

## The Problem with Service Level Agreements

*A good Service Level Agreement (SLA) should provide a sound basis for two parties to work together. If it is done well, everyone knows what to expect and there are well defined incentives for putting things right when a problem occurs.*

*In practice, the SLA often causes more harm than it does good – it is not unknown for a supplier to ignore faults on a critical service in favour of less vital tasks, simply to minimise penalties.*

*This paper explores some of the unwanted side effects of a poorly crafted SLA and explains how to design an effective one*

Presented by Intercai Consultants

Steve Hodson and Mark Norris

## 1. Introduction

Most transactions that involve money are covered by some form of guarantee. The user of a credit card is insured against loss, the users of a telephone service are covered by the provider's code of practice and most high street goods are sold on the basis that they are 'fit for purpose'. There is nothing new in this – the suppliers of goods and services have, since time immemorial, done their best to gain competitive advantage by making sure that their customers have no concerns buying from them. Customers, at the same time, tend to want assurances when they don't trust the supplier to deliver some aspect of their product, especially when financial disadvantage is involved.

So it is in the modern telecommunications and information technology business. The customer/supplier relationship between large organisations is alive and well and is growing ever more competitive. But who is to define what is 'fit for purpose'? Those offering to manage networks and systems routinely offer guarantees of performance to gain the confidence of their customers, but how much reliance can purchasers place on these promises? These guarantees take the form of Service Level Agreements (SLAs) and are the agreements upon which virtually all outsourcing deals are founded. Unfortunately, experience indicates that all too often the SLA is not soundly based and often has the opposite effect from that intended. It is not always the 'evil' supplier that is at fault either, ignorance (or misplaced enthusiasm) on the part of the customer is often a major contributor to a failed SLA.

This paper examines a few myths and disasters in the SLA world with the intention of illustrating how an SLA can add value to contractual relationships rather than being the catalyst for their breakdown.

## 2. The real purpose of a Service Level Agreement

*"Far better an approximate answer to the right question, than the exact answer to the wrong question" – John Tukey*

It is a common misconception that an SLA sets expectations about the service level that should be delivered. If it does, there are problems ahead as we shall see later on.

The level of service that a customer can expect from a supplier should be defined in the specification of the service. Alongside the specification, there should be an agreement on how the service should be used – sometimes called an Operating Level Agreement (OLA). This defines the conditions around the service that will enable it to be delivered to its specifications and could include responsibilities on the customer. Once the specification and OLA exist, the SLA can do the job it is intended to,

<i>Service Level Agreement</i>	
<i>Specific service targets – the measure by which satisfactory delivery is assessed</i>	
<i>Service Specification</i>	<i>Operating Level Agreement</i>
<i>Definition of the service</i>	<i>Normal service parameters and supplier operating procedures</i>

# White Papers

## From IML

namely managing the delivery of the service to its specification. The above diagram may seem to be little more than a bureaucratic division, but it is when the separate elements are confused that problems frequently occur. One of the most common mistakes we come across, that has its roots in this confusion, is interchanging the service level and the SLA level. They are not, and can never be, the same thing, as we explain below.

Frequently we see the absence of a service specification forcing the SLA to try to act as a substitute, when its real purpose is to provide the safety net in the customer/supplier relationship.

From the customer's point of view, the SLA should say where the limit of acceptable service lies and provide sufficient motivation for suppliers to perform within this limit. From the supplier's point of view, the SLA should define clear targets for acceptable service that can be met with an acceptable risk and which provide a continuous incentive to improve.

Is this always the case? Let's look at some real ones.

- We guarantee 100% Availability.

This is clearly unachievable. Nothing is perfect and some failure, even if rare, is inevitable. So, if the service can't be operational all of the time the supplier is bound to be liable when it goes wrong. In this case, the supplier's promise is patently impossible. And this brings into question either their competence or their judgement!

- We guarantee 100% Availability except for the first hour of downtime

By contrast with the first case, this is quite easily achievable. If a service is unavailable for, say, 40 minutes, several times a day, the customer is unlikely to be very happy. In this case, the supplier is making an apparently good offer that has a critical (and intentional) get-out clause.

- Latency across our core network is 65mS

This is another example of a partial promise. If the service works well over part of the network but is unspecified over the rest, then the end-to-end performance is entirely dependent on the unspecified element. This may be a good or a bad thing. Another empty promise from a supplier that doesn't necessarily translate into a good customer experience.

- We offer a 'gold' service that guarantees your Quality of Service

It is not clear exactly what this actually means. If the service is offered over a shared network (as most IP packet services are), it probably means that the 'gold' service is given some priority. The question then is how many other services have the same priority (and so contend for the same resources)?

They don't really meet their purpose, do they? There is no one reason why, but the prevalent theme is empty promises from suppliers and no solid basis for the claim.

## 3. Unexpected consequences

*“If you open that Pandora’s Box, you never know what Trojan Horses will jump out” – Ernest Bevan*

It is bad enough being made empty promises but there is a more insidious side to an SLA. They can do more than disappoint – a really bad one can cause significant damage to both customer and supplier, something that is emphasised by some of the high profile failures currently in the press.

Curiously enough, the most dangerous SLA is usually the one that was produced with good intent. In the previous section we looked at a catalogue of offers that were by suppliers in the knowledge that they would not be penalised by them. In this section we consider some all-too-real instances of earnestly forged SLAs that ended up doing exactly what wasn’t intended.

### 3.1 The wall of death

*How you motivate your supplier to focus on the unimportant at the expense of the vital.*

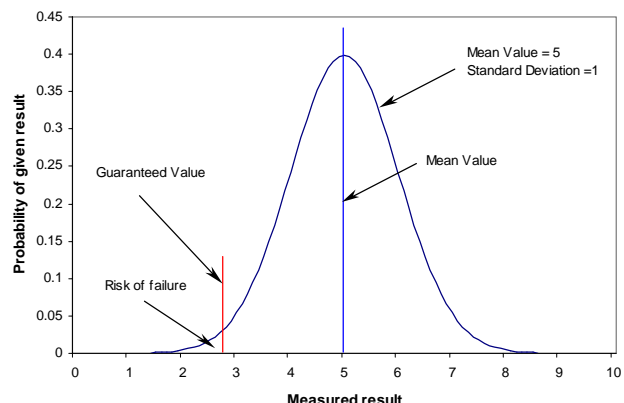
Consider for a moment a situation where a supplier is delivering several services to a client. One of these services is a core, critical, part of the client’s business and has a heavy penalty should the service fall below a defined threshold. The penalty is onerous and applies in full as soon as the threshold is crossed. The other, less important, services have lesser degrees of penalty.

The client obviously wants the absolute best efforts from the supplier in relation to the critical service, but has the SLA achieved this? If we take it as read that both client and supplier are rational and that the behaviour expected of the supplier is to minimise his penalties, then a moment’s reflection will show that the SLA is having a rather different effect. If the supplier has a bad month and the critical service fails half way through, the penalty applies in full. There is nothing the supplier can do to mitigate this so his only rational behaviour is to drop any efforts to repair the critical service and concentrate on those that can still affect his penalties. For obvious reasons we have come to call this the ‘Wall of Death’ form of SLA penalty.

### 3.2 The meaning of mean

*How a supplier can fail the SLA half the time by following the TMF handbook.*

This is where things start to get a little technical, but that’s really where the SLA should be if it is to be soundly based. Without wishing to be unduly critical of the TMF, they do have a fundamental misconception in their definition of an SLA when they state that the SLA defines the service level expected from the supplier. If this statement is true, then consider the picture that shows the results of measuring a service’s performance over a period of time and



# White Papers

## From IML

how the actual measure varies about a mean. Now if the SLA is set to the average level (i.e. the service that is 'expected') then it is pretty obvious that the service will fail its SLA on half the occasions when it is measured. That doesn't sound much of a basis for good relations between the supplier and the user. The one thing the SLA threshold cannot be is the average performance level.

In a well constructed SLA, there is a clear notion of risk. The above figure shows the likelihood of an event (such as a service failure) occurring – the area to the left of any vertical line representing the risk of failure. As already stated, if the supplier sets a line in the middle of the curve (the mean value of the event occurring), they risk failing the SLA 50% of the time.

The art of a good SLA is in setting the guarantee level for a service far enough to the left of the mean so that it represents a performance that is acceptable to the customer at a risk of failure that is bearable by the supplier.

### **3.3 The importance of time**

*How you ensure that statistics lie – and end up with a supplier who cannot win, so who gives up.*

Now things get really technical as we enter the world of statistics. We won't go into this in great depth, (but you can ask us to help you if you have a serious problem in this area), but probably one of the commonest errors that we see in SLA definitions is allowing insufficient time for any measure to reach statistical significance. What does this mean? Well, it can mean a supplier doing the exact opposite of what the SLA intends – again! This effect sometimes happens because the actual measurement time is too short, or more often, it is because the client wishes to measure services to too fine a level of granularity, such as each instance of a service.

If a supplier is delivering a service that fails fairly infrequently, say once a year for the sake of argument, then on average he may have a few hours, say 4, in which to fix a fault in order to achieve the long term service target (about 99.95% Availability in this case).

Now, if a client wishes to measure a single service on a monthly basis, then the 4 hours is evenly distributed across the months and becomes 20 minutes (it could be worse if the client only uses the working day). For most months when the service doesn't fail there isn't a problem, but in the month when the fault happens, the service provider has no chance whatever of meeting the 20 minute target so is unable to affect the size of the penalty. As a rational supplier what is his approach? The only thing he can do is to minimise the effort he puts into repairs because no other strategy will reduce his outgoings if the penalty is a Wall of Death type. Again the SLA has got in the way of good management.

### **3.4 And finally.....**

*Why performance SLAs never work*

We look at a structured approach to creating an SLA in the next section, but just take a moment to consider this common problem. Generally suppliers provide some form of telecommunications or IT service that supports some transaction or other. Taking money out of

# White Papers

## From IML

an ATM; entering data in a database; pulling down a page from the Internet, that kind of thing. The end user sees performance in terms of the time it takes that transaction to complete and that is perfectly reasonable. It is a small step from that point to deciding to place an SLA on the transaction delay that penalises the service provider if it is exceeded. Ignoring the very obvious issue of how to measure such delays (they often appear as other parameters such as jitter, lost packets, lost cells etc, but fundamentally the same argument applies to all of them) there is a fundamental principle at work here that undermines the whole basis for the SLA.

If we assume that the service is being correctly maintained (perhaps under the control of an Availability SLA) and that the service has been proved to be adequate when it was delivered (not necessarily always the case), then consider what will cause the delay to increase above the threshold. The only thing is traffic or load. The supplier has no control over this and can only react by adding capacity and that costs money. There is a business decision needed here and no matter how you dress the negotiations up, the supplier and user will, in the end, agree when the most appropriate time to add capacity occurs, and the SLA has absolutely no part in influencing this point whatsoever. SLAs that apply to delay in any of its many incarnations, never add value.

## 4. How to design a good Service Level Agreement

*“Opportunity is missed by most people because it is dressed in overalls and looks like work” – Thomas Edison*

There is no magic formula for a good Service Level Agreement but there are some useful guidelines. One useful mantra is that a good SLA needs an overall architecture at least as much as a network – it has to be carefully designed and should be based on sound engineering practice. Another is to keep separate the service specification and the SLA. Without these foundations the risk of creating a bad SLA increases markedly.

### 4.1 Structure

Perhaps the first thing to get clear is the organisation of the services that are being offered. Different types of service call for different measures, parameters and reporting. Once there is a clear discrimination between these different types of service, the most appropriate entries can be put into the SLA.

This figure shows a tried and tested structure for telecom services. It shows the typical services that a telecommunications vendor might offer. The IT vendors need not feel left out; they are part of the Applications row. Each row represents levels of technology at which an SLA could (but doesn't have to) be applied. Of course, SLAs are used for much more than just

	Implementation	Operation	Performance	Other
Application	Implementation SLAs	Operational SLAs	Performance SLAs	
Network				
Link				
Bearer				
Bandwidth				
PoP				

# White Papers

## From IML

telecommunications and IT but the same principles apply everywhere. The language of mathematics that provides the engineering basis for an SLA is universal.

The columns in the diagram separate out the different types of SLA, mainly because their mechanisms are different and the units in which they are measured are different and we all remember the lesson about mixed units when we were at school – don't we?

Implementation is to do with, well – Implementation. This is everything to do with delivery on time, whether that is a physical piece of equipment, a service or just the reaction to a telephone call.

Operational is to do with things that break and get fixed (some time later). Here we are interested in how often things go wrong (technically known as Mean Time Between Failure, MTBF) and how quickly they get repaired (the TLA here is MTTR – Mean Time To Repair, except that is an FLA). Combined, these two contribute to Availability. That is the British Standard definition and don't let anyone else tell you otherwise (that is another of our myths that we don't have room for here).

Performance relates to the way in which the system performs when it is not in a faulty state. This usually involves capacity and load and we have said above that SLAs here are seldom appropriate.

## **4.2 Measures**

Once the structure is in place, it is time to get quantitative! In the above diagram, there are a host of tempting looking places where a useful measure might be taken. For instance, an implementation measure of bandwidth may be relevant. Or perhaps a performance measure on the bearer is more telling? There is certainly an amount of judgement and experience that has to be applied at this stage but there are some overall guidelines on any measure selected.

- Is it simple?
- Is the measure the whole parameter you need?
- Can the measured parameter be controlled?
- Is the measured parameter a primary measure or a derivative?
- Is the measurement of the parameter a part of a control loop?
- Does the measurement period provide statistical significance?
- Is the measure a guarantee or a mean value?

If you meet all of these criteria but still have a couple of hundred measures to track then you have missed something somewhere. Burying the operators with statistics at the end of the month is a sure way to lose any meaning the measures may once have had. You should be asking serious questions when the measure count gets up to 30 or so.



## 4.3 Living with an SLA

Like most documentation, put it in a drawer and it will simply curl up and wither. An SLA without something to keep it active becomes shelf-ware; it has to be part of the management process with reviews built-in as part of business-as-usual.

<i>Monthly review</i>	<i>SLA review</i>	<i>Trends</i>
<ul style="list-style-type: none"> <li>- summary of performance against SLA targets over the current period</li> <li>- facts behind the figures</li> <li>- agreement of credits (if any) and actions required.</li> </ul>	<ul style="list-style-type: none"> <li>- are the targets still in line with customer need and supplier capability</li> <li>- if there is a need or desire to change, what steps have to be taken to make it happen</li> </ul>	<ul style="list-style-type: none"> <li>- alarm bells and immediate restorative action planning</li> <li>- input to long term planning (e.g. ordering of more network capacity, increase in support staff)</li> </ul>

A simple set of formal reviews that ask the questions shown in the above diagram is enough to keep the SLA live (and, therefore, useful). In most organisations, this is a straightforward adjunct to an established quality or project management structure.

In fact some of the most successful SLAs we have seen have no penalties whatsoever (other than perhaps the chairman coming along for a quiet chat) but clear aims and a desire to work together between the supplier and the user to the best business solution – whatever that may be.

## 5. Conclusion

A Service Level Agreement may be a marketing device but it should be produced by engineers. It is all too easy for a supplier to make promises that cannot be kept and end up damaging their business. It is even easier for a supplier to make promises that disappoint a customer.

Good management comes from experience and experience comes from bad management. In this paper, we have shown how our experience of some classic pitfalls can light up a path to the production of a good SLA.

## References

1. Previous IML papers on SLAs and related subjects which can be found at [www.intercai.co.uk](http://www.intercai.co.uk)
2. There is a lot of useful information on Service Level Agreements that, in spite of our comments above, can be found in the TeleManagement Forum SLA Handbook which can be found at <http://www.tmfcentral.com>. Do treat it with care though.
3. A good commentary on the evolving state of the telecoms market can be found in *The Great Telecom Meltdown*, Fred Goldstein, Artech House, 2004

*The material contained in this White Paper is based on Intercal's wide experience of SLA design. An extensive and flexible training course that expands on the ideas and techniques mentioned here is available on request.*