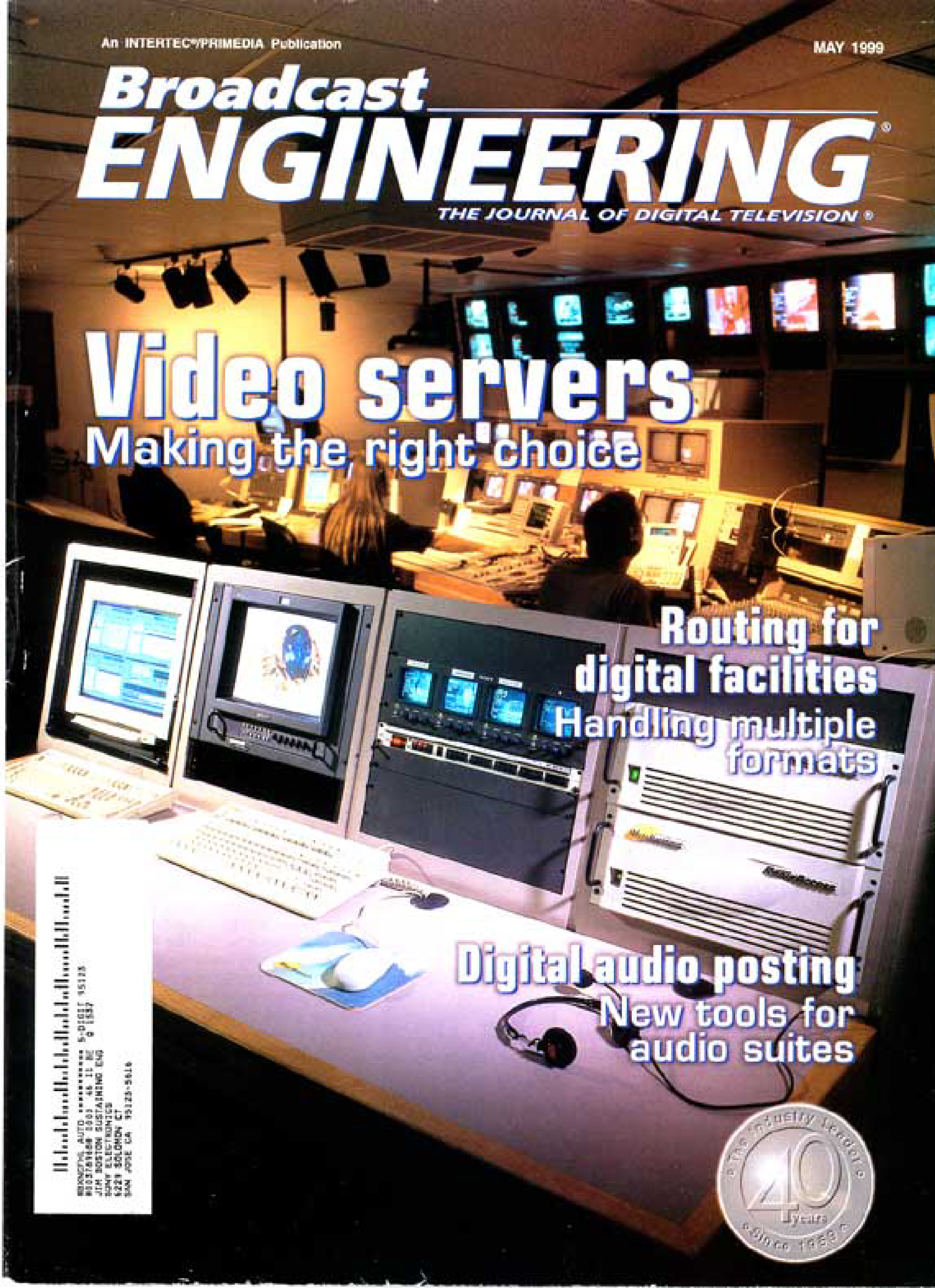
An INTERTEC®/PRIMEDIA Publication

MAY 1999

Broadcast ENGINEERING®

THE JOURNAL OF DIGITAL TELEVISION®



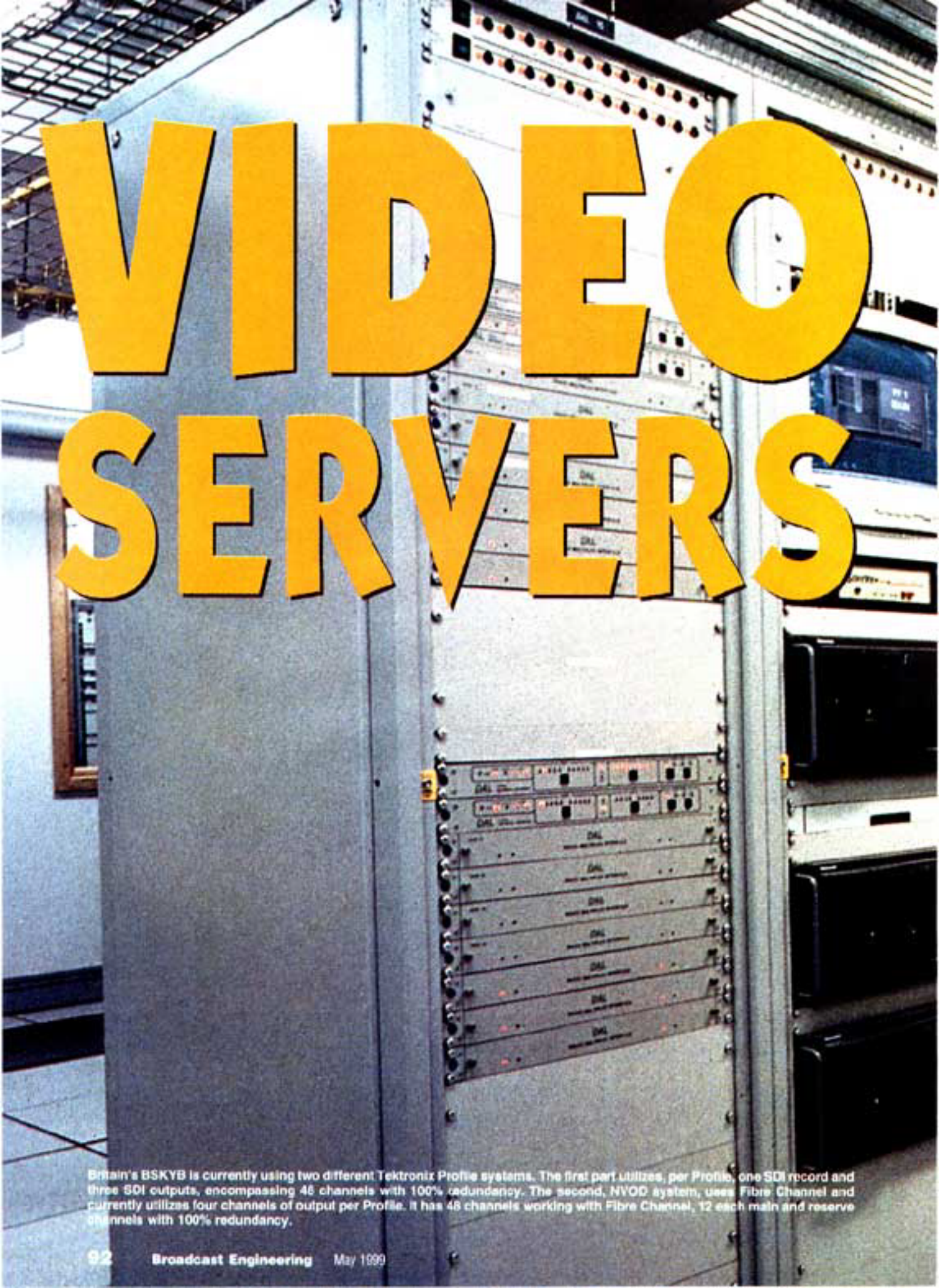
Video servers
Making the right choice

Routing for digital facilities
Handling multiple formats

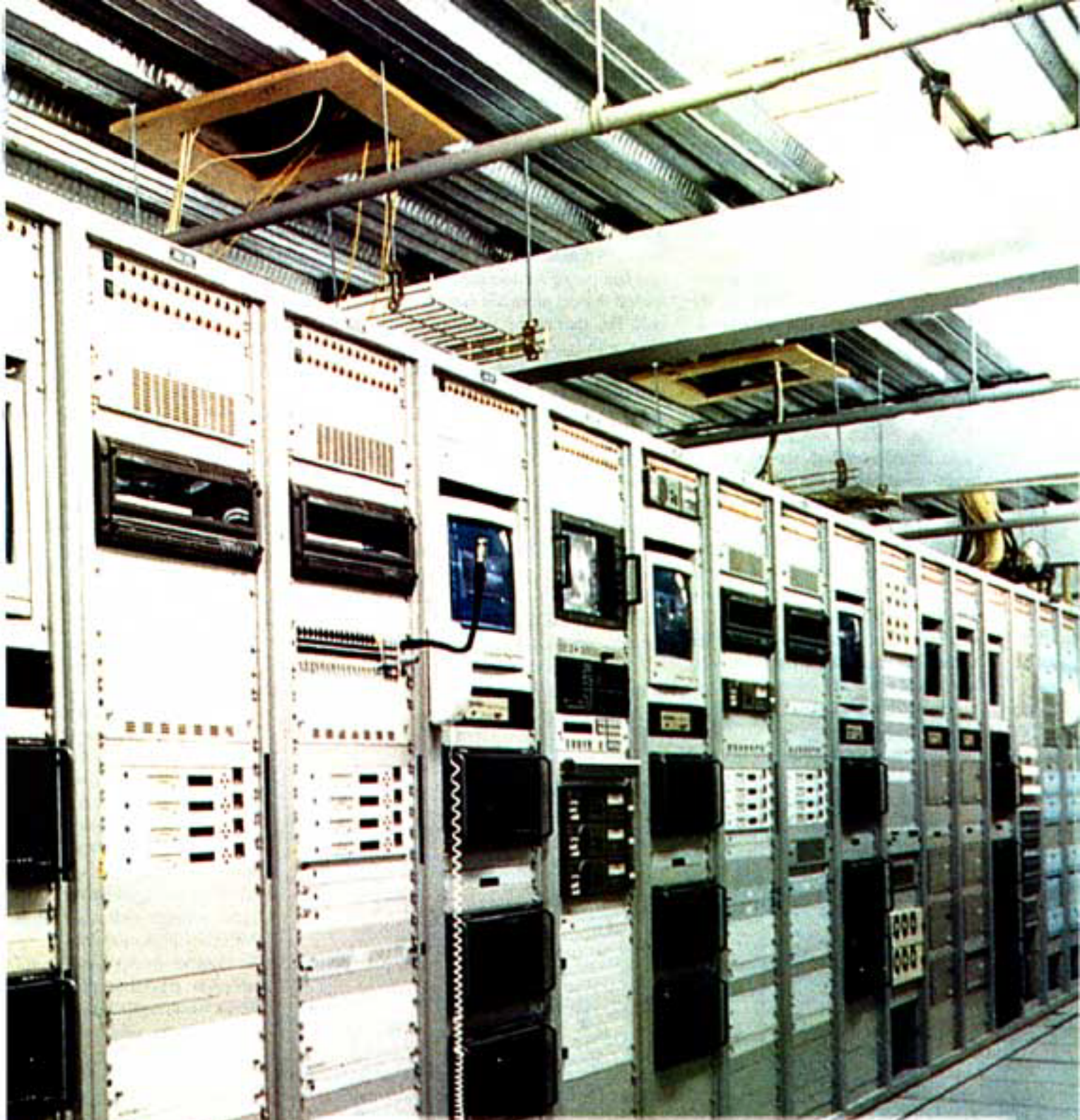
Digital audio posting
New tools for audio suites

The industry leader • 40 years • since 1959

VIDEO SERVERS



Britain's BSKYB is currently using two different Tektronix Profile systems. The first part utilizes, per Profile, one SDI record and three SDI outputs, encompassing 48 channels with 100% redundancy. The second, NVOD system, uses Fibre Channel and currently utilizes four channels of output per Profile. It has 48 channels working with Fibre Channel, 12 each main and reserve channels with 100% redundancy.



By Jim Boston

Anyone contemplating the design of a facility today will most likely base decisions around one or two basic building blocks. Until recently the router has been the cornerstone of most facilities' designs. Today, the video server has evolved to a point where it has become the centerpiece of many new installations. Not that the router is no longer a central piece of infrastructure, but the server has matured to the point where it makes sense to organize a facility's workflow around it. The video server and the router are often tightly intertwined with one another. Often, the routing system functionally wraps around the server or set of servers. The router is often controlled by the same system that has command of the server.

VIDEO SERVERS

Numerous technologies have made today's multimedia server possible, among them, rapid progress by the compression and microprocessor/DSP industries as well as disk drives with greatly increased data density. Much of the progress in that area is due to the breakthroughs in read/write head technology. Drive vendors have also utilized DSP to support the head design breakthroughs.

Server applications

Like digital television, servers are different things to different people. Servers can be used as simple VTR replacements. These are generally known as Digital Disk Recorders (DDRs). Most mimic VTRs from both an operational and interface standpoint. Many have RS-422 interfaces that use common VTR protocols. These systems work well for material that must be played out repeatedly as well as for time shift applications.

From simple VTR chameleons, many DDRs morph into more ambitious creatures. DDRs can produce multiple playout streams. While some have internal storage, many have external storage, including RAIDs. You'll find DDRs that have SCSI, Fibre Channel and Ethernet connectivity. Often the delineation between the DDR and the full-blown server is indiscernible. Systems that store and manage commercial and news story inventories are generally thought of as servers, but server systems can consist of a combination of DDRs and one or more servers to form large systems. This is the case with a number of vendors who offer news and enterprise servers.

To some users the DDR is a server. Most standalone DDRs being used for commercial spot playback are referred to by their owners as servers. A DDR used in an edit bay is most likely referred to as a nonlinear editor. But that DDR could be a part of a larger server system. Generally large higher-capacity DDRs that have client DDRs are properly referred to as a server.

The very term video server is rapidly becoming obsolete. First, they not only deliver video but also audio. Some can also be made to act as still stores. Additional development is occurring that practically demands that these devices be called multimedia servers. Currently, a number of vendors have file transfer and sharing protocols that allow not only the sharing of video/audio but data about that video/audio, or metadata. These protocols break the file into objects that can be manipulated individually. Objects can be split up and operated on at different workstations simultaneously. With the coming of MPEG-4, this evolution will continue. MPEG-4 also breaks video scenes and audio passages into objects. No longer is compression based solely on spatial and temporal relationships. Instead, the baseband scene is considered a collection of objects. Objects could be a talking head, the background wall, a chart, etc. Most objects in a scene will still have spatial and temporal compression performed on them, but that compression will be done without reference to the other objects. At the receiving end these objects will be re-associated and layered back together.

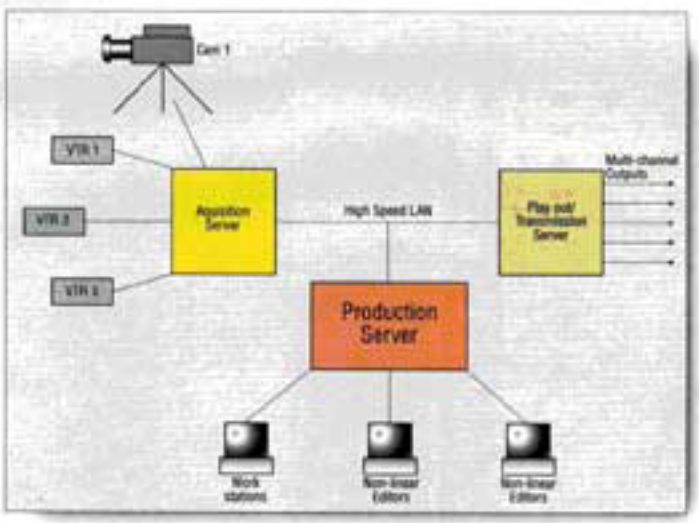


Figure 1. Depending on facility workflow, a large server system could be divided into subsystems, for example acquisition, production and playout/transmission.

Server architecture

Within most servers, the encoding/decoding subsystems often occupy the same card or box. Each pair is usually called a channel. The storage control subsystem is the gatekeeper to the actual storage system, which is usually some form of RAID (array). Controlling the storage control box, and usually the encoders/decoders, is a control system that could be a PC or workstation. Some systems have enough separate control tasks and user applications occurring simultaneously that multiple networked PCs are required. Connection between the PC and storage control can be anything from Ethernet to RS-422. Connection between the storage control and the RAID array may be SCSI, but Fibre Channel is gaining favor. If Fibre Channel is used, verify that it goes all the way to the drives. In many older designs, the actual connection to the drives was done with SCSI.

Although some servers can handle uncompressed videostreams, most apply either JPEG or MPEG compression schemes into and decompression out of the server. Besides the fact that uncompressed video consumes great amounts of disk space, it also consumes internal server bandwidth, and storage systems can only write to and read from the disk at a finite rate.

Several things define the bandwidth or throughput of a server system. The first is the compression ratio. JPEG is still widely used as it allows for editing on every frame. Every JPEG frame is, in essence, an MPEG-1 frame. However, most servers today use MPEG. MPEG provides a fourfold or fivefold compression efficiency over JPEG. Reducing the bit rate means fewer reads and writes to the storage system. Lower internal server bit rates mean more clients (channels) can be served simultaneously. Many systems let you select the trade off to be made between compression (video quality) and the number of simultaneous I/Os.

In most systems, server I/O ends up being time multiplexed for transmission to and from the storage system. Internally, video must be moved at faster than real time. RAID arrays allow throughput higher than individual drives, but as drives are added to an

array, a point of diminishing returns is reached. RAID throughput using SCSI often maxes out around 25MB/s. Currently, one of the most common systems is to use multiple RAIDs into one large server storage system via Fibre Channel. The RAID scheme ties up some bandwidth, but internal server bandwidth can be 30MB/s, with some systems offering greater than 50MB/s.

Overall throughout can be limited by a number of factors. Most disk interfaces have transmission overhead. SCSI has a 20 percent overhead. One fifth of the bytes sent over a SCSI bus are for handshaking and control. Fibre Channel has its own overhead requirements and uses eight-bit/10-bit channel coding which shrinks throughput by 25 percent. This is among the reasons why Gigabit Fibre Channel seldom has actual throughput higher than 400Mb/s.

Formatting for storage

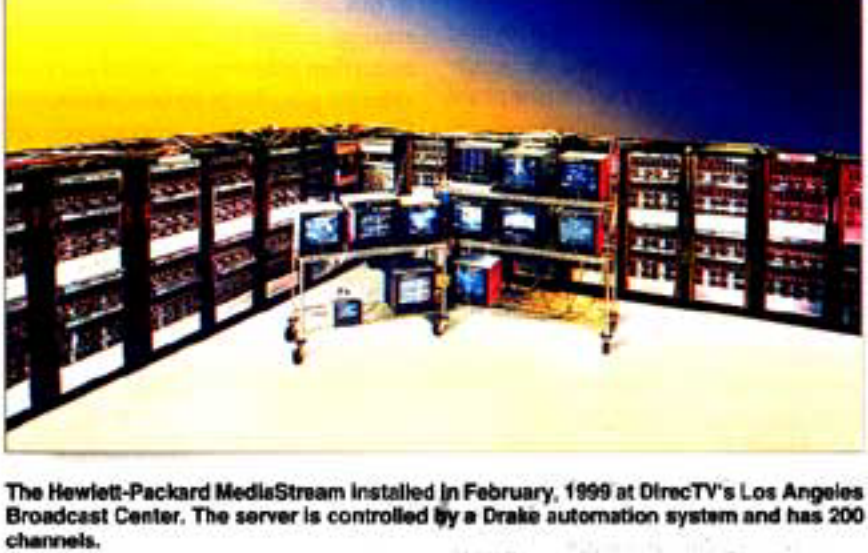
Many servers still accept NTSC video, but most immediately convert the video into one of the digital component formats. Once in the digital component domain, many systems will reduce the amount of compression needed by subsampling the chroma. 4:2:2 digital video undersamples chroma by one half vs. the luminance. However, this subsampling is only in the horizontal direction. To further reduce bit rates, some servers also undersample chroma vertically (4:2:0), reducing the active video bit rate by 25 percent. To save on bandwidth and disk storage most servers only record the active video portions of the signal. 4:2:2 has 207.4Mb/s of active video data while 4:2:0 reduces that rate to 155.5Mb/s.

Once chroma subsampling has taken place, compression is usually performed. The three main compression standards all use DCT, which converts video samples in the time or spatial domain to the frequency domain. The resulting frequency components can be scaled using scaling coefficients. The value of these coefficients effectively determines the amount of compression. Small values can be thrown away, making the compression lossy.

The next step is to order the coefficients so that lossless processes such as run length and Hoffman coding can be applied, further reducing the bit rate. This is essentially what JPEG does and what MPEG does when it produces I frames. JPEG-type compression up to 4:1 is considered essentially transparent. Video out of the server looks the same as video into the server. Compression ratios of 8:1 or even 10:1 compare with the best analog studio VTRs at a single generation. At 10:1 compression, the data rate is 20.7Mb/s for 4:2:2 video, and 15.6Mb/s for 4:2:0. MPEG takes this process one step farther by adding temporal compression to the spatial compression that JPEG employed.

Some lower-end systems store the video as GIF, TIFF, or EPS. Some organize JPEG or MPEG data into larger file structures such as AVI, QuickTime or OMF file systems. Some systems treat each clip as a separate file, while some systems create a single large file that holds all clips. The latter method is used for a couple of reasons. The first is that some systems are organized using a legacy VTR program management style. This approach has all RAIDs treated as one virtual drive. Timecode values are assigned to the entire disk storage area. The timecode is broken down into blocks with lengths such as 10 or 15 seconds. A one second clip consumes an entire block. This approach eliminates clip or program fragmentation, but enough consecutive free blocks must be available to accommodate a new clip being inserted into a vacated area. The second reason this approach is used is that it allows a simpler file structure. Video and audio are easier to synchronize. Because the file is not fragmented, average disk seek time is reduced resulting in a higher system throughput.

Whereas SCSI drives might sustain transfer rates below 1MB/s with reads of random blocks only 1K long, transfer rates of 8MB/s can be realized with 64K-long sequential block reads. Even if the read cycle was 64K blocks long, sustained transfer rates of over 2Mb/s would be unlikely if the blocks where randomly read. Regardless of the file format used,



The Hewlett-Packard MediaStream installed in February, 1999 at DirecTV's Los Angeles Broadcast Center. The server is controlled by a Drake automation system and has 200 channels.

many systems use external PC workstations to maintain a database for managing the file system on the RAID. If this database is lost or corrupted, programming could be lost. Because of this, many systems and users make use of a second RAID system to hold the file database.

Control systems

Large server systems can require fairly elaborate methods for controlling the entire process. Simple servers usually have a PC or workstation that controls encoding/decoding parameters, file management chores, and the operator/user interface. Many user interfaces provide the illusion that the server system is essentially a bank of VTRs. Larger systems can have other assets to manage and may have multiple servers to break up the workload, and to organize the workflow. Numerous router levels can be used to tie these servers together. Video, audio, timecode, and even machine control routers are often employed to integrate multiple servers and editing workstations into a single system. Besides controlling video and audio paths, many servers rely on more than one network topology. Some use Fiber Channel for transfer of video/audio between servers, while using Eth-

VIDEO SERVERS

ernet for control purposes. Ethernet can also be used for sending the metadata that describes a clip.

Workflow

Large server systems are often divided into three subsystems based on facility workflow (see Figure 1). The first phase is the gathering or acquisition of program material. Material is recorded with metadata being added, either automatically or manually by an operator. The metadata is typically stored in a database that is common to the entire system and not just the acquisition server. Acquisition servers typically have equal or even greater input than output capability. Bandwidth management usually gives the input greater priority than the output to ensure live incoming material is not interrupted. Clients requesting material from the acquisition server might experience anything from faster-than-real-time downloading to much slower than real time depending on the bandwidth needed for input. Clients requesting material from the acquisition server would be considered part of the production or manipulation subsystem.

The production subsystem is where the stories are edited. There are several uniquely different approaches to this phase. One uses shadow servers. Shadow servers record everything the acquisition server sees incoming, but at a much higher compression rate. These greatly reduced bit rate images are then sent via the LAN to client workstations requesting the material. The video/audio quality is good enough for the editor at the workstation to generate an EDL describing scene cuts and effects. When the editor is done, the control system uses the EDL to play the material out of the acquisition server to the transmission or playout server in the appropriate order. The finished piece can be moved at a rate determined by the available bandwidth. Another method is to have a client workstation directly manipulate the acquisition server to view high-quality video and audio. The local workstation still builds an EDL, and the finished piece is moved to the transmission server. However, the most common method of editing in large systems is via online editing. In this scenario, the client workstation requests a copy of the acquired video/audio then generates a finished story and ships the piece to the transmission server. In most cases, these workstations are full-blown nonlinear editors.

Transmission servers are typically symmetrical copies of the acquisition server. They give priority to playout rather than incoming feeds from production clients or the acquisition server. To make these three subsystems work together efficiently, a layer of control is needed to manage these subsystems.

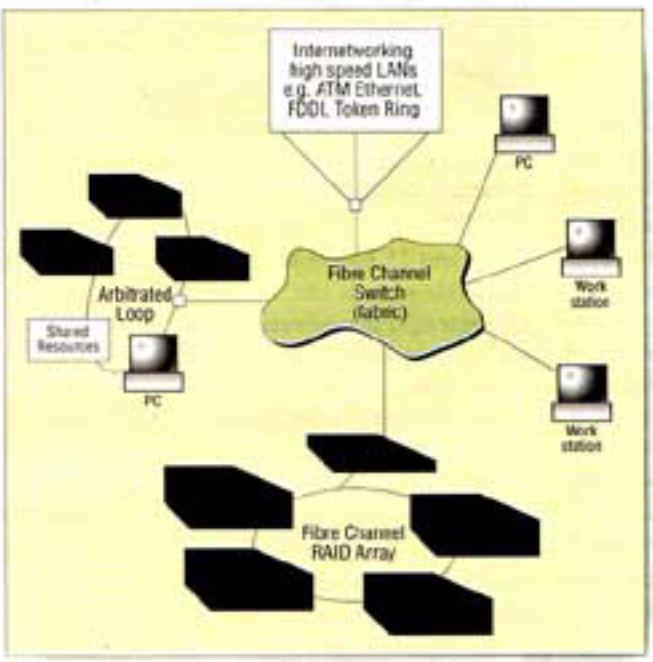


Figure 2. Using today's networking technology, a server "system" can be built to replace or supplement traditional routing

Server storage

Individual servers consist of a number of subcomponents. Disk storage is usually an external device. Almost all storage means a stand-alone RAID of some type.

Until recently, internal PC buses were not up to moving the data required for video applications internally through a computer or workstation. Today, even with new bus technology allowing much higher throughputs, that bandwidth must still be shared with internal PC housekeeping and operating system traffic. Servers that seem to house video/audio storage internal to the box usually have separate storage for the PC use and the video/audio data. To get I/O to and from the internal video/audio RAID, over the top architectures are often used. In these systems, peripheral cards plugged into the PC's expansion slots use a separate bus to move data from I/O to storage. Most video server systems have stand-alone encoders/decoders that talk straight to the storage control system, which in turn talks to the actual storage subsystem. Several methods are currently in use.

Today, Fibre Channel is a common method of communicating with the storage sub-system. The physical layer can be either coax or fiber. There are two Fibre Channel (FC) topologies: the arbitrated loop (AL) and the switched channel (S) (See Figure 2). The arbitrated loop is akin to a token ring LAN approach. At any given time one device has the loop to send data to another device. When the ring is free, devices arbitrate to use the loop. There is a set of rules as to who wins the right to use the loop. As with most things in life some devices play by the rules while some don't. With FC-AL, the loop bandwidth is shared by all devices on the bus. Switched Fiber Channel (FC-S) uses a switch, much like a router in television, to setup connectivity from one device to another. At any given time one device talks to another device, and the two devices have all the available bandwidth to communicate. The throughput is higher but the system needs expensive switching hardware to make this work. Looking physically at the boxes involved in a Fibre Channel installation won't usually tell you which topology is in use. Many FC-AL installations have every FC node connected to a concentrator that provides what appears to be a star topology, but, in reality, the loop is electrically preserved. The concentrator is used to shunt around FC devices that fail, thus preserving a closed loop. For storage applications, FC may be used to carry SCSI com-

mands. It can be thought of as a serial implementation of SCSI. The SCSI commands and data are wrapped in FC packets. Except for managing the transport and flow of FC packets, FC has no command device control language of it's own. FC can also be used to carry TCP/IP packets and thus can be used as a LAN. There are server systems that use FC in both configurations.

SSA is a point-to-point store-and-forward technology developed and mainly used by IBM in their drives. SSA allows strings of attached peripherals to transfer data simultaneously between each other and to a host. SSA uses a dual-ring topology for redundancy if one ring fails. Each node or device on the SSA ring has four ports, two for each ring. Both rings allow for 40MB/s through a node. SSA also usually carries SCSI protocol.

Also known as Firewire, P1394 is intended as a storage and peripheral linking topology. It also can carry SCSI commands. This is known as Serial Bus Protocol. There are already camcorders, CD-ROM drives, and hard drives that have P1394 interfaces. This means that video can be off-loaded from the camera to hard disk with no intervention from a PC. Currently P1394 runs at 200Mb/s, but much higher speeds are on the way. The P1394 cable consists of two copper pairs, one for data and clocking, the other for power. Although cable between nodes has been limited to 4.5m, that limit is also being addressed, as are issues regarding cable/connectors. P1394 divides its time between two modes, one geared to sending data, the other to shipping video/audio. The data mode is asynchronous where one device sends data to another node. About 20 percent of P1394 bandwidth is devoted to this opportunistic transfer of data. The rest of the bandwidth is used to send video/audio. No handshaking or flow control is used. One device simply sends the video/audio data isochronously to any and all devices interested. No devices acknowledge its receipt and there is no data retransmission.

FDDI has a token ring structure with fiber usually used instead of copper. It allows large area networks with hundreds of stations or nodes. Like SSA, FDDI allows for a second redundant ring. Interestingly, FDDI has an isochronous mode like P1394 for the deterministic transfer of data such as video.

SDTI uses the same data format as SMPTE 259M except that video data is replaced with MPEG data. The TRS signals, SAV and EAV are sent as usual. Instead of 10-bit words, the payload is eight-bit bytes. The other two bits are used to ensure that data values reserved for only TRS signals are not sent. The ancillary data space between the EAV and SAV signals contains data format information and error checking. SDTI also has a destination address sent in the ancillary space so SDTI devices downstream know whom the data is intended for. There are a number of manufacturers that use this as a mezzanine compression layer. Because it has proper TRS signals, SD digital component devices pass the signal with no problems as long as they do not process the active video. As such, production switchers and proc amps are off limits. The data pattern can even be displayed on a SDI monitor. A couple of systems use SDTI between the encoders/decoders and the RAID.

Except for managing the transport and flow of FC packets, FC has no command device control language of it's own.

HIPPI is the High Performance Parallel Interface. It is a high-speed point-to-point connection technology that operates at 800Mb/s or 1600Mb/s over cables 25m or less. It uses 4B-wide data transfers and as such, has clocks of 25MHz or 50MHz. Like SCSI, it allows one connection at a time. HIPPI has not found use in television as a storage connection, but some devices such as telecines are using this to transmit video data to computer workstations.

RAIDs

RAID makes today's server possible. The acronym RAID is said to mean either Redundant Array of Independent Drives or Redundant Array of Inexpensive Drives. The concept of the RAID array was developed at the University of California-Berkeley in 1987. RAID arrays can provide protection against data loss caused by a single drive failure. This is done by generating extra data (typically parity info) that is stored and used to regenerate lost data if necessary. The various schemes are commonly referred to as RAID levels. The various RAID schemes offer various levels of data bandwidth and protection. RAIDs can become quite large and expensive. Commercial RAIDs containing as many as 192 drives, offering 4TB of storage, are available and cost well over $500,000.

Today, most RAIDs provide a means of rebuilding the RAID online once the failed drive is replaced. Depending on size of the array, most can be rebuilt in a few hours offline. Online background rebuilding can take much longer, depending on usage. Some RAIDs can be unforgiving with operational mistakes. On some if you take a drive offline, the only way to put if back in service is to perform a rebuild of that drive. Also it should be noted that if for some reason you take a drive offline to pull it from the array, it can take 30 seconds for it to spin completely down and park its heads.

Drives now have spindle RPMs generally above 7k. Some new ones are moving into the 10k range. These units generate considerable heat. Many RAIDs have backup fans to ensure adequate cooling in the event of a fan failure. Cooling fans, being mechanical creatures (just like the drives themselves), will eventually fail, it's just a matter of when. Most RAIDs provide over temperature indications, often more than one warning. One as a warning and the second for automatic shutdown, but server vendors may override the over temperature shutdown. Drives running hotter than 50° C are undergoing an unwanted stress test. An interesting subtlety in some RAIDs is that the internal SCSI bus is often activity terminated by the last drive in the chain. If that is the case you will disrupt activities of all drives in that chain if the end drive is turned off or removed.

Servers are one of the most obvious examples that the computer and television industries are indeed converging. Besides grasping the new technology required to understand and manipulate digital video, today's television engineers need to come up to speed on the computer industries methods of processing, controlling and storing television data. ■

Jim Boston is an engineer at KICU-TV, San Jose, CA.