
Everyone wants to record CCTV video. And although everyone wants a reliable recording system, the truth is that this reliability is based on a number of factors. This white paper briefly examines the key factors for centralized storage and compares the most common options. It does so by answering several simple questions, not all of which have simple answers.

Question: What's wrong with using the hard drives inside my DVR?

Nothing. If it meets your needs, then it may be the most economical solution and quite likely, the easiest to set up. Your typical risk is that in the event of a drive failure, you will lose the footage stored on that drive. You experience this result with Bosch's Divar MR or a Bosch VideoJet with a directly attached USB hard drive or embedded CF card.

With multiple drives that have been linked together as a JBOD (Just a Bunch of Disks), you may find that certain cameras write to certain drives. In this case, several cameras' recording will be affected and others' will not.

Question: What about RAID inside the DVR box?

Some DVRs, such as Bosch's Divar XF hybrid recorder, offer built-in RAID. With this capability, one drive can fail without any loss of video, assuming it is repaired (disk is replaced) before a second drive fails. You get the same advantage when using a recorder as a DVR or NVR.

Question: What about RAID outside the DVR box?

Other DVRs, such as Bosch's DiBos hybrid recorder, use directly attached SCSI disk arrays, which are typically configured as RAID 5 devices. This method generates larger amounts of storage and the option to buy your storage independently from the DVR provider. You also get the same benefit when using a recorder as a DVR or NVR.

Question: What about recording to a remote RAID?

Some recorders are able to stream their video to some kind of networked storage. This ability may be desirable when the recorders are distributed, but the storage is centralized. Some recorders including Bosch's DiBos hybrid recorder are able to record on any networked storage that can appear as a mapped drive i.e. can be assigned a drive letter such as "E:".

Additionally, some recorders, typically NVR servers, can send their video directly to an iSCSI disk array by activating the iSCSI initiator inside the NVR server PC. Because this task involves encapsulating SCSI commands inside IP packets for transmission over the network, it is processor-intensive and may conflict with the NVR's primary duties. In these circumstances, it is possible to delegate this onerous task to a TCP Offload Engine or TOE, which is a piece of dedicated hardware that easily handles the job previously assigned to the operating system software.

It is important to remember the distinction between this use of iSCSI disk arrays and Bosch's Direct-to-iSCSI architecture. With the latter, the iSCSI initiator is built directly into the IP camera or IP encoder so that no NVR or hybrid recorder is necessary – the edge device streams directly to the iSCSI RAID.

Question: Why bother with RAID?

Back in the '80s, RAID stood for "Redundant Array of Inexpensive Disks" and was a way for smaller computer systems to achieve the kind of storage reliability reserved for very expensive and high-performance mainframes. RAID achieved this range of reliability through various levels, each one delivering a unique combination of Data Reliability (or failure protection), Read Performance and Write Performance. By compensating for relatively inexpensive – and therefore less reliable – drives with intelligent software in the form of RAID controllers, RAID storage manufacturers have given us 5 important RAID levels.

Before examining them, it is important to understand the concept of “parity”. Parity is redundant information – extra data that is never used unless something goes wrong. In an event where data may have been lost or corrupted, the parity information is your insurance policy, allowing you to automatically recreate the missing data as if nothing happened.

RAID 0. Video is striped across all the drives.

- Pros: All drives are used equally, which improves read and write performance. Also, no disk space is lost for parity data so you achieve full disk utilization.
- Cons: There is no parity information. Any disk failure results in the instant loss of all video on the entire RAID. For this reason, it is a step above JBOD (where some drives get used more than others), but offers little value for recording video.

RAID 1. Drives are divided into 2 mirrored sets. Video is striped across all the drives.

- Pros: Extremely fast write performance. Will survive a single disk failure.
- Cons: You lose half your disk space for redundancy. Consequently, it is rarely, if ever, used to record video that is particularly hungry for disk space. However, it is extremely appropriate for storing the OS for a DVR.

RAID 4. Video is striped across all the drives except for one designated parity drive.

- Pros: Will survive a single disk failure and only one drive is “lost” as a parity drive, which offers a massive storage utilization improvement over RAID 1. It is also computationally economical to implement, especially compared to RAID 5 (see below).
- Cons: In *conventional* RAID 4, the parity drive is hammered because every time you write to any of the other drives, you have to write to the parity drive. If your RAID has 12 drives, you’re driving the parity drive to the end of its life 12 times faster. It offers little advantage for recording video compared to its big brother, RAID 5.

RAID 5. Video is striped across all the drives. Parity is striped across all the drives.

- Pros: Will survive a single disk failure and all drives are used equally. This quality makes it the perfect compromise of load balancing (wear all drives equally) and single parity reliability (only one drive’s worth of capacity is lost for parity information). The hero of modern video storage, it beats RAID 1 and *conventional* RAID 4.
- Cons: Calculating and then distributing the single parity across all drives is processor-intensive. It is expensive to write data since RAID 5 is not optimized for writing data. CCTV is all about recording video; therefore, almost all writing and negligible reading poses a problem. This is bad enough with RAID 5 where you are protected against a single drive failure – with RAID 6, it becomes a terrible challenge (see below) because you are seeking protection against a double drive failure.

RAID 6. Video is striped across all the drives. Double parity is striped across all the drives.

- Pros: Will survive a double disk failure and all drives are used equally.
- Cons: Implementations based on RAID 5 implementations, as well as calculation and distribution of the double parity across all drives, are crippling on processor performance. These processes are so challenging that distributed parity RAID 6 is almost never offered in disk arrays – the price is prohibitive. The consequence is that double drive failure protection is essentially unavailable to the masses. Instead, manufacturers recommend RAID 5 with a hot spare. This method delivers the same net capacity, but does not survive a double failure during the prolonged rebuild.

Question: Who wants double drive failure protection?

Almost nobody wants double drive failure protection because the chances of two drives mechanically failing have historically been very small.

The point is that it's the wrong question.

Over the years, the mechanical reliability of drives has not changed. However, in line with Moore's Law's prediction, their capacity has doubled every 18 months. 2TB SATA drives are already widely available whereas 250GB drives were common only 4 years ago. The bottom line is that we know that if we lose data, we're going to lose a lot of it.

Secondly, drives don't fail because they crash, break or blow up. In fact, the major threat to reliability is a read-write error. The disk controller thinks it wrote one thing when in fact, for various rare reasons, something completely different was written. RAID's are designed to mark a sector as failed if this event occurs. If you're wondering why this is the case, try to imagine your bank account as part of the corrupt data.

With SATA drives, this event statistically happens one bit in 10^{14} . 10^{14} bits or 100,000,000,000,000 may seem like an infinite amount, but it is equal to only 11TB. In the world of CCTV where vast amounts of data are constantly being written, a disk array typically gets written over every 7 days. Which means today's drives frequently experience read/write errors in CCTV applications.

The good news is that RAID manufacturers know this fact. So when such a failure occurs, they use the redundant parity information to correct the error without you even noticing. One challenge is that the RAID does not correct the error as soon as it happens; rather, the RAID does so when it happens to be in the vicinity, trying to do another write. Therefore, the error may be detected at an unpredictably delicate moment, possibly a moment when you can't afford to realize it.

The bad news is that if you have a failed drive, until you have discovered it, replaced it and given it 12-24 hours to rebuild, you lose all data on the entire RAID. This period of vulnerability, which exponentially worsens with larger drive sizes, is the reason why those who cannot afford to lose video need double drive failure protection.

Question: What are the options for double drive failure protection?

Hot Spare. This is a hot stand-by unused drive in the disk array that leaps into action as soon as a failed drive is detected. This option shortens the vulnerable period by eliminating the "wait for a human to discover the failure" and "wait for the hard drive to be swapped out" steps.

- Pros: Can be used with RAID 5 disk arrays and does not require additional horse power when in normal operation.
- Cons: In any networked environment, devices such as RAID's should be constantly monitored by software that instantly alerts someone by email or SMS when there is a problem. In these circumstances, the Hot Spare solves a problem that largely does not exist. Also, you still have the rebuild period of 12-24 hours – during this period, the RAID is under heavy stress, so performance can be degraded and a second failure can lead to total loss of video.

RAID 6. Video is recorded with double parity.

- Pros: Protects against a double failure.
- Cons: Too processor-intensive and therefore not affordable to the masses.

RAID-DP. NetApp's RAID 6 that uses double parity arranged on dedicated parity drives.

- Pros: Protects against a double failure. Needs only 10% more horsepower than NetApp's single parity solution (RAID 5 solution). For more information, please visit http://en.wikipedia.org/wiki/Non-standard_RAID_levels#RAID-DP.
- Cons: Reduced net capacity per RAID as is the case with RAID 6 or RAID 5 with a hot spare.

Question: Is RAID-DP the same as RAID 6?

Yes. The SNIA (Storage Networking Industry Association) defines RAID 6 as: "Any form of RAID that can continue to execute read and write requests to all of a RAID array's virtual disks in the presence of any two concurrent disk failures. Several methods including dual check data computations (parity and Reed Solomon), orthogonal dual parity check data and diagonal parity have been used to implement RAID Level 6."

RAID-DP is NetApp's implementation of RAID 6. Because it is so efficient, it has brought RAID 6 reliability to the masses.

Question: What is NetApp's RAID 5 implementation called?

The benefit of RAID 5 is protection against a single drive failure. RAID 5, by definition, spreads the parity across all the drives, allowing all the drives to be load balanced.

Traditional RAID 4 provides the same benefit. But because of the dedicated parity drive, it wears out faster than the data drives – it is not load balanced.

NetApp developed a RAID 4 solution that used a concept called "stripe at a time" where data is buffered in non-volatile RAM until a single neat stripe of data (and parity) can be written equally across all drives. NetApp delivered the benefit of RAID 5 with the computational efficiency of RAID 4.

This characteristic makes it a more efficient solution. But much more importantly, in order to achieve the impressive RAID-DP performance, NetApp built it on top of their Stripe-at-a-Time RAID 4 by simply adding one extra dedicated parity drive. The extra computational power required is negligible, but you get the otherwise hard-to-reach benefit of RAID 6.

Without NetApp's Stripe-at-a-Time RAID 4, RAID-DP could not have been accessible to the masses.

Question: Is Stripe-at-a-Time RAID 4 unique to NetApp ?

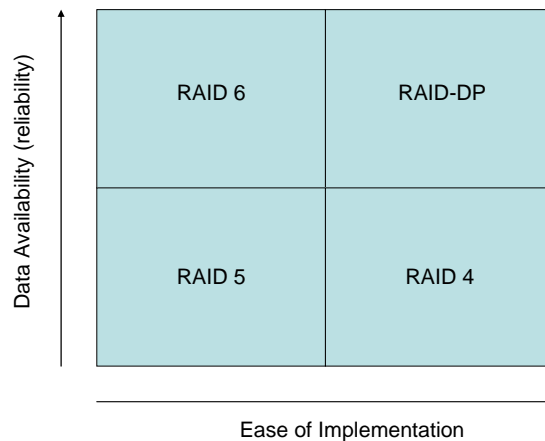
No. In fact, Bosch's Divar XF hybrid recorder has embedded RAID 4 and writes a stripe at a time. Which means it also achieves the same RAID 5 availability while maintaining load balancing across all the drives. The Divar XF does so to allow the hybrid recorder to focus its horsepower on its main job – to compress, record and playback video. And lots of it.

Question: Which one should I use ?

Bosch and NetApp have collaborated to bring to market iSCSI RAID's that can be configured for either RAID 4 or RAID-DP.

RAID 4 is more appropriate to use when you need RAID 5 protection, but need to maximize your usable capacity.

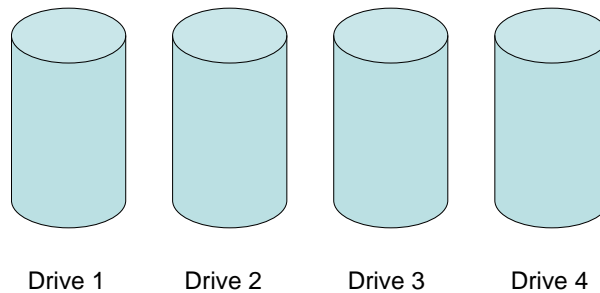
RAID-DP is preferable when you want to maximize your availability as drive sizes continue to grow. It is possible to use RAID 4 with a hot spare. However, the usable capacity is the same as for RAID-DP because in both scenarios, 2 drives are lost. Also, the availability is lower because you cannot survive a double failure during rebuild. Consequently, RAID-DP is invariably preferred over RAID 4 with a hot spare.



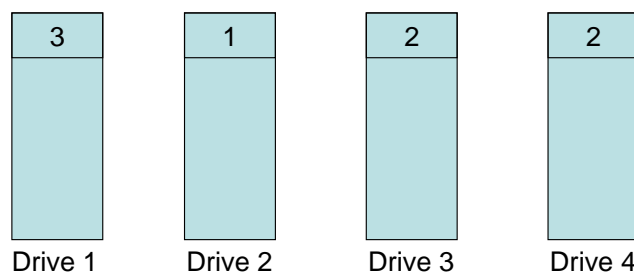
Question: Can you explain parity in plain English?

Parity is a cool name for redundant information – extra data that we use as an insurance policy in case we lose a hard drive. If we lose a drive, we can fill in the missing gaps and not lose any data, or in the case of CCTV, video.

Imagine a disk array with 4 hard drives.

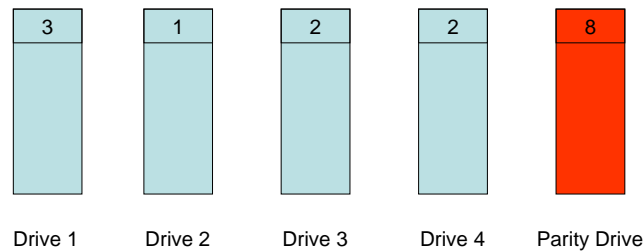


If we write a stripe of data across all of them, which we do to avoid overloading any one particular drive, we get something like this:



Here's the problem: When a drive, say drive 2, fails, everything on it is lost. We lose the data "1" forever. This is essentially a RAID 0 implementation, which is really a misnomer – there is no redundancy, yet the "R" in RAID stands for Redundant.

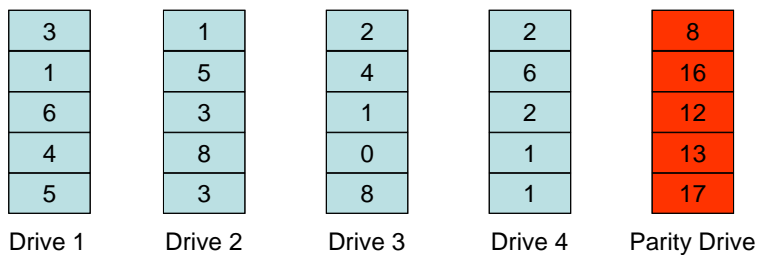
Instead, imagine we have a parity drive (this is RAID 4) indicated by the red drive. We can manually sum the data on the other drives and save the answer on the parity drive.



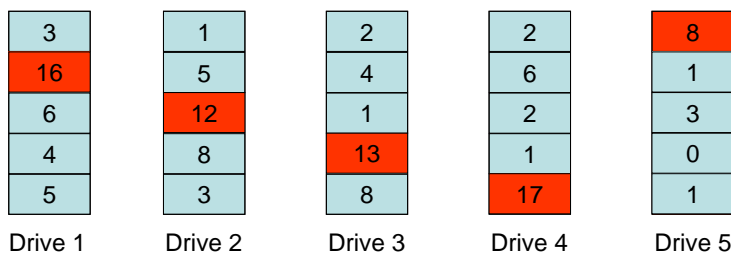
So, $3+1+2+2 = 8$. So the parity for that row is "8". Now, if drive 2 fails, we can replace it with a fresh new drive and "reverse engineer" the missing data. This will be $8 - (3+2+2) = 1$.

In the real world, parity is not exactly calculated using simple summation, but the concept is the same.

For completeness, here are 5 rows of data.



The example above assumes a dedicated parity drive, which is RAID 4. In comparison, with RAID 5, the parity is spread across all the drives, known as distributed parity.



Both RAID 4 and RAID 5 use the concept of single parity or one red block per row. With this concept, we can lose one drive without losing any data. RAID 6 and NetApp's implementation, RAID-DP, introduce another "red block" per row, giving us twice the protection. For an excellent explanation of how RAID-DP works, please visit http://www.netapp.com/us/library/white-papers/wp_3298.html.