect of this requirement is a need to reduce 'quick and dirty' solutions to problems, which often arise for good reason, but are sometimes implemented with insufficient thought to the requirements, to unforeseen side effects, and without documentation and testing. Systems must be built so that small modifications and extensions can be implemented by the system users, not developers. It is unlikely that many projects will have the resources to build such a system from scratch with custom code, so the use of toolkits, with well-tested, already existing software components will be the norm. One surprising by-product of this trend will be a reduction in staff necessary to maintain these systems. In the extreme, if a system has no downtime it does not need dedicated maintenance staff. The challenge to the accelerator controls community will be the adoption of best practices of industry without incurring a large cost or manpower overhead.

INTRODUCTION

As the reliability of accelerator subsystems, including controls has increased, downtime of a facility has become less accepted as being unavoidable. The high construction cost of large facilities as well as the travel and other costs of users of the facilities mean that the true cost of facility downtime is becoming recognised. Accelerators can be seen as a non-mature technology, in that unscheduled downtime is seen as a natural consequence of complexity, and a very large support staff is needed to keep the facility operational. Other industries with large, complicated systems to build and operate, such as the aviation or chemical industries, have long ago passed this phase. Industries such as automotive and robotics have increasingly complex computer-controlled systems that must operate reliably for long periods without maintenance. These industries have reliabilities much higher than we achieve in the accelerator industry.

REASONS FOR HIGH AVAILABILITY CONTROLS

As well as the direct and indirect costs of downtime to a facility, a number of requirements influence the need for high availability controls:

Large facilities

The problems of scale of existing and future large facilities pose a significant problem to providing very high availability controls. This is due both to the large number of components (reducing overall MTBF) and time needed to get to the location of the problem and replace failed components (increasing MTTR). At the same time the cost of operating these facilities is such that demand for high availability will increase. It is likely that it will no longer be possible to promise that the system will be reliable without quantifying what that means. It is also likely to be unacceptable to view controls downtime as a 'growth opportunity' - an excuse to demand more resources.

Remote operation

Possible future large accelerator facilities may very well be built as collaborations between existing laboratories. However such collaborations, in order to provide continuity of expertise at the partner laboratories, may extend beyond the construction phase and into the operational phase. It may very well be therefore that the system experts are simply not available on site, requiring a significant increase in reliability of all systems including controls. It is also possible that the control system will have to span continents in order to allow the possibility of remote operation from different control rooms located at different facilities.

Competition between facilities

It cannot be assumed that the demand for experimental time by users will always exceed the beam time available. And while some facilities may be in a monopoly situation (LHC?), most are not. For these reasons, and that there is already competition to attract the 'best' science, facilities need to provide a reliable service. User groups are often mobile, and if they do not get the beam time they expect, when they expect it, they may well go elsewhere.

Medical and industrial application of accelerators

Perhaps the requirement that will most drive the need for very high availability controls is that of medical and industrial accelerators. Accelerators for these purposes will be installed and operated in non-laboratory environments, without the traditional support infrastructure on hand. Patient treatment and production schedules will not allow unscheduled or perhaps even scheduled downtime. Accelerators of this type will become common in hospitals and industry, and will be operated by non-specialist staff, nurses and factory technicians rather than physicists and accelerator operations technicians.

HOW DO WE ACHIEVE VERY HIGH AVAILABILITY?

When we talk about very high availability, we should be considering and designing for no unscheduled downtime over years of continuous operation. A control system is a complex collection of many interrelated components, so how do we achieve this?

Quality control

It will become the norm to apply the highest professional standards to the design and implementation of controls hardware and software. As we are usually working in a research institute, we have tended to treat the development of controls as an interesting research activity, rather than an engineering activity. It will necessary to follow an engineering plan, and apply quality control to all activities. At the simplest level, do what you document and document what you do.

Re-use components

It will not be possible to treat each project as an opportunity to start designing a new system from scratch. Components, both hardware and software, can only be considered reliable when they have been extensively used operationally in the same or very similar environments. It will be necessary to use components with a proven history, which probably means industrial controls components, or components used for a long period of time at another accelerator. This will mean in effect that the controls group will have little development to do. It is important to stress that what has to be taken are the components themselves, without being re-implemented (for instance in a different language), and not just the concepts, protocols, etc.

Redundancy

Where appropriate it will be necessary to apply redundancy, even in some critical cases down to the sensor level. But often the need to make every component redundant can be avoided by good system design. It is better to eliminate the need for a single point of failure, rather than make that device complicated with redundant components. We should ask ourselves “why do we need a file server” rather than “how do we make our file server more reliable”. Other simple measures can be taken; for instance making all consoles identical so if one fails another can take over its function.

Don’t modify systems

If it is not broke, don’t fix it. Introducing any change introduces risk of failure. It is very hard, if not impossible, to fully test the effects of changes to systems. Perhaps one of the more difficult changes in the approach to control system maintenance will be for us to realize that it will not be possible to introduce a change to the operational system and wait for feedback from operators if it worked, or if any unsuspected side affects were observed. Testing will not be possible on the ‘real’ accelerator, unless a complete copy is available.

Allow users to configure parameters

Changes to the operational parameters of a system, when needed, should be possible on-line by the users. It is not practical in a high availability system to require a programmer to change a hard coded parameter in an application. Changing an alarm limit, or maximum motor position, for example, should be a user function. This also introduces a need for access control on a parameter-by-parameter basis, and recording of all changes.

CONCLUSIONS

There is a clear need for high availability accelerator controls. This requirement will only be met by changing the way we design and build our control systems.

- Applying strict quality control to all our activities.
- Using existing, well-tested software and hardware components.
- Designing redundancy into the systems.
- Minimising changes to the running system

A side effect of these trends will be the reduction of the size of control groups, during both the development and operational phases. As it is hard to retain expertise under these circumstances, a logical conclusion to this would be to outsource controls, either to a specialist company or another laboratory.