grass valley
A BELDEN BRAND

K2 Media Server Design Guidelines for Building Cost-Effective, Redundant Media Server System

Matt Allard, Market Development Manager Grass Valley, a Belden Brand — March 2011

APPLICATION NOTE

TABLE OF CONTENTS

In designing major systems there are inevitably many decisions and trade-offs. The cost of a system is always a concern, but the cost of its failure also needs to be considered as a key factor in your overall design. Understanding the available storage-system options is critical to achieving the right balance between performance and cost.

Introduction

When designing any system composed of a number of elements, one of the key issues to address is the proper level of redundancy. Given that any component could fail at any time, what course is the wisest?

The answer lies in balancing security on the one hand, and cost and complexity on the other. The less downtime that can be tolerated, the more backup equipment and signal paths are required.

When it comes to broadcast server systems, downtime requirements can vary. Premium channels require five nines, or 99.999 percent availability. Other channels might be content with a slightly higher level of risk. Still other operations, such as multichannel facilities, might dynamically reallocate a server system's availability during different times of the day. Still others might assign different reliability constraints to different parts of their workflows.

The key to designing server systems for redundancy is to satisfy the required levels of availability without the over-provisioning of resources and signal paths. This white paper examines the necessary approach to such a system design, and illustrates the advantages of using Grass Valley K2 server systems to do so..

K2 System Overview

Ideal for transmission and production environments, the K2 system supports a wide range of applications to address specific user needs in the most cost-efficient and cost-effective ways possible.

The power of the K2 platform lies in its architecture. It can be used in a variety of system designs, from standalone and mirrored configurations to distributed and shared storage-area network (SAN) topologies — or in any combination. Redundancy can be added to an existing non-redundant configuration easily and cost effectively. The system also scales very well, so a solution can start small and then expand as channel count, storage hours and/or bandwidth needs increase.

A K2 SAN consists of clients, servers and RAID-protected storage systems. The client performs all the video/audio input/output using built-in encoders and decoders. A standalone client can also incorporate the server functions with integrated internal storage. In a K2 SAN, separate servers manage the file system, the data flow between storage and clients and control file transfer operations. The unique built-in quality of service (QOS) capabilities dynamically allocate bandwidth where it's needed, always ensuring that on-air channels do not display loss of video.

The K2 platform comes standard with AppCenter software (Figure 1), a suite of tools for basic transport control, media management, configuration control and playlist management. Unique to the AppCenter application is its ability to run locally and over a remote connection that controls any channel, anywhere in the world.

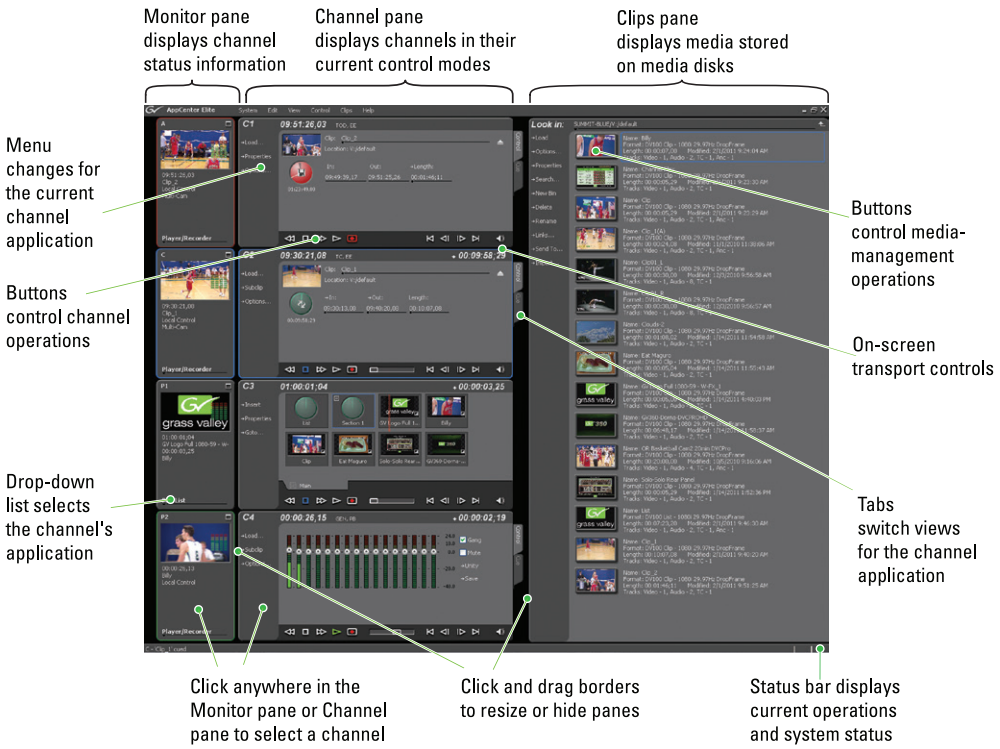


Figure 1 – The K2 system includes AppCenter Elite software, which features a powerful, easy-to-use interface for file management, transport control and configuration management.

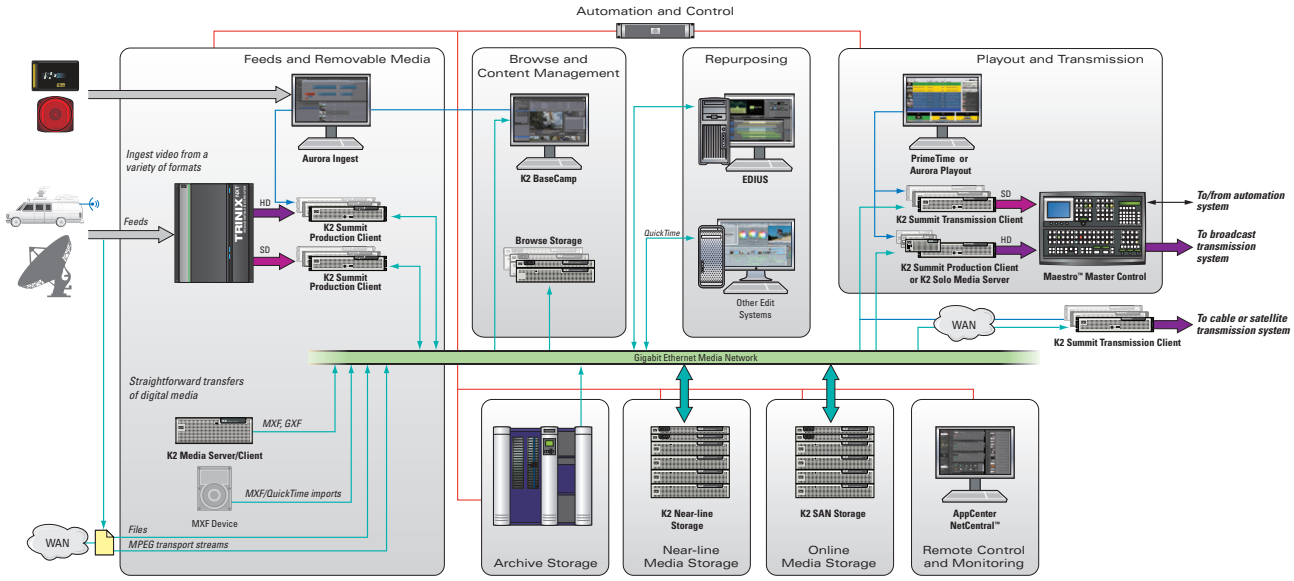
Evaluating A System's Redundancy

Redundancy of a system can be evaluated at four levels:

- **Operational requirements.** Does the use case allow for alternate ways of performing all required operations, perhaps with reduced functions, in the case of any failure?
- **System architecture.** Having defined the usage model, it needs to be determined as to how it should be implemented. The architectural options include a SAN, a distributed storage model or a combination of a SAN and distributed storage. The highest level of redundancy is to mirror everything, but the issues using this approach include failure detection, switchover time, operational procedures and of course- cost.
- **Data-flow analysis.** This step involves an analysis of the usage model with the system architecture. The purpose of the analysis is to determine the number of paths, the routing from one point to another, and to determine the level of protection for each data path.
- **Component analysis.** This step involves considering each individual component and determining how much redundancy is required. It is important that this is the fourth and final stage of the process: too often designers start at the component level without a good understanding of the rest of the structure, and end up with a poorly constructed solution.

As stated earlier, the goal of this analysis is to develop a redundancy model that balances a system's cost against the consequences — operationally as well as financially — of its failure. If redundancy is accounted for at the use case and system architecture levels and combined with a flexible operational model, significant costs can often be saved.

Understanding Your Workflow



The term workflow is now widely used in various industries, but in this whitepaper it is taken to mean what it says: a model of how a specific media operation works, how people do their jobs and how content flows through the facility. Part of the workflow involves understanding the actions performed on content and how that content is packaged for viewers.

There are many ways to design a workflow. It is not tied to specific technical solutions: facilities that perform the same overall process with the same equipment may have different internal workflows.

At the top level, though, and for purposes of illustration, a common workflow for a broadcast facility starts at ingest, moves through quality assurance, editing, playout and ends at archival (Figure 2).

In such a workflow, these general statements can be made about the various steps:

- **Ingest**. This step involves getting the content into a facility. In years past, facilities received an analog video signal and encoded it into a preferred compressed digital video format which could be stored on a server network.

Today, it is common to ingest a file that has been encoded elsewhere, and transferred to a facility via LAN or WAN network. Increasingly, even cameras are capable of transferring content as files from their tape, disk or solid-state storage.

When ingesting a file, consideration must be given for the essence itself (video and audio), its compression format, the file wrapper (QuickTime, MXF, GXF-SMPTE 360M or some other flavor) and the metadata (descriptive information about the content), to ensure full compatibility. Provision may also be needed for some content transcoding, unwrapping and rewrapping at the ingest stage.

Understanding Your Workflow (Cont.)

- **Quality assurance.** This step involves logging materials and verifying them to be correct and within house technical standards. This is also the point at which metadata may be added to identify the content, and usually manage the low-resolution proxy for browsing. Traditionally, this step has involved inspecting content visually to determine its quality. Today tools are available for performing file-based quality checks against a test template that automates this step — or, at the very least, greatly reduces the manual effort required.
- **Editing.** This stage covers a broad range of tasks. Video may need to be edited for duration or for content. It may be desirable to add special effects, local branding or graphics. Voiceovers and language dubs may need to be added. Other services such as subtitling/closed captioning, signing and audio descriptions may need to be accessed.
- **Playout.** Because it involves delivering the content to air, or distributing it to another location for playout, this stage is generally the most critical and it is here that reliability needs to be the highest — as close to 100 percent as possible.
- **Archival.** This final or first step places or retrieves content with a long-term, offline storage system. It can involve a single stage — such as the transfer of content with server and data tape — or a dual stage architecture that uses a nearline disk for intermediate storage and a robotic data-tape library for deep storage.

In a dual-stage approach, a fast-access RAID array generally provides short-term (typically two to four weeks) storage and then material is sent to the robotic tape library for permanent archiving. This approach is another consideration in creating a redundancy model as it lets users rapidly access all online content from nearline storage in the event of a problem.

To design for redundancy at each stage, there should be alternative methods should a particular piece of equipment fail.

For example, in a news application, a nonlinear editing (NLE) suite could be chosen as the primary point of ingest, with local storage feeding a SAN as a background task. Should the NLE system fail, ingest can be done directly into the SAN and a desktop editing tool used to make rough cuts on the server.

Achieving this capability would require little more than the addition of extra router ports. However, it is a good example of thinking through redundancy at the usage and operational levels rather than just adding spare equipment that will rarely, if ever, be used.

The important thing to remember is that, while there are always alternative ways to complete a task, improvising a solution under pressure rarely works well. Designing a fallback position into the workflow, and training staff for such a situation, is a much more secure solution.

Designing a System Architecture

The first decision when designing a system architecture is the storage structure: it can be SAN-based, distributed or a mixture of the two.

- **SAN.** In this model, all ingest, editing and playout can take place on the SAN. This architecture makes most sense where many people and systems need to access the same content, particularly in production. For playout, if there is a lot of common content for many channels, SANs eliminate the need to make many duplicate copies of the same material, saving disk space and communication paths.
- **Distributed.** In this model, each server has its own internal storage. Playout servers receive content as file transfers from archive, SAN, servers or other storage arrays. The playout server is typically loaded with enough content for one or two days of transmission, with an intelligent asset-management system keeping the server content up to date. This architecture is appropriate in a multichannel playout facility where there is little common material across the channels.
- **SAN/distributed storage.** This model is increasingly seen as the most popular solution (Figure 3). In it, all ingest, quality assurance and editing is performed on the SAN; content is pushed to playout servers when required for broadcast. This provides a natural firewall between production and playout, and allows better load balancing.

In this design, only the playout servers are mirrored for full redundancy rather than the entire SAN. As a third level of backup, content could be played directly from the SAN if for some reason the distributed servers are down — or if last-minute changes are made to content and there is no time to transfer it for playout. K2 systems can be configured in these topologies with production level storage for the SAN. The higher capacity drives in production storage are precisely matched to offer a lower cost alternative to SAN systems that do not need to play to air.

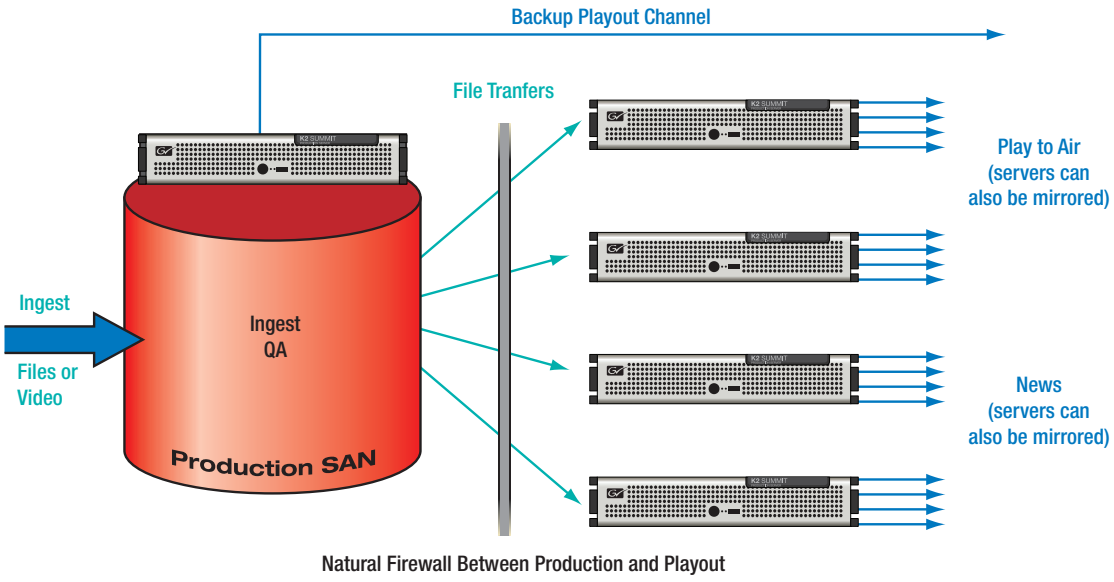


Figure 3 – In a mixed SAN and distributed design, all ingest, QA and editing/production is performed on the SAN. When a clip finishes, it is pushed to the distributed playout servers. This provides a natural firewall between production and playout.

- **Mirrored Systems.** Many large facilities will design their systems for capacity then simply mirror it: install two of everything. This, of course, provides excellent redundancy, but at a cost not all facilities can afford.

When mirroring, you can create maximum redundancy, minimal redundancy or something between the two. Adding as much redundancy to each side of a mirrored system is the best solution, if for no other reason than to provide a great deal of security and to be able to take one side down for routine maintenance. On the other hand, this approach involves considerable capital investment.

Conversely, minimal redundancy in a mirrored system implies that each side of the mirror is fully functional, but there is little redundancy within each side, thus putting it at greater risk of failure.

In reality, the most logical approach is to take a middle path somewhere between these two extremes. Some redundancy options — power supplies are an obvious example — are so inexpensive that to omit them would make little difference to the overall project cost.

Whichever architecture is chosen, the next step is to look at how to add the required redundancy. The following sections take an analytical, step-by-step approach to redundancy planning using the K2 media server/media client system.

Data-Flow Analysis

In performing a data-flow analysis, the goal is to examine all data paths for content and control systems to ensure the right degree of system resilience.

In a K2 server system, there are three sets of data paths: control, file transfer and media. Media can be moved with iSCSI over Ethernet or Fibre Channel.

- **Control network.** The control network is the low-traffic network that connects all component parts via Gigabit Ethernet connections. No content data goes over this network; rather, it is used for controlling all devices so it must be very reliable and deterministic. Since broadcast video requires frame accuracy, a command stalled by other network traffic could create bad on-air video.

Ideally, this network should be redundant as the cost is usually only a second Ethernet switch and some cable infrastructure. But, it is vital that this path is isolated from other networks, and especially corporate networks, to minimize traffic and to eliminate the potential for viruses.

- **File transfer network.** The file transfer network moves content data over a Gigabit Ethernet connection using FTP or CIFS protocols. It is the equivalent of a video-router network, but instead of moving video in real time it is moving files between locations at high speeds. It is this ability to move content faster than real time — a real benefit in workflow and network architectures — that drives the need for speed on this network.

For a file transfer network there can be redundant design, or use multiple networks each with dedicated servers. In the case of multiple networks, if one network goes down an alternate network path or server can be used. Here the operational design needs to account for how to switch networks or failover to a backup network quickly and without losing data.

Since a file transfer network frequently requires access to the outside world to exchange content with other systems in other locations, it must be isolated by a strong firewall with strict access rules. Nothing but expected video content should be allowed to pass through this secured checkpoint.

- **iSCSI/Fibre Channel network.** The iSCSI/Fibre Channel network is the connection between a storage system and a video server or other attached devices, such as editing stations. It is designed to do one task: move data between storage and client applications. It is absolutely critical that this network be highly deterministic and set up for high quality of service control. It should also be one of the first areas to consider for redundancy.

Data-Flow Analysis (Cont.)

Grass Valley K2 servers and Aurora nonlinear editors are designed with built-in QOS control to guarantee that system bandwidth always goes first to video traffic rather than to file transfer traffic. A directly attached editor such as Grass Valley EDIUS or Apple Final Cut will require multiple streams of content concurrently — and in real time. The design of the K2 server software is that it will provide three distinct levels of QOS. The top level is real time with system bandwidth priority for media clients. The bottom level is additional bandwidth for non real time uses such as FTP transfers or common Internet file system (CIFS)-attached tasks. A third level in between is for edit systems with allocated bandwidth to specific clients that are not allowed to exceed their allotted bandwidth settings.

While many video server systems may claim QOS, the K2 platform has this capability at the very core of its file system and software management layer. It dynamically manages and allocates bandwidth between all connected devices so that available bandwidth is being used intelligently. In this way, the K2 platform delivers the required performance levels necessary for all critical content moves and ensures the maximum use of the available bandwidth at all times. By contrast, conventional QOS approaches fix specific amounts of bandwidth for different levels of users but cannot reallocate any unused bandwidth between them. This sophisticated management layer ensures that every type of bandwidth user, gets the bandwidth they need, when they need it.

The K2 system supports both iSCSI and Fibre Channel options for its internal network connecting clients to the K2 SAN. As a general rule, each of these networks has its own characteristics:

- K2 uses iSCSI protocol across Gigabit Ethernet networks with tuned TCP/IP offload engines on the servers for optimal Ethernet performance. Dedicated processor cards control data traffic between the storage and video server client or editor, minimizing the CPU load so that load in turn can be used for other services. As more bandwidth is required, more servers with TCP/IP offload engine cards can be added to the network.

Additionally, Ethernet based networks are generally less expensive than Fibre Channel versions, as Ethernet cards and switches are much more mainstream devices than Fibre Channel ones. Most engineering staff tend to be much more comfortable with installing and extending Ethernet infrastructures.

Fibre Channel networks require Fibre Channel HBA cards in each server or client, as well as a Fibre Channel switch. This approach is a more expensive solution than an Ethernet based one, but the bandwidth will be greater, so it may be best for very high-performance systems.

When considering the connectivity for an internal network in a redundant system, it's important to consider how the network will failover to a backup path. One advantage of a Fibre Channel network is that it detects many faults more quickly than an Ethernet one. As a result, systems will sometimes failover much more quickly.

The expression no single point of failure is sometime used in regard to data flow and redundancy requirements. Be aware that this phrase means different things to different people. To be technically correct, no such thing exists: a catastrophic power outage or even a simple human error could do significant damage to a facility and/or data. Thus the goal in designing a system is to achieve as few points of failure as possible, and that means looking at every piece of equipment.

For example, some devices have dual power supplies but they are both fed from the same cord; this approach protects against a failed power supply but not a failed power source. Dual power supplies require dual power sources to eliminate the point of failure. Similarly, if there is a single-circuit cooling system with a single power supply, it could be a point of catastrophic failure. The point is this: while all potential failure sources cannot be eliminated, a thorough analysis will help eliminate as many as possible.

Building A Redundant Design Using The K2 System

A basic K2 SAN system is composed of a client (video I/O), a server (FTP and file management) and storage (Figure 4). In this system, data flows from the:

- Client to the switch over a Gigabit Ethernet connection using iSCSI
- Switch to a TCP/IP offload engine on the K2 server
- K2 server to the RAID controller over 8 Gb Fibre Channel connection
- RAID controller, with data striped across the disk drives

In a redundant design (Figure 5), a second connection from the client is connected to a backup switch, a backup server and a backup RAID controller. In the event of a failure anywhere in the main data path, the system switches to the backup path.

A further level of redundancy can be provided by putting the file system and database on separate redundant servers with separate redundant RAID drives that contain only the file system and database information.

Rather than duplicate all of the clients, as it is unlikely that two will fail at the same time, one option is to add a spare in an N+1 design. Should a client fail, a single backup should suffice, provided that the automation system is configured to use the spare server in the event of a failure.

The advantage of this approach is that the most expensive components — server and storage — do not need to be duplicated.

While the N+1 design can protect the clients, the storage system can be protected with dual RAID controllers. The data itself is protected with a RAID-5 or RAID-6 implementation.

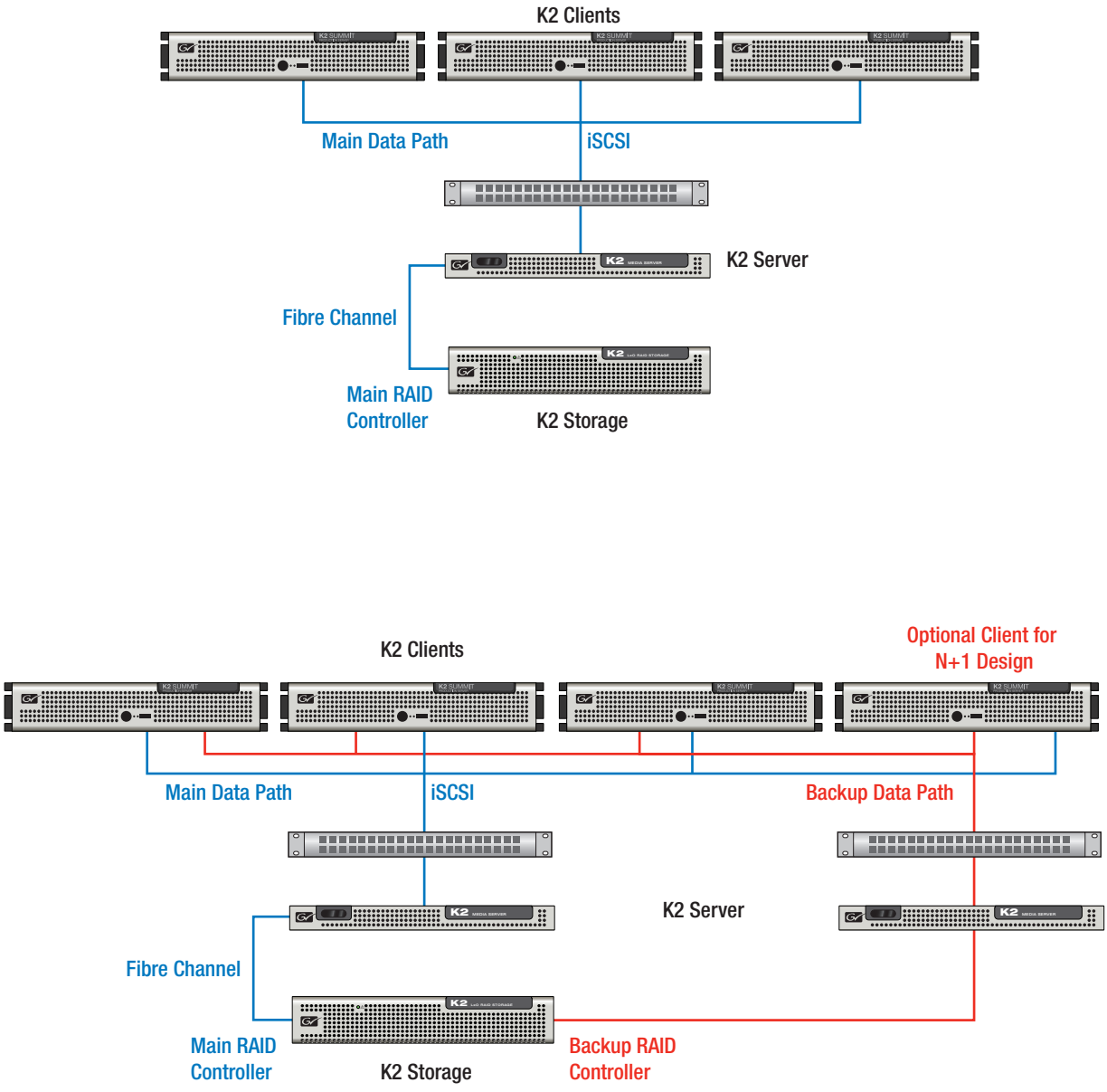


Figure 5 – In a redundant K2 SAN design, a backup data path is added with backup switches and servers. If a failure is detected in the main data path, the system switches over to the backup data path.

Performing A Component Analysis

After analyzing and designing the basic system architecture, the next step is to look at individual components to see how they can be made as reliable as possible. Generally, a fair amount of redundancy can be added to individual components at a very low cost. These components include:

• Power supplies. Redundant supplies are generally a low-cost option on most broadcast equipment. Remember, to be effective, supplies require two independent power sources and a good UPS system to be comprehensively protected from external power problems.

• Cooling. As with power supplies, redundant fans and extra cooling are either standard or very low-cost options.

- **System drive.** As almost all devices are now built on an operating system and applications software, the system drive becomes as critical as the power supply. Adding a mirrored system drive is a very cost-effective option. However, you should remember that this is only a protection against drive failures: if the main drive has a software bug, the mirrored copy will, too. Another method of providing a more robust device is to use solid state drive such as Compact Flash to host the operating system and device software.
- **Ethernet ports.** Dual network connections are a must. The IT term for redundant ports is teaming, where the Ethernet driver will treat two ports as one and will seamlessly switch to the backup port if the main fails without changing IP addresses.

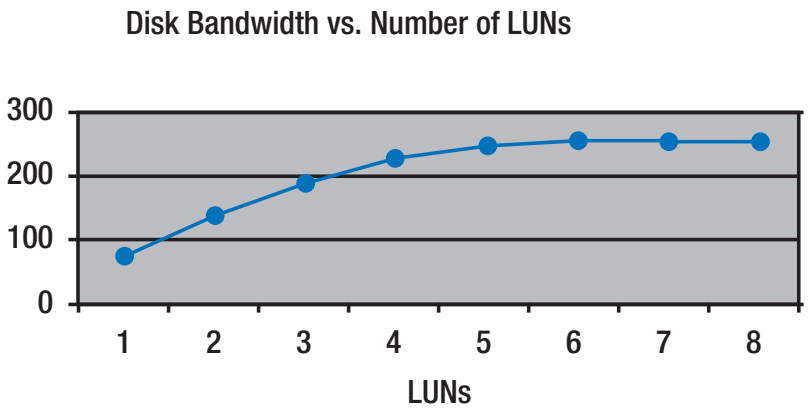
All K2 clients, servers, storage and switches have redundant power supplies, cooling, Ethernet ports, along with CompactFlash system drive as standard features.

Storage Redundancy

The next consideration is RAID storage components. These components play two critical roles: they provide hours of data storage and they determine the system bandwidth.

The limiting factor in bandwidth in video server systems tends to be the speed at which content can be moved off the disks. By spreading data over a number of disks, system bandwidth is improved because it is no longer constrained by the data rate of a single disk.

Disks are grouped into logical unit numbers (LUNs), with the disks in each LUN protected with one or two parity disks. The K2 RAID storage system uses a 5+1 LUN (five data disks and one parity disk) at the RAID-5 level or a RAID-6 option with a 4+1+1 LUN (four data disks and two parity disks). The more LUNs in the system, the higher the system bandwidth (Figure 6), until the maximum capacity of a single RAID controller is reached, in which case more controllers are added.



Performing A Component Analysis (Cont.)

In the early days of RAID storage, the wisdom was that RAID-3 was optimal for video (large files, few transactions) and RAID-5 for IT (small files, many transactions). Today, with higher-performance disks and RAID controllers, this is not an issue: RAID-5 works well with video and is provides better storage efficiency.

The concern with RAID-5 systems is that there is only one parity drive, so that should two drives fail in a LUN (a rare event in a well-maintained system), data could be lost. One way to minimize this risk is to keep the LUN size small. While other manufacturers use LUNs as large as 9+1, the K2 online systems use 5+1.

RAID-6 storage offers two parity drives (4+1+1 LUN in the K2 system), so three out of six drives would have to fail before data was lost, which is an extremely low probability; thus RAID-6 is an excellent, low-cost way to provide additional security. Compared with the alternative of mirroring the entire storage system, RAID-6 provides a substantial cost saving for virtually the same level of reliability.

One concern is that, should a disk fail, there will be some bandwidth impairment while the LUN is rebuilding itself. In truth, the reduction in performance is small, and K2 systems are specified with the RAID system in rebuild mode so under normal circumstances they will perform better than specified.

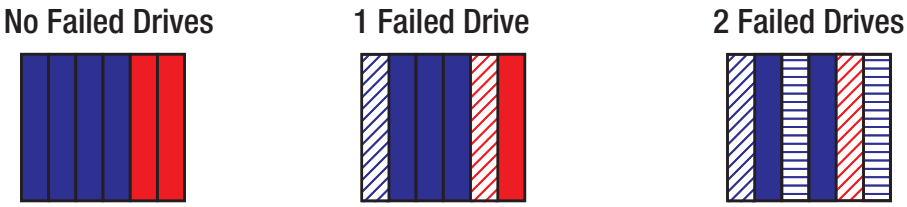


Figure 7 – RAID 6 provides two parity drives per LUN – in this case there are 4 data drives and 2 parity drives. When data drive 1 fails, parity drive 1 takes over. When data drive 3 fails, parity drive 2 takes over.

Using Nearline Storage And Archival In Your Redundancy Plans

Another method of providing storage redundancy at a very cost effective level is to incorporate near-line storage in a facility. The difference between online and near-line storage is that near-line systems do not need the deterministic high performance required of online schemes: that is, the ability to meet all requests for video to air under all conditions. Thus online storage is inevitably more expensive than other storage options because of the need to guarantee its performance.

Near-line storage still needs a fast pipe to move data, but it does not need to be as deterministic. If there is a short stall in a transfer for any reason, users are unlikely to notice. This means SATA drives can be used for a much lower cost per gigabyte. K2 storage systems compensate for the lower reliability of SATA drives by always using RAID-6 protection.

Near-line storage can be used in conjunction with data-tape archives for a two-stage solution. Since RAID provides very fast access at a low cost, content can be stored for as long as users are likely to need access to it. Then, as usage goes down, material is offloaded to data tape. In practice, a well-managed system will spread content across online, near-line and tape storage to provide maximum protection as soon as possible after ingest.

This model can then also be used to reduce the cost of online storage. By adding the capability to move data when and where it is needed, the SAN can be sized for immediate use only — maybe one week's worth of content — and an intelligent data mover introduced to control the flow between elements.

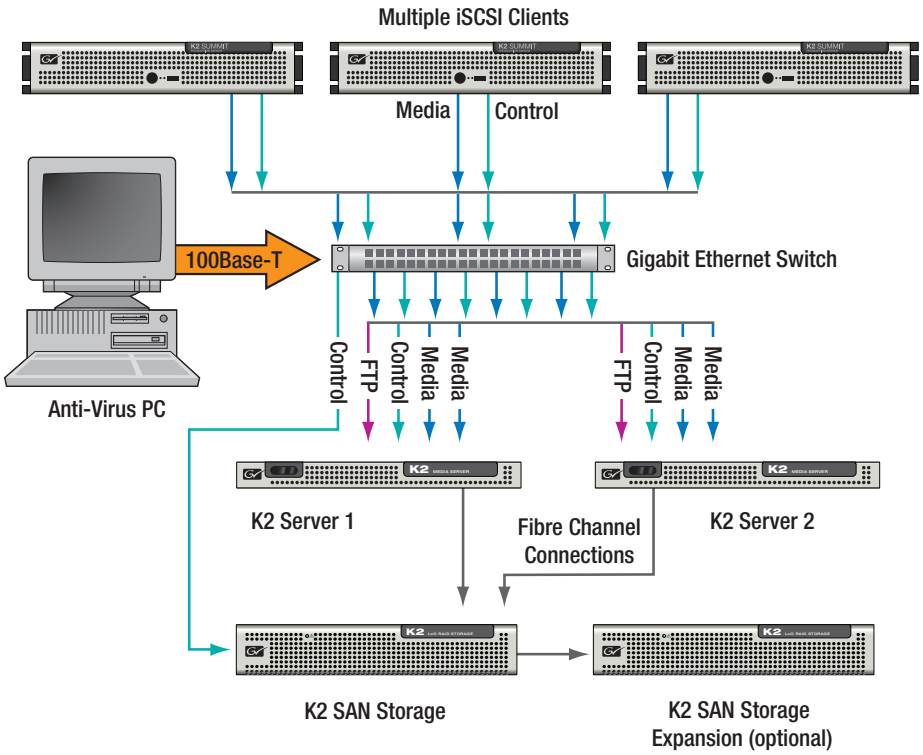
This approach does call for a degree of sophistication in the design of the system, and for intelligent archive management; but it can save costs and add another level of data redundancy.

System Protection

No discussion of system reliability can neglect system protection from such issues as virus infiltration or denial of service attacks. There are at least three levels of protection that should be designed into a system from the start:

- **Network isolation.** There is no reason why many broadcast networks should ever be connected to corporate networks, and this provides the best protection possible. However, some services — news is the obvious example — will need to connect to other computer networks and this may represent a risk.
- **Firewall construction.** When access to the external world is unavoidable, a strong firewall is required. This will restrict network access to trusted systems and individuals. K2 systems can use most any commercial firewall as long as it is configured properly to provide proper access to specific ports.
- **Virus protection.** There are a number of well-known and very successful anti-virus programs on the market. However, it is critical that these are used with care. Even running in the background, they may take a significant number of CPU cycles, and if they spot a virus they may take virtually all CPU cycles while they kill it, causing visible disturbance to the on-air video.

The solution to this problem is a dedicated virus-checking PC to monitor all network disks remotely (Figure 8). This PC mounts all the drives in the network, continually monitoring them for viruses. By connecting the PC through a bandwidth-limiting switch port you can stop it from swamping the network. The K2 system recommendation is that the virus checker be bandwidth- limited to no more than 100 Mb/s. The virus checker will carry out all the processing, so there is no CPU load on critical devices.



Advantages Of K2 Servers In Designing Cost-Effective Redundancy

In designing major systems there are inevitably many decisions and trade-offs. The cost of a system is always a concern, but the cost of its failure also needs to be considered as a key factor in the overall design. Understanding the available storage-system options is critical to achieving the right balance between performance and cost.

The K2 system offers many advantages compared to other products on the market. So when designing a cost-effective, redundant system for broadcast and production applications, here are 10 advantages of the system worth considering:

- K2 ability to work with SAN, distributed or combination networks provides maximum flexibility.
- The architecture of K2 clients can be either standalone or SAN configurations so support and maintenance is greatly simplified.
- The level of redundancy can be specified at each of its component levels to create the system that best meets particular performance and cost criteria. K2 dual data ports (iSCSI or Fibre Channel) enable redundant data paths without duplicated clients and storage for cost effective security.
- K2 core components contain redundant power supplies, cooling, system drives and Ethernet ports.
- K2 RAID-6 protection provides an affordable alternative to storage duplication, with a small LUN size (4+1+1) to provide virtually the same statistical reliability as mirroring — at a potential saving of hundreds of thousands of dollars.
- K2 production storage and SATA-based near-line storage offers practical alternatives to complete online drive data access.
- K2 integrated AppCenter interface provides utilities for ingest, editing and playout control.
- K2 offers a safe and non-intrusive alternative for virus protection.
- K2 supports a full set of tools to support workflows and architectures, including the Grass Valley Aurora fast turn production suite, EDIUS craft-editing and K2 Dyno Replay System.