WHITE PAPER · ENTERPRISE NETWORK ARCHITECTURE

Your SLA Is Measuring the Part That Isn’t Failing

What carrier service level agreements actually cover, and what to buy instead

Roughly 98 percent of our circuit outages happened in local access — the portion of the service our contracts protected least, and frequently the portion the SLA measurement span excluded entirely. Here is how that changed where we spent the money.

BY **JACK NASH** 500 SITES · 60 COUNTRIES ABOUT 10 MIN READ

THE ARGUMENT IN BRIEF

- SLA risk isn't free money from the carrier. It's a modelled cost, and the customer generally **funds part of the risk it believes it transferred**.
- The strongest commitments often cover the **provider's own network** — while roughly 98% of our circuit outages happened in local access.
- Performance is frequently measured **PE to PE, not CE to CE**. A carrier can report an excellent result across a span that excludes where your failures live.
- Even a valid credit **arrives after the outage**. The business needed connectivity during it.
- So we bought less contract and more resiliency: **an SLA is a contractual response to failure; resiliency is an engineering one.**

The theory is reasonable

Between approximately 2007 and 2009, while at Cadbury, I was responsible for a global corporate network of roughly 500 sites across 60 countries. Negotiating carrier agreements at that scale meant negotiating service level agreements at that scale, and it made me examine what we were actually purchasing.

Start by giving the SLA its due, because the underlying logic is sound.

A provider commits contractually to a particular result. Failure to achieve that result creates a financial consequence. That consequence is intended to influence behavior: fix my problem rather than pay the penalty. It aligns incentives, at least on paper.

SLAs also carry real organizational value beyond the technical. They establish measurable contractual expectations, which is genuinely useful when two large organizations need a shared definition of acceptable. They create financial remedies. And they provide internal reassurance — a demonstration that management has contractually protected the business, which matters in environments where that question gets asked.

I am not dismissing any of that. What I began questioning was narrower and more practical: when we paid more for a stronger SLA, did the operational result actually change?

Four things I found suggested it often did not.

Who really funds the risk

In the large multinational telecommunications agreements I was dealing with, SLA exposure was not simply free financial risk absorbed by the carrier out of confidence. It was another variable in the commercial model.

The provider understood its own historical performance. It understood the commitments being negotiated. It understood its likely financial exposure across the term. All of that could be modelled, priced, and incorporated into the economics of the overall agreement.

Which means the customer could end up funding at least part of the very risk it believed it had transferred to the provider — paying, in effect, an insurance premium set by the party who knows the claim history and you don't.

That doesn't make the arrangement improper. Carriers are entitled to price risk. But it does change how you should think about a stronger SLA. It is not free protection acquired through negotiating skill. It is a product with a cost, and the cost is somewhere in your rate.

The commercial level and the operational level

The second issue was organizational, and I saw it repeatedly during actual outages.

The operations personnel restoring our service were generally driven by their own internal metrics, escalation procedures, and restoration objectives. They were not, as a rule, working from detailed knowledge of the financial provisions in our master telecommunications agreement. In my experience the technician dispatched to a failed circuit does not know what our contract says, and it would be strange if he did.

The SLA existed at the commercial level. Restoration happened at the operational level. Those are different parts of a large organization, operating on different timescales, and the connection between

them is considerably weaker than the theory of incentive alignment assumes.

So the question became concrete: when a circuit fails at 2am, does a stronger contractual penalty cause anything different to happen on the ground that night? I could not persuade myself that it did.

Does it cover the failure domain?

The third issue was the one that mattered most, and it came directly from our own operational data.

Depending upon the carrier, country, service, and contract, the strongest SLA commitments could apply primarily to the service provider's own network rather than to every component of the local access path. Access arrangements, particularly where the provider was reselling or interconnecting with a local incumbent, were frequently covered differently or more weakly.

That mattered enormously, because our incident data showed where failures actually occurred. Approximately 75 percent of our network outage tickets involved circuits or service providers — and of those circuit outages, approximately 98 percent involved the local access circuit rather than the provider's core network.

~75%	~98%	~2%
of outage tickets involved circuits or providers	of those circuit outages were local access	involved the provider core — the best-covered part

Figures are approximate and drawn from internal Cadbury incident analysis of the period.

The provider's core was not our reliability problem. The edge was. And the edge was the part the contract protected least.

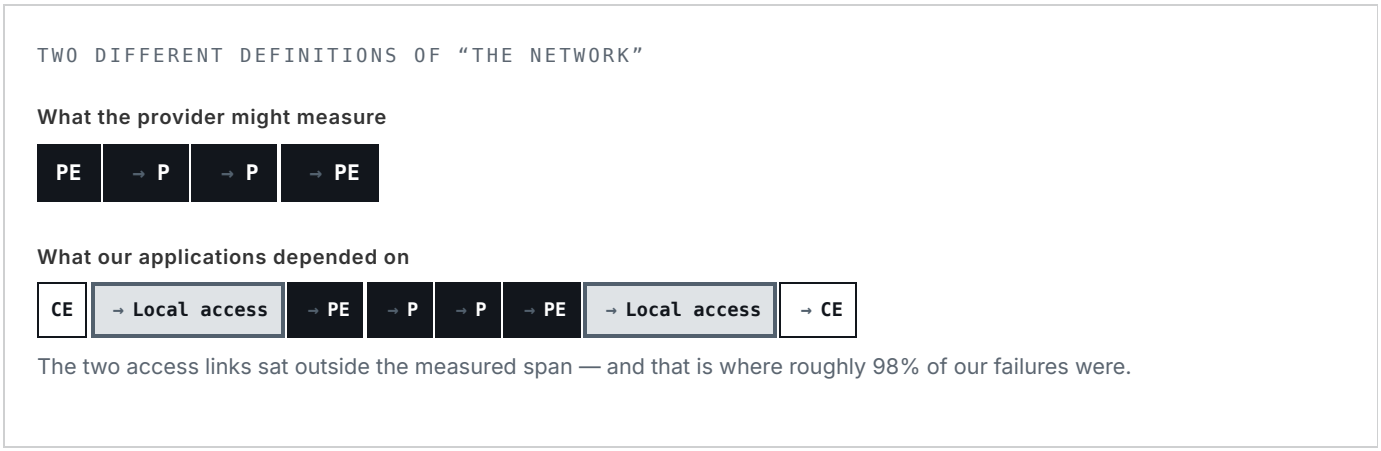
> **Does the SLA actually protect the portion of the service that is failing?**

A contractual guarantee covering a highly reliable provider backbone has limited operational value if the overwhelming majority of the customer's outages are occurring somewhere else entirely. You can hold a strong commitment on the 2 percent and a weak one on the 98 percent, and the contract will still look robust in a summary slide.

What the SLA is actually measuring

The fourth issue was the measurement boundary, and it is the one I would encourage anyone to check first, because it is easy to verify and frequently surprising.

Depending upon the provider and service, SLA performance could be measured between provider-controlled points in the network rather than across the customer's complete end-to-end path.



Those are not equivalent measurements. They are not close to equivalent.

A carrier could legitimately report excellent latency, packet loss, and availability across its backbone while a customer simultaneously experienced poor end-to-end service because of an access problem. Neither party is lying. They are describing different spans. And the span the carrier is describing systematically excludes the two hops where our failures concentrated.

An excellent provider-core SLA result therefore did not necessarily describe the network experience our applications or our users were receiving. The SLA was measuring the provider's definition of its network. I needed to measure the business's definition of the network.

> **The question that mattered wasn’t “did the provider meet its SLA?” It was “can our applications reliably communicate from one endpoint to the other?”**

The credit arrives after the outage

One more point, and it is the simplest of all.

Even when an access failure did qualify for an SLA credit, the credit did not restore service. It arrived afterward, as a line item, usually weeks later and often months later after the claim process concluded.

The business did not need a credit. It needed connectivity during the failure. A plant that cannot reach its systems for six hours has lost six hours of production, and the remedy is a proportional reduction on a circuit invoice that is small relative to what the outage cost.

Once I framed it that way, the economics of buying stronger SLAs stopped looking like risk transfer and started looking like a modest rebate on the wrong number.

Engineer around failure instead

All of this changed how I thought about SLAs. Not into uselessness — I still negotiated them, and I still expected providers to be held to them. But I stopped treating a stronger SLA as a substitute for resilient architecture, and I stopped spending incremental money there.

Rather than spend more on contractual consequences after a provider failed, I preferred to invest in making the provider's failure less important.

If a circuit failed, I didn't primarily want a service credit. I wanted the network to stop using the failed circuit. If a path became unavailable, I wanted another path.

That is what we built. Sites had multiple Internet circuits where appropriate, so a failed circuit meant failover rather than an outage. We engineered failover between our global points of presence as well, so a site's preferred path into the global network could use one PoP while an alternate path provided access through another. Circuit failover protected local access; PoP failover protected access to the core. Different failure domains, addressed separately.

> **An SLA is a contractual response to failure. Resiliency is an engineering response to failure.**
>
> **I preferred the engineering.**

Spend against the failures you actually have

The same reasoning ran in the other direction too, and it saved us real money.

Standard enterprise practice was to deploy high-availability routers and firewalls at every site. Two routers. Two firewalls. Redundant everything. It sounded safe, and it was easy to justify in a review.

But our incident data said the failures were elsewhere. A second firewall doesn't restore a failed circuit. A second router doesn't repair a carrier outage. We were spending capital protecting against a failure mode that accounted for a minority of our incidents, while the dominant failure mode sat outside the equipment rack entirely.

So we changed the approach. Instead of automatically buying redundant edge hardware everywhere, we examined each site's actual availability and recovery requirements. How long could this location operate without this device? How quickly did we genuinely need to restore it? Could strategically positioned spares satisfy that requirement instead of duplicating production equipment at every site?

Frequently they could, so we built in-country sparing strategies covering multiple locations according to geography and required recovery time. Where a business requirement genuinely justified immediate hardware failover, we still deployed local HA. We simply stopped treating two of everything as the automatic definition of availability.

> **Availability is the requirement. Redundancy is one possible solution.**

The connecting idea is the same one behind the SLA analysis. Look at where your failures actually occur, then spend there — whether the spending is contractual or architectural.

What to ask for instead

If you are negotiating or renewing carrier agreements, these are the questions I would put on the table before discussing the size of the credits.

Where exactly is performance measured? Ask for the measurement endpoints in writing. PE to PE, CE to CE, or something else. This single question reframes most SLA conversations, and the answer is often not what the summary suggests.

Is local access covered on the same terms as the core? Ask specifically, per country. Where the provider is reselling or interconnecting with a local incumbent, the answer frequently differs from the headline commitment.

What does your own incident data say about where you fail? Pull two years of tickets and classify them by failure domain. You cannot evaluate whether a contract covers your failures until you know what your failures are. Ours took an afternoon and changed the architecture.

What did last year's credits actually total? Compare that against the premium paid for the enhanced commitments. Then compare both against what the outages cost the business. The three numbers are usually not in the relationship people assume.

What is your mean time to restore, in practice, by country? This is the number that describes your operational reality. A contract does not change physics, local labor practices, or how long it takes to get a technician to a site in a difficult market.

What would the same money buy in resiliency? Price a second diverse circuit at your most outage-prone sites against the SLA premium across the estate. In our case that comparison was not close.

You may still conclude the SLA is worth buying. In regulated environments, or where a contractual commitment is required for reasons that have nothing to do with engineering, it plainly is. The point is to make it a measured decision rather than an inherited one.

What this is and isn’t an argument for

An SLA is not inherently useless. It establishes shared expectations, creates remedies, and in some organizations it is a governance requirement rather than an engineering one. Those are legitimate reasons to have one.

This is an argument against a specific assumption: that purchasing a stronger SLA is a meaningful way to improve availability. In my experience it frequently is not, because the commitment is priced back into your rate, the incentive rarely reaches the technician, the coverage often excludes your dominant failure domain, the measurement span often excludes it too, and the remedy arrives after the business already absorbed the loss.

Ask what the SLA measures, ask whether it covers where you actually fail, and then decide how much of your money should go to the contract and how much should go to the architecture.

Don’t contract around failure when you can engineer around it.

Measure the network your business depends on, not the network your carrier is describing.

A note on the account. This is a first-person recollection of work carried out between approximately 2007 and 2009. Percentages are approximate and drawn from internal analysis of the period. SLA structures vary by carrier, country, service, and contract; the patterns described are those I encountered and are not claims about any particular provider's current terms.