Planning for the Worst

Redundant Power

A redundant power supply is simply an enclosure that provides two or more complete power supplies inside of one. You learned about this back in your A+ studies. Now, most servers are going to utilize two individual power supplies in the server's case to ensure that power is always available to that server at all times. This eliminates a single point of failure that really does exist inside your desktop computer. If you look in your desktop computer, you only have a single power supply. If it fails, the entire computer is going to fail with it. But, with servers, we have two power supplies, those redundant power supplies. And so, this is going to ensure that we can mitigate the threat of power supplies failing on us and taking down the entire server or system.

Surge

A surge in electrical power means that there is an unexpected increase in the amount of voltage that's being provided. So, here in the United States, our power supply is 120 volts. If that went up to, say, 124 or 125 volts, that's only a little bit of an increase in power. So, that would be considered a surge.

Spike

Now, a spike is going to be a short transient voltage that's going to be due to a short circuit, a tripped circuit breaker, a power outage, or even a lightning strike. This might jump from 120 volt up to maybe 140 or 150, or even more. Now, to protect against a surge or a spike, you should use a surge protector. A surge protector is going to help you against those little surges, but if you have a really good surge protector, it can help when you have those large spikes, as well.

Sags

A sag is kind of like a surge, but in reverse. Where a surge went up, a sag is an unexpected decrease in the amount of voltage provided. So, it's going to go down.

Typically, sags are only for a short duration of time, and usually, it's not even going to make the power get lost to your computers.

Brownouts

However, when that voltage reduces for longer than that, it becomes known as a brownout. A brownout is when a voltage drops to such an extent, that usually your lights start dimming and your computer would even shut off.

Blackouts

Now, a blackout occurs when there is a total loss of power for a long period of time. So, if you're sitting in your house and the lights all go out and the computers turn off and it happens for 30 seconds or a minute, that's considered a prolonged amount of time for a blackout.

Backup Power

In the last lesson, we discussed the different types of power conditions that can affect our systems. Now, to mitigate these, we used different forms of backup power.

UPS

The first type of backup power we have is called a UPS or an Uninterruptible Power Supply and this is going to combine the functionality of a surge suppressor with a battery backup. Now, the great thing about these is they can also provide line conditioning. So, they can protect against things like brownouts, sags, and surges. So, if you have short durations of times where the power goes down or goes up, that line conditioning function can help protect your machines and keep them running smoothly. Now, backups like these are good for short durations of time but they usually don't last more than 15 or 30 minutes. Some of the largest ones I've seen and most expensive ones I've seen can actually last up to about 60 minutes.

Backup Generator

A backup generator is part of an emergency power system. It's used when there's an outage of your regular power supply for the electric grid. Now, some emergency power systems might include things like lighting for your hallways or special fuel cells. Larger commercial backup generators can actually power the entire building or large portions of it. Now, it just depends on how much fuel you have available and how much power you need to generate. There's really three types of generators that

we're going to talk about. There's portable gas-engine generators, permanently installed generators, and battery-inverter generators.

A portable gas-engine generator is the least expensive type to run and it usually uses gasoline or sometimes even solar power. These tend to be noisy when they're gas engines. They're have high maintenance and they have to be started manually and you'll usually plug in an extension cord and run that into your building.

Now, when you start going to larger generators, you start talking about permanently installed generators. These are much more expensive and much more complex to install. But they're always there. Generally, these will run on natural gas, propane, or diesel fuel. They tend to be quieter and they can be connected directly to your organizations' electric panel. So, if you lose power to the building, somebody can go and turn on this generator and bring it back online.

The third type of generator that we have is known as a battery inverter generator. These are based on lead acid batteries. They're super quiet and they require very little user interaction, aside from, maybe, an uncommon restart once in a while and changing out the batteries every couple of years. They are well matched to environments that require a low amount of wattage or are the victims of short power outages only. They can't withhold your whole facility or data center for a long period of time. But if you combine something like the battery generators with the diesel commercial generators, you can actually have the battery take over the short period of time and the diesels take over the long period of time.

Data Redundancy

If you remember, a RAID is a redundant array of independent disks which is essentially going to allow you to combine multiple physical hard disks into a single logical hard disk drive inside of the operating system. Now, for the Security+ exam, you need to know about a couple of RAID types including RAID 0, RAID 1, RAID 5, RAID 6, and RAID 10.

RAID 0

A RAID 0 provides data striping across multiple disks and is used to increase your performance. The keyword here is striping. For example, you might use a RAID 0 when you need performance but you don't care about fault tolerance, so, a good example of this is I do a lot of video editing and so, I really care about performance there as I'm editing these raw videos and so, by having these two drives working

together, I can do things much quicker than I could with a single drive. Now, to do a RAID 0, you do need at least two disks to work in tandem with each other.

RAID 1

The next one we have is a RAID 1, and this is going to provide redundancy by mirroring the data identically to two hard drives. So, if one drive fails, the other can continue to operate because it has a full copy of everything that was on there. This provides the least amount of downtime because there is always that complete copy of data ready at a moment's notice to take over. This provides wonderful fault tolerance, but it can only be used with two physical hard disks and that provides you with one single logical hard disk inside the operating system. A good example of this is once I'm finished editing all my videos and I have the final product, I want to make sure I don't lose it, so, I can actually move that over to a RAID 1 where I get two identical copies of that file one on each of those drives.

RAID 5

Now, the next one we're going to talk about is a RAID 5. A RAID 5 is known as a striping disk with parity. It requires at least three physical disk drives to work, and it provides fault tolerance by striping the data across multiple disks and writing parity data to the multiple disks, too. If one disk fails, the other two can reconstruct the data based on the parity and they continue to operate. This means that if one of those drives fails, I can pull it out, put in a new drive, and it will rebuild itself inside the RAID as it keeps moving and operating for the rest of the system.

RAID 6

Next, we have a RAID 6, and a RAID 6 is a modified form of a RAID 5. In fact, it's one better than a RAID 5 that's why we call it a RAID 6. Now, it's going to use data striping across multiple disks just like a RAID 5 did, but instead of having one stripe for parity data, it's actually going to have two stripes for parity data. This requires another disk in the array to work so, you need at least four physical disks, but that does provide you additional fault tolerance because you can lose up to two of these four disks and the RAID will still continue to function.

RAID 10

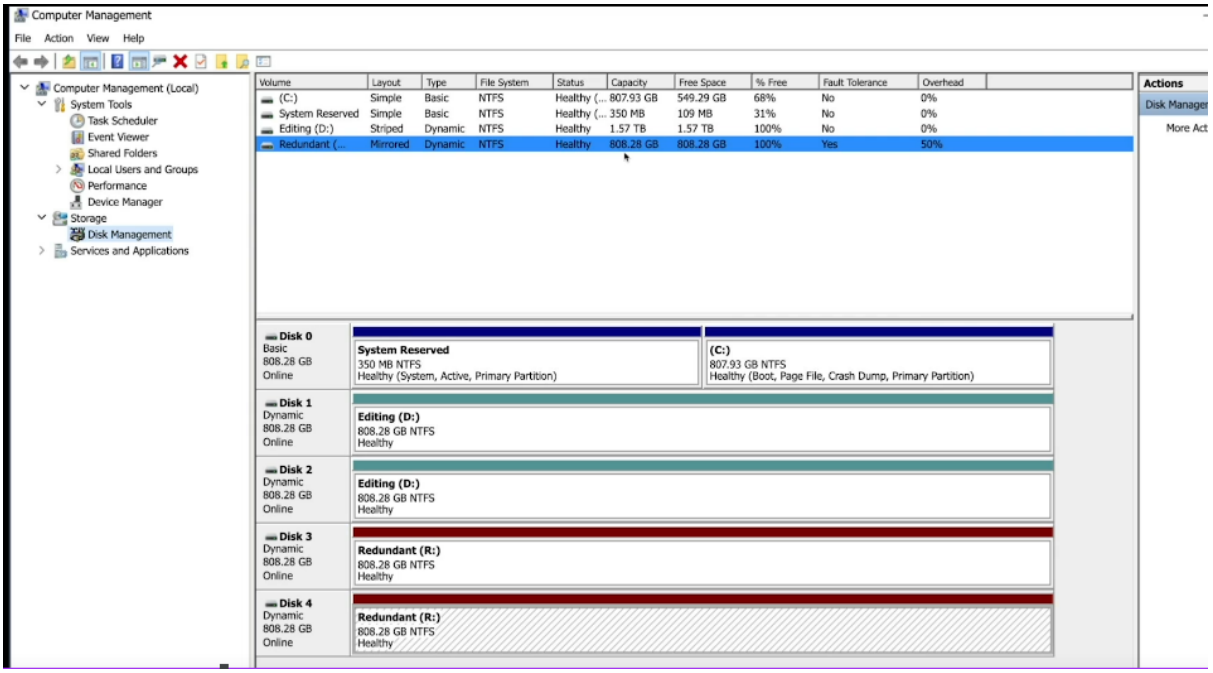
The last RAID we have is known as a RAID 1-0 which is written as RAID 10. This combines the advantages of a RAID 1 and a RAID 0 because one plus zero equals 10. This requires four physical disks, just like a RAID 6, and it's going to provide you with a redundant mirror of striped drives and it is fully fault-tolerant. This gives us all the speed of a RAID 0 by splitting up the load across two sets of RAIDs, but it also gives us the full redundancy of a RAID 1 by having those two RAID 1s in there. Now,

this all works as one combined logical drive, even though it uses those four drives split up into two pairs of two.

So, when we think of RAIDs, they can be categorized as failure-resistant, fault-tolerant, and disaster-tolerant. These are our three categories for RAIDs. Now, if you have a failure-resistant RAID, that's going to be something like a RAID 1 or a RAID 5 because it's going to protect against the loss of the array's data if a single disk fails inside of it. Now, when we talk about fault-tolerant disk systems, this will be something like a RAID 1 or a RAID 5 again, or even RAID 6, because even if a single component fails, one of those drives or even one of the cards inside of it, then that RAID can continue to function properly. Now, our final category is known as disaster-tolerant, and so, if we call a RAID disaster-tolerant, this means that the RAIDs has two independent zones with full access to the data at all times. A RAID 10 is a good example of a disaster-tolerant RAID.

Features	RAID 0	RAID 1	RAID 5	RAID 6	RAID 10
Minimum number of drives	2	2	3	4	4
Fault tolerance	None	Single-drive failure	Single-drive failure	Two-drive failure	Up to one disk failure in each sub-array
Read performance	High	Medium	Low	Low	High
Write Performance	High	Medium	Low	Low	Medium
Capacity utilization	100%	50%	67% – 94%	50% – 88%	50%
Typical applications	High end workstations, data logging, real-time rendering, very transitory data	Operating systems, transaction databases	Data warehousing, web serving, archiving	Data archive, backup to disk, high availability solutions, servers with large capacity requirements	Fast databases, file servers, application servers

Demo



Network Redundancy

Network redundancy is focused on ensuring our network remains up and running at all times to increase its availability. This includes the server's connections to the network, the connections between our switches and our routers, and our connections to the Internet. To accomplish this, our servers often have two or more network interface parts, and they can be operated as a pair or in a low balancing configuration. Now, this can paired for redundancy, or you can split them and put them into two groups so you can have additional throughput by operating more than one at a time, simultaneously. In this example, I might create two groups, one with four network cards in it and one with a single network card by itself. That gives me the total of five. Four network cards can operate in a shared load capacity, meaning, that all four will work together. So, if they were each 100 megabytes per second each, that gives me a combined 400 megabytes per second, but that fifth one is sitting there by itself and it has 100 megabytes per second to use. We use that as our backup redundancy.

Server Redundancy

To create redundancy for our servers, we're going to use a concept known as clustering. A cluster is when you take two or more servers and have them work together to perform a particular job function. We can cluster our servers as either failover clusters or load-balancing clusters.

Redundant Site

What’s a redundant site? Well, let's consider if your office building was flooded. You're going to need a new place to work. That's a redundant site. Now, if you're really concerned about downtime, you could have a redundant site up and ready to go at all times. Redundant sites are classified as one of three categories. They're either hot sites, warm sites, or cold sites.

Now, a hot site is a near duplicate of your original location. It's going to have it where the organization can move in and be up and running within minutes. That means, they have servers, phones, desks, lights, power, connectivity, everything. It's just as if you've picked up and went to a different building that day.

A warm site, instead, is going to have some of those capabilities, but not all of them. It's going to have things like the computers and phones and servers there, but they may not be configured or patched or updated. And so, when people show up, you're going to have to install their user accounts or set up their configurations and things like that before the users can start working.

Now, a cold site is going to have things like tables and chairs and bathrooms and maybe some technical setup like basic phones and data and electric lines. But it doesn't have computers, it doesn't have servers. And none of it is configured. And so, if you have a cold site, that might take you a couple of days to get back online.

Data Backups

Data backups can be conducted using full backups, incremental backups, or differential backups.

1. Now, the first kind we have is known as a full backup. When you do a full backup, all of the contents of your drive are backed up, that's every single file.

2. Now, when we go to an incremental backup, which is our second type, this is going to back up only the contents of the drive that have changed since the last full backup, or since your last incremental backup.

3. Differential backups will only back up the contents of the drive that have changed since the last full backup.

Now, you can see the tough choice we have between using incrementals and differentials, because incrementals take a lot less time to backup, but the differentials make it a lot quicker when we need to restore, and so, this going to be a choice you have to make. Do you want quick backups and lengthy restores, or do you want long backups and quick restores?

Tape Rotation

Now, there are three main rotation schemes that we're going to cover in Security+. We have the 10 tape rotation, the grandfather-father-son, and the towers of Hanoi.

10 tape rotation

Now, the 10 tape rotation is a simple method that provides easy access to the data that's been backed up. It could be accomplished during a two-week backup period. Why is it called a 10 tape rotation instead of a 14 tape rotation? Well, because most companies are open Monday through Friday, so, weekends don't count. If you wanted to do it seven days a week for two weeks, you could simply call it a 14 tape rotation instead and use 14 tapes. Either way, the concept is the same. Each tape is going to be used once per day for two weeks, and then, the entire set is reused again. This means, after two weeks, you don't have any more backups, though.

Grandfather-father-son

Now, the second method is known as a grandfather-father-son, and it's a backup rotation system that's very commonly used. It's actually one of my favorites. When attempting to use this design, there are three sets of backup tapes that have to be defined. Usually, we call these the daily, the son; the weekly, the father; and the monthly, the grandfather. These tapes are then rotated on a daily basis, and the last one of the week will be graduated to father status. Then, these tapes, the weekly ones, are then rotated on a weekly basis, and after four weeks, they become the monthly, or the grandfather, and that is how we get our grandfather, fathers, and sons. Generally, your monthly tapes are kept offsite, and this will allow you to ensure

that they're safe in case of an emergency at your regular facility or site. I mean, it would be horrible if your site burned down and you couldn't do any backups because all your tapes were in the server room, right? So, you want to make sure you have some good offsite backups, as well.

Towers of Hanoi

Now, the third type we have is called the towers of Hanoi, and this a rotation system that's based on the puzzle called the towers of Hanoi that you might've played as a kid. Much like the grandfather, father, and son, this system also uses three sets of backups, but they're rotated a bit differently. Basically, your first tape is used every second day, and the second tape is used every fourth day, and the third tape is used every eighth day, and so, this system helps prevent tapes from being worn out as quickly as the 10 tape rotation does, and it does allow for three different categories of backups like the grandfather, father, and son method, but because of this complexity, it makes it harder to remember what tapes do I use to backup and in which order, and then, when I go to restore, I have to figure that out, as well.

Snapshots

With a snapshot, all of the applications, the hard drives, and even the operating system is backed up to create a full backup of the system as a virtual disk image. This makes it very quick to redeploy that system onto a cloud server or another offsite location, but it does take up a lot of storage space, so, you need to plan for that extra storage resources and costs that are involved in using snapshots instead of regular backups. Snapshots are also commonly used with virtualized systems, so, if you're running VMware or VirtualBox, you can take a snapshot of your server and create that full backup of the entire virtual system.

Disaster Recovery Planning

Disaster recovery planning is a development of an organized and in-depth plan for problems that could affect the access of your data or your organization's building. So, if you think about the fact that you might have a cyber attack, or a flood, or a fire, all of these things might be things that are covered by your disaster recovery plan. Now, planning should also include information regarding redundancy, such as what sites you have, are they warm sites, cold sites, or hot sites. How your backups are

done and where they're going to be restored from, but it shouldn't include any information that deals with day-to-day operations of your organization.

Now, a good disaster recovery plan should always be written down, it shouldn't be here in my head. Everyone in the organization needs to know what those policies are. We should have clearly outlined disaster recovery policies, procedures, and information.

Business Impact Analysis

Now, when I talk about a business impact analysis, this is also abbreviate as a BIA. This is a systemic activity that identifies organizational risks and determines their effect on ongoing mission-critical operations.

Now, when we start talking about these metrics there are lots of different ones we have to consider. We have things like our Maximum Tolerable Downtime or MTD. We have Recovery Time Objective RTO. The Work Recovery Time WRT, or Recovery Point Objective RPO.

MTD

Now, when we talk about a Maximum Tolerable Downtime or MTD, this is the longest period of time a business can be inoperable without causing irrevocable business failure. Essentially, how long can you be down without going out of business? Now, the MTD is going to be different for each organization and even within each organization, each of your business processes can have its own MTD. For example, some may be just a couple of minutes or a couple hours for critical functions. You may have up to 24 hours for urgent functions and up to seven days or longer for normal functions.

RTO

Now, the next one we want to talk about is our RTO. This is our recovery time objective. Now, this is the length of time it takes after an event to resume your normal business operations and activities. When you start thinking about recovery time objective, I want you to think about the fact of something went down. We lost power. How quickly do you need it back? In my case, we have a 60-second time for power. We want to make sure our power is back up and online within 60 seconds. Now, is that achievable? Yes. If you have a backup diesel generator, it will turn on in about 45 seconds and transfer power to the diesel generator. Now, my wife wasn't

happy with 45 seconds or 60 seconds and she wanted a recovery time of zero. Now, can I achieve that? The answer is yes. And that's one of the reasons why we have those battery backup systems. Because if power goes away, those batteries come on instantly. There is zero lag time there. And so, we're able to hit a recovery time objective for power of zero seconds. Now, the overall power of getting it back to the grid, we can't control that. That's up to our local power company. But we can make sure that we can recover our business and make sure we're on battery, on solar, or on generator within zero seconds.

WRT

Now, the next one want to talk about is work recovery time or WRT. This is the length of time in addition to the RTO of individual systems to perform re-integration and testing of a restored or upgraded system, following an event. So, let me give you an example. Let's say in my organization, we had a power outage and we didn't have the batteries yet. We had to rely on those generators. So, we had a 60-second recovery time. Well, in 45 seconds, power comes back up. But if those systems went down because of a power surge, and I had to replace one of my servers, well, that's going to take additional work recovery time. I fixed the main problem, the recovery time objective of getting the power up, but now, I have to fix the second and third order effects to get work product going again, which might be rebooting your computer. It might be rebuilding a computer. It might be replacing a hard drive. Whatever those things are, I have to perform that re-integration and testing to bring those systems back online in an upgrade or restored state to be able to get us back to regular work product.

RPO

And our final one we want to talk about is RPO. This is our recovery point objective. This is the longest period of time that an organization can tolerate lost data being unrecoverable. Now, the way I like to think about this one, when I think about RPO is think about ransomware. If you have ransomware on a system, it's going to encrypt your files. Now, you've got a couple of choices here. You can pay the ransom, which we never recommend. You could try to crack the ransomware key, which could take you days, weeks, or months, or years, depending on how strong it is, or you can actually wipe that system and recover from a known good backup. Well, that's great. Let's go ahead and choose that option. Well, if we do that, what is the longest period of time that we can tolerate data loss? Well, there's going to be time that we're going to be lagging as we're recovering all that data back. And that data may have several hours since it was last backed up. For instance, if you run your backup once a day at midnight, that's when your data was backed up. And if this ransomware hits you at six in the morning, you have six hours worth of lost data because you don't have a backup of that. This is what we're talking about when we're talking about recovery

point objective. That six hours is going to be lost period of time. And so, if your RPO was 12 hours, that's fine. If your RPO was four hours, you've just broken your RPO.