WHITE PAPER · ENTERPRISE NETWORK ARCHITECTURE

Rainy Day Bandwidth

The economics of WAN redundancy, and designing availability to business requirements

Providers discount the network half of a backup connection and charge full price for the local access circuit — which was 75–80% of our WAN cost. I started calling the result rainy day bandwidth: billed 8,760 hours a year, used for six. Here is how we priced it, and what we changed.

BY **JACK NASH** ~70% WAN COST REDUCTION ABOUT 11 MIN READ

THE ARGUMENT IN BRIEF

- Providers discount the network portion of a backup connection, but **somebody still has to sell you the local access circuit** — and that was 75–80% of our WAN cost.
- So a “discounted” backup ran **$1,500 against a $2,000 primary**. Across hundreds of sites, standby capacity became one of the largest line items in the network.
- I started calling it **rainy day bandwidth**. Billed 8,760 hours a year; at many sites, used for a handful.
- The useful measure isn’t how often it’s used. It’s **cost per avoided hour of downtime** — $18,000 a year against six hours of outage is $3,000 an hour, which is obviously right for a plant and obviously wrong for a sales office.
- Matching availability to what each site actually required, rather than duplicating everywhere, contributed to a **WAN cost reduction of roughly 70%** without giving the business less than it needed.

PART I **The Standard and Its Cost**

Two of everything

Traditional enterprise WAN design followed a principle that seemed too obvious to examine: if a connection is important, buy two of them.

Large sites received redundant circuits. Critical locations received redundant routers and firewalls. MPLS providers offered backup ports and services at discounted rates, which made the whole arrangement look economically sensible. Redundant power supplies, redundant carrier paths, sometimes redundant providers.

For genuinely critical locations, all of that can be exactly right. The problem was that redundancy had quietly become a design standard rather than a business decision.

If two circuits were correct for the data center, two circuits must be correct for the regional office. If redundant routers suited a major manufacturing facility, surely every significant office should have them. Over time the availability architecture came to rest less on quantified business requirements and more on a general conviction that downtime should always be eliminated.

Eliminating downtime has a price. On a global network, that price is very large.

Where the money in an MPLS service actually is

The observation that reframed this for me was where our WAN spend actually sat. The provider's MPLS backbone was not the expensive part. The local access circuit connecting each location to the provider edge was — roughly 75 to 80 percent of our total MPLS WAN cost.

That matters enormously when evaluating redundancy, because of what a provider can and cannot discount.

A provider could offer attractive terms on the network portion of a secondary connection. The MPLS port might be discounted. Class of Service charges might be reduced. Other provider services might carry favorable backup pricing. All of it genuine.

But someone still had to supply the physical access circuit connecting that building to the network. The local loop was not cheaper because we intended to use it only during an outage. The trench, the pair, the last-mile provider — none of that cares about our intentions.

So the economics came out something like this:

PRIMARY	$2,000 per month — full service, in use every day
BACKUP	$1,500 per month — discounted network, full-price local access, idle

The backup service was genuinely discounted. Economically, we were still spending an additional $18,000 a year at that location for something we hoped never to use. Multiply across hundreds of locations in dozens of countries and standby capacity becomes one of the largest components of the network.

Rainy day bandwidth

I began thinking of those secondary circuits as rainy day bandwidth. Bought for an emergency, sitting in the corner, waiting for a day that mostly never came. It did nothing most of the time, and we paid for it every month.

That alone does not make it wasteful. Insurance also sits unused, and insurance can be excellent value. Which means the obvious question is the wrong one.

> **The question is not how often we use the backup circuit. It is what business loss the circuit prevents, and whether that avoided loss is worth its annual cost.**

That reframing changes the conversation entirely, because it can be answered with numbers rather than with instinct.

PART II **Measuring What It Buys**

Measure the outage, not the fear of the outage

When I examined our incident data, the pattern was clear. We were not failing over to backup connectivity every week. At many sites we used the secondary circuit a few times in an entire year.

So take a backup circuit at $1,500 per month — $18,000 a year. Suppose the location suffers three primary outages that year, totalling six hours. The organization has spent $18,000 to avoid six hours without connectivity.

$18,000	**6 hrs**	**$3,000**
annual cost of the standby circuit	of primary outage it covered that year	per avoided hour of downtime

Illustrative figures, drawn from the cost and incident patterns we observed across the estate.

Three thousand dollars an hour might be an outstanding investment at a manufacturing facility where an hour of lost connectivity halts production worth hundreds of thousands. It might be an absurd one at a small administrative office where people can work around a few hours of disruption.

The technology is identical in both cases. The business requirement is not.

Cost per avoided hour

The calculation itself is trivial, which is part of why it is useful — anyone in the room can follow it.

THE ONLY ARITHMETIC THIS DECISION REQUIRES

Annual backup cost → **÷ hours of downtime avoided** → **= cost per avoided hour**

Both inputs are already known: the circuit is on the invoice, and the outage hours are in the ticket history. Nothing needs to be modelled.

Now the business has something concrete in front of it. Is avoiding one hour of downtime here worth $3,000? At one site, absolutely. At another, plainly not. At a third the answer might be that they would happily pay $30,000, at which point the conversation turns to whether one backup circuit is even sufficient.

The point is that the decision has stopped resting on an assumption that redundancy is always good, and started resting on the economics of the specific location.

Paying for 8,760 hours to protect six

There are 8,760 hours in a year. A backup circuit is billed for all of them and may be needed for a handful.

That does not automatically make it wasteful. It does mean the organization should know what it is buying. If a site pays $18,000 a year for backup connectivity that prevents six hours of downtime, it is purchasing those hours at $3,000 each — and that may well be correct.

> **What it should never be is invisible — a decision nobody made, buried in an architecture diagram under the heading of best practice.**

PART III **Availability as a Business Requirement**

Availability is a business requirement

This became one of the principles that increasingly governed my infrastructure decisions: availability should be engineered to the business requirement, not maximized as an abstract technical objective.

IT organizations naturally want systems to stay up. Engineers are trained to eliminate single points of failure, and that instinct is genuinely valuable. Taken to its logical conclusion, it produces

two of everything — two circuits, two routers, two firewalls, two power supplies, two providers, two paths, and eventually two data centers.

Every layer of that reduces some category of risk. Every layer also carries a cost. Which means the correct amount of redundancy cannot be determined by the network team alone, however good it is. It depends on the business impact of failure, and that information lives somewhere else in the company.

Not every site is a data center

A multinational enterprise has radically different availability requirements across its estate.

A manufacturing facility may need extremely high availability because losing the network stops production. A distribution center depends on connectivity for warehouse operations and shipping. A regional headquarters carries real business continuity obligations. A sales office can often operate temporarily by other means. A small administrative location may absorb several hours of downtime with almost no financial consequence.

Treating all of those identically is not standardization. It is a failure to notice that they are different.

> **The architecture should start with one question per site: how long can this location reasonably operate without WAN connectivity?**

Once that is answered, the appropriate technical solution usually becomes obvious — and it is frequently not the same solution as the site next door.

Redundancy versus recoverability

Working through this produced a distinction I found useful well beyond the WAN.

Traditional availability engineering concentrates on redundancy. But redundancy is only one route to acceptable availability. The other is recoverability. If a component rarely fails and can be restored inside the window the business can tolerate, paying continuously for a dedicated standby may deliver very little economic value.

So the question becomes: should we pay continuously to prevent the failure from causing downtime, or design an efficient way to recover when it happens?

For genuinely critical services, prevention through redundancy is usually right. For less critical ones, rapid recovery is often far more economical. The principle applies equally to circuits and to hardware.

The most expensive component was also the one that failed

One further piece of our data sharpened all of this. Approximately 98 percent of our circuit outages occurred in the local access portion of the network.

Which makes sense. The provider’s backbone was heavily engineered and heavily redundant. Local access is where physical infrastructure reaches into an individual building, and it is where things get dug up, cut, and knocked over.

THE REDUNDANCY CONTRADICTION

Local access was the most expensive part of the service and the part most likely to fail.

Traditional redundancy answers that by buying a second one at every protected location.

At global scale, duplicating the most expensive and least reliable component everywhere becomes extraordinarily costly — and it is worth noticing that a second local loop into the same building is not always as independent as the design assumes.

Remove the redundancy the business does not require

The answer was not to strip redundancy out everywhere. That merely replaces overengineering with underengineering, and the first serious outage proves it.

Instead we evaluated redundancy against site availability requirements. Locations that genuinely required highly available connectivity kept redundant access. Locations whose functions could absorb the downtime we had historically observed did not need an entire second WAN connection standing by.

That let us eliminate a great deal of rainy day bandwidth. Combined with the broader redesign of the WAN, it contributed to reducing our WAN costs by approximately 70 percent.

The savings were not the important part. We had stopped buying availability the business had never asked for.

The same question, asked about hardware

The same analysis changed how I thought about redundant equipment. During the same period, roughly 75 percent of our network outage tickets related to circuits or service providers rather than to failures of routers and firewalls. Yet standard designs routinely placed a second router and a second firewall at site after site.

So I asked the same question: what failure are we paying to prevent, how often does it actually occur, and what does the business require when it does?

If hardware failures were comparatively rare, two expensive devices at every location was not obviously the most economical way to buy availability. We developed in-country hardware sparing instead. Rather than duplicating routers and firewalls everywhere, spare equipment sat within a country or region, and each site’s availability requirement determined how quickly a replacement had to arrive.

A critical location might still justify on-site redundancy. Another needed four-hour replacement. Another was content with next business day. Availability became a requirement to be met rather than a standard to be applied.

Engineering to requirements instead of maximums

Engineers gravitate toward maximum availability. Finance gravitates toward minimum cost. Neither objective is correct on its own, and an organization that lets either one win produces a bad network in a different direction.

The right architecture sits where technical capability meets business requirement, which means knowing, for each site: the business functions performed there, the financial impact of losing connectivity, historical outage frequency and duration, available workarounds, required recovery time, hardware replacement capability, circuit restoration expectations, and the annual cost of preventing that downtime.

Only then can anyone say whether redundant connectivity makes economic sense there.

Knowing why

Rainy day bandwidth taught me something durable about infrastructure economics: redundancy has value only in relation to the business impact of the failure it prevents.

A second WAN connection may improve availability dramatically. But when local access is most of the WAN cost, that redundancy can nearly double the price of connecting a location — and across a multinational estate the numbers become enormous. Many of our backup connections were used a few times a year, while local access accounted for roughly 75 to 80 percent of MPLS WAN cost and roughly 98 percent of our circuit outages.

So instead of automatically buying two of everything, we matched availability architecture to actual site requirements. Some sites kept redundancy. Others did not. Hardware was evaluated against real failure rates and recovery requirements, which produced in-country sparing rather than universal duplication.

The result was not a less reliable network. It was a network whose cost and availability were deliberately aligned with the business.

The goal of infrastructure architecture is not to eliminate every possible minute of downtime regardless of cost. It is to deliver the availability the business actually needs at a rational price. Sometimes that means two circuits. Sometimes it means one.

Buy the availability the business requires.

Know why you bought it.

A note on the account. This is a first-person recollection of work carried out at Cadbury during approximately 2007–2009. Percentages and cost figures are approximate and drawn from internal analysis of the period; the per-site figures are illustrative of the patterns we observed rather than a single location’s invoice. Product and category names are used descriptively and belong to their respective owners.