OPN Resilience
Introduction and executive summary

For the last three years, the GridPP PMB has been monitoring the need for a backup optical link between RAL and CERN to provide resilience for the OPN network that delivers the raw LHC data to the UK. At this point we feel that the balance of risk-benefit-cost ratio justifies the expenditure: The long LHC data run planned reduces the chance of recovering, or catching-up, from a significant outage; the volumes of data to be transferred (particularly the T1-T1 traffic that also uses the OPN) have been demonstrated by STEP09 to be larger than previous anticipated; the incidence of outages has proved to be larger than expected from earlier experience; the threat to the UK reputation, being the only major Tier-1 without a backup link, is now increased; and the provisioning cost has fallen by at least a factor of two. GridPP proposes to spend £52k of our existing hardware budget to install the link, supported by a recurrent cost of between £40k and £60k per annum, depending on negotiations about the end-point costs. These sums, though large, represent a few percent of the GridPP hardware budget and can be accommodated within the existing award as the delay of the LHC-start has pushed back the experiment resource requirements.

The LHC Network (LHCOPN) – risks and resilience planning

The LHC Grid (LCG) consists of a Tier-0 centres at CERN connected to Tier-1 centres spread throughout the world. Each of the Tier-1 centres has many associated Tier-2 centres. The Tier-0 to Tier-1 connections are via the LHC Optical Private Network (OPN) which utilises “lightpaths” (more correctly dedicated bandwidth circuits) in a star configuration from CERN to each Tier-1. This is depicted in the Figure-1 below. In addition to the primary connections (mostly at 10 Gbits/s) many sites have backup connections to mitigate the risk of the primary connection being lost. The 10 Gbit/s primary connections are unprotected and as such the costs are very much lower than a Telco grade connection. In the UK the cost is currently covered by a agreement between JISC and the Research Council.

In the UK we currently have a single connection from RAL to CERN and are the only Tier-1 without a backup link. The GridPP PMB has been monitoring this situation on at least an annual basis since 2007 with a view to identifying the point at which a back-up link was justified by the cost, benefits and level of risk. We have very good relations with JANET and have access to the information needed to assess outage risk through Robin Tasker. The PMB has weighed this assessed risk against the consequences of such an outage over different assumed periods of time up to a week. Appendix A shows the first of these assessment considered by the PMB.

In 2007 the LHCOPN was used primarily for Monte Carlo data transport and for “data challenges” designed to prove the readiness of the entire LCG infrastructure. The PMB weighed the risk of failure against cost and consequence at the time and decided that there was no compelling case for allocating funds to mitigate possible link failure.

In 2008 the PMB considered the case again in the light of the imminent start up of the LHC. At this time more information was available on risk of failure and experience of actual failures on JANET. After a careful analysis (Appendix B) the PMB again concluded that whilst the consequence of an outage was greater, that the realistic expectation of data volumes was low enough that outages of even a week could be dealt with by both “catch up later” and use of the production network ~ 1 Gbit/s rates. The PMB therefore again decided not to put in place a resilient link at that time.
In 2009 the PMB considered the case again. The analysis is presented in Appendix C. Due to both the more significant consequence of an outage on the competitiveness of UK physics analysis groups, and due to the latest planning of the experiments in respect of network volumes the PMB decided that it was now timely to put the resilient link in place. The PMB has always seen this as an expected part of LCG planning which it was able to delay until now. The PMB has also taken advantage of the fact that the provisioning costs have dropped due to the re-engineering of the JANET lightpath service.

The LHCOPN resilience planning and experience.

The LHCOPN encourages all countries to install resilient back up links. The figure below shows the current provisioning. Every participating country, except the UK, now has a resilient back up link.

![LHCOPN – current status](image)

Figure-1: The LHC OPN Network Topology

As reported in Appendix B there was a significant outage in 2008. During tests the main links provider suffered a fibre cut between Frankfurt and Geneva, resulting in connection problems to the following Tier-1 sites for 36 hours around the 11th June: Taiwan, Canada, Italy, Germany, and Nordugrid. All were cut off but used their backup links to maintain connectivity. BNL and FNAL were running with reduced capability. This event served to emphasis that (i) breaks do happen and (ii) the can take out more than one Tier-1 connection.
During recent STEP’09 in 2009 exercise there was another major fibre outage which did not affecting RAL but affected several other T1 sites. The LHCOPN resilient infrastructure kicked in and the traffic continued to flow. This figure below taken form a recent LHCOPN meeting illustrates this. If the fibre outage had affected the RAL/CERN circuit then RAL would have been dead for the duration.

Forward Planning

GridPP has also carried out a network forward planning exercise. This is included as Appendix D. This has considered the actual and planned network traffic between all types of site. The conclusions were that whilst at present rates the single 10 Gbit/s circuit is expected to be adequate until the end of 2010, traffic is rising and this may not remain adequate much longer. The LHCOPN rate will be kept under close scrutiny and if the rest of LCG moves to greater capacity in 2010+ then GridPP should be prepared to follow quickly.

Good forward planning based upon normal operating conditions is fundamentally different to disaster planning and mitigation. Nevertheless the forward look indicates further capacity is likely on the scale of 12-18 months. The resilient link that GridPP intends to install will allow for development of diverse routes and load balancing as a step towards a second circuit.

Resilient Solution

The proposed solution will be diversely routed everywhere except for the connection to Geant in London. This is unavoidable but presents little practical risk as this is housed in a Telco standard exchange and the router in question is an essential part of the JANET to Geant peering.
Summary

The GridPP PMB has been monitoring the status of the LHCOPN since 2007. The risks and consequences of outages have been considered along with expected traffic rates. The PMB decided that a resilient link should be installed to ensure that the UK would not become disadvantaged in the event of a failure. The PMB considered that this could be delayed in 2007 and 2008 and that it should now be installed for late in 2009.
APPENDIX A: 2007 Assessment

RAL-LCG2 OPN Link Resilience

Robin Tasker (r.tasker@dl.ac.uk). 20 August 2007

1. Introduction

This document discusses the issues associated with the provision of resilience for the CERN-RAL-LHCOPN circuit used for Tier0/Tier1 LHC traffic and for peering between RAL and other Tier 1 Centres. It is to be noted that this lightpath connection is operated over an unprotected circuit across both JANET and Geant.

2. Reported Outages on CERN-RAL-LHCOPN Circuit

Trouble Tickets are generated by the end-2-end Network Operational Centre (NOC) of Geant and reported via JANET(UK) to the Networks Group at RAL. The tickets are there routed onwards to the RAL Tier 1 Centre for distribution to the GridPP community as appropriate.

The following table shows the reported outages for the CERN-RAL-LHCOPN circuit since it was brought into service towards the end of 2006.

<table>
<thead>
<tr>
<th>Ticket #</th>
<th>Problem Start</th>
<th>Problem Finish</th>
<th>Problem Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>998059</td>
<td>07/03/2007</td>
<td>03/07/2007</td>
<td>Alcatel maintenance on Paris to London circuit</td>
</tr>
<tr>
<td></td>
<td>08:34:54 CET</td>
<td>13:56:47 CET</td>
<td></td>
</tr>
<tr>
<td>1003054</td>
<td>20/03/2007</td>
<td>20/03/2007</td>
<td>The replacement of a module in one of the Cern LHCOPN routers</td>
</tr>
<tr>
<td></td>
<td>08:00:00 CET</td>
<td>08:30:00 CET</td>
<td></td>
</tr>
<tr>
<td></td>
<td>10:52:11 CET</td>
<td>10:59:48 CET</td>
<td></td>
</tr>
<tr>
<td>1021499</td>
<td>24/04/2007</td>
<td>24/04/2007</td>
<td>Alcatel reported a problem on the Supervision channel on the circuit.</td>
</tr>
<tr>
<td></td>
<td>01:47:41 CET</td>
<td>01:50:42 CET</td>
<td></td>
</tr>
<tr>
<td>1049958</td>
<td>20/06/2007</td>
<td>22/06/2007</td>
<td>A fibre break in France due to road engineering. Colt Telecom reported that a third party (Cegetel Telecom) had been contacted to arrange</td>
</tr>
<tr>
<td></td>
<td>10:17:54 CET</td>
<td>18:34:33 CET</td>
<td></td>
</tr>
</tbody>
</table>
It would appear that in the first 6-months of operation this circuit has been in general reliable with two scheduled maintenance outages, two short outages lasting in total around 10 minutes. However one major failure as a consequence of a fibre break on a circuit in France has occurred which resulted in an outage of ~56.5 hours whilst ~2km of fibre was repaired/replaced.

3. Risk Assessment of Fibre Breaks

Fibre outages will inevitably happen and while the lightpath service is delivered across unprotected circuits this will lead to a loss of connectivity between the end points.

It is difficult to base a generalisation on the first six months of operational experience specifically as one would reasonably expect a fibre break as experienced over the period 20-22 June to be the exception and certainly not an event expected with a 6-month frequency.

Based upon the JANET(UK) experience of operating a large scale network (SJ5) over the period January 2007 to June 2007 the risk to the fibre infrastructure is assessed as follows,

<table>
<thead>
<tr>
<th>Actual Fibre Break</th>
<th>one break per 182.5km per year</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean Time to Repair (MTR)</td>
<td>19.8 hours</td>
</tr>
</tbody>
</table>

On the basis that the fibre route RAL – CERN is ~1400km, for the CERN-RAL LHCOPN,

| Expected actual fibre breaks per year | 7.7 (with 19.8 hours MTR) |
| Total Outage | 152.6 hours (6.4 days) |

Note: Please treat the data from JANET(UK) shown above in confidence to GridPP PMB

4. CERN-RAL LHCOPN

4.1 Primary CERN-RAL LHCOPN Circuit

To provide resilience an entirely diverse and physically separate fibre route is required between RAL and CERN across both the JANET and Geant networks. In addition separation is required at the route end-points and at peering points on the route. It is assumed that end-site and intermediate termination equipment is covered by maintenance contract (or similar) limiting the scope of resilience discussed here to the physical fibre infrastructure.

The existing advertised tariffs for the cost of a 10G lightpath circuit are as follows,

**JANET(UK) Costs**

Installation: £30,451

Monthly recurrent charge: £2,204

i.e. about £57k for the first year, and about £27k for subsequent years.
GEANT Costs

40K Euro per year, i.e. 20k Euro per end point

Together these costs equate to **£84.2k for the first year** and **£54.2k for subsequent years**.

It should be noted that at present the cost as detailed above for the existing circuit is carried by the STFC but if this should change then GridPP would be liable for that amount in addition to any cost associated with a second resilient route or for any additional capacity on this route

**Note:** It is my understanding that the Research Council agreement with the JISC, for which STFC pays its share, includes the cost of the primary connection. I have started discussion with JANET(UK) to confirm this statement

However these costs are based upon the most direct route for such a circuit and do not take account of an additional equipment/circuit costs needed to provide the required resilience. These additional costs would be calculated on the basis of a route survey.

4.2 Proposed Resilient CERN-RAL LHCOPN Circuit

The following solutions have been identified by JANET(UK) together with the indicative costs to provide the required resilience,

**RESILIENT SOLUTION 1**

Backup: RAL over TVN to RNEP-2, Warrington, Dublin, GENEVA

Indicative cost of backup only:

Year 1: £163,000 Inc VAT

Subsequent years: £90,000 Inc VAT

**Note:** JANET(UK) have yet discuss this solution with HEANET or DANTE

**RESILIENT SOLUTION 2**

Backup: RAL over TVN to RNEP-2, London UK5, London Telecity, GENEVA (over GEANT using a diverse route to the primary path).

Indicative cost of backup only:

Year 1: £195,000 Inc VAT

Subsequent years: £122,000 Inc VAT

**Note:** There is one common point of failure in Telecity with this solution

5. Conclusion: GridPP are invited to set the probability of an outage to the CERN-RAL-LHCOPN circuit and the consequential service loss (of data transfer from CERN, and between Tier 1 centres) shown in Section 3 against the indicative cost of mitigation of such a risk (Section 4) and to determine the cost that
Network Disaster Planning

Document identifier: GridPP-PMB-xxx-NETWORK DISASTER_PLANNING
Date: 28/02/2008
Version: 3
Document status: Final
Author: P.Clarke, R.Jones, D.Newbold, R.Tasker
**Introduction**

RAL is connected to CERN by a 10 Gbit/s OPN link which is crucial for LHC operations.

The question of the resilience of this link has been raised. There is pressure from the WLCG-OPN Group for the UK to provide a redundant link for resilience, in common with some other WLCG-OPN partner countries (e.g. using a T1-toT1 fallback).

This issue has been studied and debated by the GridPP-PMB. This document sets out the results of that discussion and the policy adopted.

The status of this document is “for information and discussion with interested parties”

**Discussion**

1. Robin Tasker has received indicative costs from JANET(UK) for the provision of a diversely routed redundant 10 Gbit/s link (Appendix 2). The cost supplied was approximately £180k for the first year and £100k per annum subsequently with the possibility of additional cost based on the outcome of a route survey.

JANET(UK) have recently revised their tariffs to allow for (a) a 50% rebate on the DANTE costs for the GEANT services, and (b) on the assumption that an NREN/project in another country would pay for its end of the service.

This latter point (b) above is unlikely to be correct in the case of the RAL-CERN lightpaths, so in practice the overall figures from Appendix 2 will be reduced by the sterling equivalent of €30,000, which is half of what DANTE now charges an NREN for a 10G lightpath (full cost is €60,000 and includes both ends). At current exchange rates this will be in the region of £22,000. Such costs are too great for GridPP simply to decide to divert funds, or make a special case for extra funds to the STFC without carefully considering the risk/benefit and timing.

2. There is an alternative option to provide resilience for the Tier 0 – Tier 1 traffic only using multiple 1Gbits/s circuits via the JANET Lightpath / Geant+ services. This would be provisioned between RAL and CERN routed through Amsterdam (and therefore diversely routed to CERN).

The UK component of this costs comprises an element for the Regional Network and the JANET Core. However the published JANET(UK) tariffs do not include the Regional Network component as this necessarily is dealt with on a case-by-case basis.

For budgetary purposes the cost for each 1Gbits/s path would be charged at £8,800 for circuit setup and £15,200 for annual running within the UK; an additional £3,500 per annum for the GEANT cost; and in all likelihood add some additional cost for any far-end infrastructure. This is summarized in the Table below.

<table>
<thead>
<tr>
<th></th>
<th>TVN</th>
<th>JANET Core</th>
<th>GEANT</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Setup</td>
<td>£6,800</td>
<td>£2,000</td>
<td></td>
<td>£8,800</td>
</tr>
<tr>
<td>Annual Operation</td>
<td>£14,600</td>
<td>£600</td>
<td></td>
<td>£15,200</td>
</tr>
<tr>
<td>Annual Operation</td>
<td></td>
<td></td>
<td>£3,500</td>
<td>£3,500</td>
</tr>
</tbody>
</table>
3. Robin Tasker has obtained reliability figures from JANET(UK) based on the first 6 months operation of SuperJANET5. He has been able to derive for the RAL / CERN OPN route an estimation of fibre outages due solely to fibre breaks. This suggests an outage rate of 8 breaks per year each taking an average of 20 hours to repair. RT will work with JANET(UK) to provide revised figures as more SuperJANET5 operational data becomes available.

4. The above expectation should be seen in the light of two factors:
   a. Today such an availability is higher than availability of the endpoint storage systems themselves.
   b. Such an availability is already 98% (8 days in a year) hence violating the Tier-1 99% availability aspiration.

5. It has been agreed that the JANET Lightpath Service will be added to the SLA between the JISC and JANET(UK) as a part of the JANET services. This will include a statement with regard to the re-establishment of service following an outage although it must be noted here that the Lightpath Service is operated over unprotected circuits and that the re-establishment of the JANET IP Service will take precedence.

6. Each of the experimental representatives have been asked to write down the risk and consequences to the experiment of (i) 6 x 1 day breaks (ii) 2 x 3-day breaks during 2007/8. The results are appended.

7. The PMB will monitor this situation, and note carefully the outages in 2007/2008. The matter will be reviewed in the light of experience in the second half of 2008.

8. Consideration must also be given for the disaster scenario where a catastrophic network outage is experienced on the RAL – CERN route. The OPN circuit is carried across JANET(UK) and DANTE infrastructure which is known to carry IP production traffic. It is reasonable to expect that both JANET(UK) and DANTE have well developed service continuity plans in place which would aim to restore the physical network integrity within days. Clearly their priority will lie with the production IP traffic against which they are formally judged. RT is working with JANET(UK) (and by extension with DANTE) to understand the planned priorities within their service continuity planning to restore the lightpath circuits but see Item 5 above.

   On a similar basis RT is working with RAL and CERN to understand their service continuity planning following disastrous site outages and their timetable for service recovery.

9. For completeness the PMB is reminded that this document is only considering resilience. It does not address ongoing capacity planning. It should be noted that there is the expectation that a single 10 Gbit/s link will not be adequate on the scale of a year. The PMB needs to budget for such provision based on the costing provided in this document.

Conclusion

1. Based upon consideration of these statements, the PMB does not feel that at this stage it would be warranted in diverting the funds from the GridPP3 project in 2008 for a permanent fully resilient 10 Gbit/s link. Nor does the PMB feel it to be appropriate at this stage to seek to open special negotiations with STFC. STFC should however be fully briefed on the issue.
2. At the same time the PMB should consider carefully whether (and when) to commit ~£20k per annum for a 1 Gbit/s backup.

APPENDIX 1: Consequence of likely outages upon experiments

In this appendix each of the experimental representatives have been asked to write down the risk and consequences to the experiment of

1. 6 x 1 day breaks
2. 2 x 3-day breaks

break during 2007/8. [Originally 6 days was also requested, and this is left in for information]

ATLAS

The ATLAS computing model has two primary data transfers from CERN to RAL: the RAW data, which is moved when it arrives in the input buffers for the CERN Tier 0; and the processed data, which is shipped to RAL ~2 days later, after processing. If the network path is not available when the RAW arrives at the Tier 0, it will have to be redirected to another Tier 1 to ensure data security. This implies both that the other Tier 1s have an over-capacity in network to receive the extra data flow, and that RAL has a correspondingly larger pipe to take a larger than usual share when recovering from the outage. This applies independent of the length of the outage; a reasonable planning assumption is that the input pipe to RAL and receiving systems should be able to handle double the usual capacity. (We note that this extra load will apply to all experiments, as we will all be off at the same time, and all taking data at the same time.)

The handling of the reconstructed data is more difficult. This must be sent to the same site as its parent RAW data, and so cannot be diverted to an alternate site if the RAW data has been transmitted before the outage. This data must be buffered at CERN until it is shipped. If the processing delay is 2 days and the buffer can hold ~5 days worth of data, this means that the maximum outage allowed is limited. Assuming the UK takes 10% of the output and is the only Tier 1 unavailable, a 30 day outage would be possible, but would be at great risk. As we cannot guarantee processing within 2 days, especially in 2007/2008, a 6 day outage is a significant risk.

The other data transfer between CERN and RAL that would be at risk concerns services (conditions data, possibly the TAG database, probably the synchronization of the LFC with CERN etc.) The impact on these services depends on the maximum update speed allowed during the recovery. This has yet to be evaluated. The other consequence would be the lack of availability of the required conditions and other information to the ATLAS UK cloud. This requires work both from the experiment and the networking experts to allow a fail-over for the UK cloud to another Tier 1 or to CERN.

Turning to Tier 1-Tier 1 traffic over the OPN, if this is stopped then both RAL and its partner sites must provide buffer space to store datasets produced in the Tier 1 reprocessing. The reprocessing will be a near continuous activity. The ability to cope with a 6-day outage implies an output buffer able to hold more than 6 days’ worth of files at each of our partner Tier 1’s and at RAL, and double the Tier 1-Tier 1 bandwidth. However, as all files produced are to be archived and to be available on disk at the producing site (data class t1d1), then such a buffer will exist. The implication is that the Tier 1-Tier 1 link must be able to provide the catch-up bandwidth.

If an alternate route is available at 1/5 of the normal bandwidth, the first priority for ATLAS would be the transmission of processed data matching RAW data already received from CERN, then the service and database communications. The RAW and AOD files associated would be sent to the partner Tier 1, and copied from the
partner Tier 1 when the normal network is available, assuming there is the required over-capacity in bandwidth.

The impact of outages grows sharply with the length of outage. A 1-day outage may fall into machine development periods and is not too dissimilar to the likely machine and processing delays. A three-day outage will require significant buffering and considerable human effort to redirect the normal activities, both during the outage and during the recovery. It would effectively mean a disruption of 1 week per outage. It would also put the Tier 2 cloud activity at risk; the provision of a secondary route that can handle the transmission of the RAW data for CERN to RAL and some significant fraction of the processed data over the same route, which may be possible over the proposed 2Gbps link, will significantly mitigate the effects of a three-day outage. A six-day outage would have a major impact on our services, and will require significant precautionary measures in our partner sites; the proposed 2Gbps link would only somewhat mitigate the effects of such a prolonged outage. It is important in each case that a realistic estimate of the likely outage time is given, to allow the system to react appropriately.

CMS

The PhEDEx transfer system, which steers the data flow between CMS computing centres, is fault-tolerant. If the connection to a Tier-1 centre breaks, then the system will back off and allow buffers at the Tier-0 to fill. These buffers should be large enough to cope with a six-day outage of any single Tier-1 centre; in practice, it is likely that the data would be diverted elsewhere for an outage of this length. Tier-1 to Tier-1 dataflows also use the OPN in the CMS computing model; a six-day outage during this period would also not be critical.

However, the implications of an outage in terms of catch-up should be considered. In order to restore the nominal data fractions of the centres, RAL would be required to take a substantially higher data rate for some period after recovery of the outage; there is some doubt over our ability, with current middleware, to control data rates on the network. Given that all three major experiments would be attempting catch-up at the same time, there is a serious possibility of congestion on the OPN, with potentially serious effects on throughput.

LHCb

The LHCb computing model has a failover dispatching system built into it. This is predicated upon the certainty of planned downtime for upgrades ...etc.. Should a target Tier1 centre be unreachable another is automatically sought for the data transfer. At the same time an entry is made in a scheduling system at the failover site to keep trying to place the data in its correct location as soon as possible.

This system will (by construction) cope with 6 x 1 day outages without serious consequence.

This system should cope with 2 x 3 day outages without serious consequence.

A period as long as 6 day has not been explicitly considered, and would give some cause for concern. We are seeking further clarification.
Appendix C: Assessment 2009

LHCOPN Resilience

26 March 2009 Robin Tasker (robin.tasker@stfc.ac.uk)
Updated 16th June 2009 Peter Clarke

Issue
1. The connection between the UK Tier 1 Centre and CERN across the LHCOPN has no resilience. The figure below shows how most other major countries achieve a resilient backup link. The UK is now the only major country (with the exception of Spain) without such a backup. The UK shows as “UK-T1-RAL”.

2. During recent tests the main links provider suffered a fibre cut between Frankfurt and Geneva, resulting in connection problems to the following Tier-1 sites for 36 hours around the 11th June. This affected the following Tier-1 sites: Taiwan, Canada, Italy, Germany, and Nordugrid. All were cut off but used their backup links to maintain connectivity. BNL and FNAL were running with reduced capability. This event served to emphasis that (i) breaks do happen and (ii) the can take out more than one Tier-1 connection.

Recommendation
3. It is recommended that a 4Gbits/s backup provision be established between the UK Tier 1 Centre and CERN to carry LHC data between those two sites only in the event that the primary 10Gbits/s path fails. It is noted that the 10Gbits/s path is currently used for both CERN / UK Tier 1 traffic and inter-Tier 1 traffic and in the first instance this traffic would be dropped on the assumption that such outages are in general of short duration. It would be feasible to manually restore inter-Tier 1 routing to mitigate against the effects of longer duration outages.

Risks and Costs
4. Once the LHC becomes operational loss of the existing path from UK Tier 1 Centre into the LHCOPN would increasingly cause difficulty for the UK dependent upon the length of such an outage.

5. If GridPP wish to provide a resilient connection into the LHCOPN the indicative the worst-case cost for the 4Gbits/s provision between RAL and CERN will be an install cost of £51,726 with an annual recurrent cost of £60,974 (all ex VAT). It is to be noted that the cost is likely to be less based upon the information provided in Table 1 and the associated Notes.

Background
6. Currently a 10Gbits/s lightpath is provided between the UK Tier 1 Centre at RAL and CERN which forms the UK connection into the LHCOPN. The LHCOPN has been established for the transfer of LHC data between CERN and the UK Tier 1 Centre, and additionally for the transfer of LHC data between Tier 1 sites worldwide.

7. The UK is currently the only member of the collaboration that has not provided any resilience in its connection to the LHCOPN. This has been a positive decision for 2008/9 based upon the assessment of the risk of circuit outage when set against the cost and the impact such an outage to GridPP’s interests.

8. This paper provides the cost for the introduction of a resilient 4Gbits/s path from the UK Tier 1 Centre to CERN which would be used in the event of failure on the primary circuit, and then only for the traffic between those two sites. All inter-Tier 1 LHC traffic would by default (and design) be dropped for the duration of the outage although for longer term problems manual intervention to determine routing could be considered.

9. 4 Gbit/s is somewhat arbitrary. It is chosen as the minimum which is comfortably above the OPN peak rates in 2008/09 (approx 2 Gbit/s). The costs for a full 10 Gbit/s link are given for comparison and it is though that we can upgrade pro rate at any time (although this is unlikely to be relevant in the middle of a short fallback scenario).

10. In arriving at the costs, JANET(UK) have taken the opportunity to take advantage of their recent re-engineering of the JANET Lightpath service (also called UK Light) to make use of Ethernet technology using an MPLS control plane to manage the actual bandwidth provisions. In so doing, it is proposed that the existing 10Gbits/s path from RAL to CERN is re-engineered across the Thames Valley Network to terminate in the Reading(Neos) PoP while the 4Gbits/s provision terminates at the Reading(Verizon) PoP. The paths from Reading to London Telecity will therefore be diverse, and JANET(UK) has conferred with DANTE to ensure the 4Gbits/s backup path takes the “northern” route to CERN.
11. Table 1 presents the figures provided by JANET(UK) for the indicative cost of the entire path although the service is offered through a combination of JANET(UK), DANTE and Switch. The costs shown will be subject to some change at the time of purchase as a consequence of actual component cost, the Euro exchange rate, and the negotiation with Switch of the attribution of cost of “their end” of the DANTE path across Europe.

12. Based upon the costs in Table 1, the worst-case cost for the 4Gbits/s provision between RAL and CERN will be an install cost of £51,726 with an annual recurrent costs of £60,974 (all ex VAT)

<table>
<thead>
<tr>
<th></th>
<th>Full 10Gbits/s cost</th>
<th>Pro-rata 4Gbits/s cost (@50%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Install</td>
<td>Recurrent</td>
</tr>
<tr>
<td>RAL to Neos(Reading) - 10Gbit/s [1]</td>
<td>60,000</td>
<td>6,000</td>
</tr>
<tr>
<td>Verizon(Reading) to Telecity [2]</td>
<td></td>
<td>26,448</td>
</tr>
<tr>
<td>Reading C-PoP Lighpath port (1 of 4)</td>
<td>6,000</td>
<td>500</td>
</tr>
<tr>
<td>Neos(Reading)(RNEP) to Telecity [3]</td>
<td>30,451</td>
<td></td>
</tr>
<tr>
<td>Telecity patching</td>
<td>1,000</td>
<td>12,500</td>
</tr>
<tr>
<td>Geant Access (1 end @40k Euro) [4] [5]</td>
<td>38,000</td>
<td></td>
</tr>
<tr>
<td>Totals [6]</td>
<td>97,451</td>
<td>83,448</td>
</tr>
</tbody>
</table>

Table 1: Itemised costs for the provision of 4Gbits/s on a resilient path between RAL and CERN

Notes:
1. Based up a new 10G wave going long way around TVN
2. Making use of separate RNEPs – Neos and Verizon - in Reading
3. Across different paths on the JANET core to London
4. SWITCH may charge additionally for their end of the connection @40k Euro or pro-rata – to be determined at time of order
5. Exchange rate 0.95
6. Excluding Vat
LHC Network Forward Look
2009/10
Introduction

This document is the first of an annual review and forward look for UK based networking capacity required for LHC operations. The scope includes performance and capacity for

- The UK Tier1 to OPN (for CERN and other Tier1s)
- The UK Tier1 to UK Tier2
- The Tier1 to JANET
- A view from the Tier2 sites themselves

UK Tier1 to OPN

Performance:

The UK Tier 1 connection to the LHCOPN operates at 10 Gbit/s and the usage statistics in Figure 1 show the inbound and outbound throughput over the LHCOPN during 2008/09. Over this period the network has been reliable with the usual maintenance outages. There have been no major outages which have impacted on the transfer of data between CERN and the UK Tier 1 Centre (note that this includes Tier 1 to Tier 1 traffic via CERN). A picture of the current LCG OPN network is included at the end of this report.

Figure 1b shows the same traffic as Figure 1 but this time during the STEP 09 tests. This shows rates peaking around 4 Gbit/s.

Note: Although no outage occurred for the UK link, on June 11 2009, during tests, the main links provider suffered a fibre cut between Frankfurt and Geneva, resulting in connection problems to the following Tier-1 sites for 36 hours. This affected the following Tier-1 sites: Taiwan, Canada, Italy, Germany, and Nordugrid. All were cut off but used their backup links to maintain connectivity. BNL and FNAL were running with reduced capability. This event served to emphasis that (i) breaks do happen and (ii) the can take out more than one Tier-1 connection.
Figure 1a: Traffic measured at the OPN Router at RAL between RAL Tier 1 Centre and CERN for 2008/09. Includes T1-T1 traffic.

Figure 1b: Traffic measured at the OPN Router at RAL between RAL Tier 1 Centre and CERN for STEP 09 tests. Includes T1-T1 traffic.

Forward look

In 2009 through to the end of running in 2010 the experiment forecasts with respect to current data rates are

- **ATLAS statement**: T1 to CERN traffic will be similar to 08/09 with a modest increase expected. T1-T1 traffic is expected to rise more steeply due to plans for more frequent re-processing.

- **CMS statement**: T1 to CERN traffic will be similar to 08/09 with a modest increase expected. T1-T1 traffic is expected to rise more steeply due to plans for more frequent re-processing. CMS plans assume a shared 10 Gbit/s OPN link.
• LHCb statement: Current rates are modest. LHCb also expects an increase due to more frequent re-processing resulting in approximately 0.6 Gbit/s.

In summary we see that data rates look like rising to take a large fraction of the 10 Gbit/s capacity by the end of 2010. The situation should be kept under close scrutiny. If the LHC OPN group recommend an increase from 10 Gbit/s the UK should be prepared to follow quickly.

Emergency planning

In 2008 the GridPP PMB carefully considered the likely traffic load from early LHC running and the likelihood of failures. It concluded that it would not be warranted in diverting funds to set in place a fallback mechanism in 2008.

In 2009, given the forecast and the importance of protecting the data flow in the light of the 2008 delays, the PMB has decided to investigate costs of fall-back solutions with diversely routed paths between the UK Tier 1 Centre and CERN for the LHCOPN.

To manage the cost, the option to maintain CERN / Tier 1 traffic in the event of a outage has been considered with the provision of 1 * 4Gbits/s on the backup path. The discussion with JANET(UK) has coincided with a re-engineering of the JANET Lightpath service which will make the service more cost effective and better able to offer flexibility. As a consequence the option readily to extend – at additional cost - the back-up provision to a full 1 * 10Gbits/s would be available in the future.

Both paths terminate at the “lightpath” router at RAL which makes fail-over routing straight forward. In the first instance it is intended that in the event of such a failure, the back up connection would carry the CERN (Tier 0) / UK Tier 1 traffic only with any Tier 1 / Tier 1 traffic being dropped. Were the incident to be prolonged then the Tier 1 / Tier 1 traffic could be manually re-routed.

External Tier 2 via JANET

Performance:

The external RAL connection to JANET is provided by a diversely routed 10Gbit/s provision connected directly to the JANET Core. By default all Tier1/Tier2 traffic makes use of this route.
Figure 2a shows the inbound and outbound throughput between routers on the RAL site that accounts for all Tier 1 / Tier 2 LHC traffic flows. For 2008/09 and Figure 2b during the STEP 09 period. Again we note that the scale has doubled between these periods.

![Figure 2a: Traffic measured at the SAR Router at RAL showing Tier 1 / Tier 2 LHC traffic flows during 2008/09](image)

![Figure 2b: Traffic measured at the SAR Router at RAL showing Tier 1 / Tier 2 LHC traffic flows during STEP 09](image)

Over the period JANET has been reliable and there has been ample headroom on the connection out of the RAL site as is shown in Figure 3a and 3b.
Figure 3a: Traffic measured at the SAR Router at RAL showing all site / JANET traffic flows during 2008/09

Figure 3b: Traffic measured at the SAR Router at RAL showing all site / JANET traffic flows during STEP 09

**Forward look**

In 2009 through to the end of running in 2010 the experiment forecasts

For Tier1 ⇔ Tier2 with respect to current data rates are

- **ATLAS statement**: No major change expected in 09/10.

- **CMS statement**: No major change expected in 09/10.

- **LHCb statement**: No significant data rates forecast.

The Thames Valley Network (TVN) is be the subject of an infrastructure re-procurement by JANET(UK) to deliver network services to institutions in the Thames Valley area covering the period 2010-2015. TVN has been very successful with its use of leased fibres and an optical transmission system which JANET(UK) manages directly. This mirrors what has been done in SuperJANET5 and
has enabled JANET(UK) to deliver both high-capacity resilient IP services and Lightpath services for specific projects.

JANET(UK) plans to take a similar approach for the new TVN, and as part of the preparation process JANET(UK) has consulted organisations which connect to JANET via the TVN about future service requirements.

As a part of that consultation, JANET(UK) were informed of the likely LHC network needs over that period. In essence, the RAL site expects its JANET traffic to increase markedly with 100Gbits/s provision likely to be required by the end of the new TVN contract period. Similarly its lightpath provision will grow, and for the LHC, multiple 10Gbit/s paths are likely to be required on the path to CERN with the additional capacity for specific lightpaths between the Tier 1 Centre and UK Tier 2 sites should that need arise.

## Tier-2

The table below summarises the comments made by local Tier-2 site admins in respect of the external WAN (internal problems are not reported here)

<table>
<thead>
<tr>
<th>WAN</th>
<th>Comments on WIKI</th>
<th>Forward look</th>
<th>Note/Action required</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKI-LT2-Brunel</td>
<td>Capped at 400 Mbit/s</td>
<td>Expected to increase to 1Gbit/s</td>
<td></td>
</tr>
<tr>
<td>UKI-LT2-IC-HEP</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKI-LT2-IC-LESC</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It can be seen that in almost all cases there are no reported problems in the last year. Most problems reported on the WIKI were largely internal.

The only common theme which emerges is that some sites note that they are limited to a 1 Gbit/s link shared with other users. Whilst there is no compelling case that this is a problem at the moment, it would seem prudent to recommend that all sites discuss this with their campus and have a view as to what would happen if the Tier-2 rate started to take > 50% of the shared 1 Gbit/s campus capacity.

Several campuses will already be connected to their MANs at N x 1 Gbit/s Ethernet where N >=2 and some have a Tier-2 connection which does not contend with other campus traffic.
<table>
<thead>
<tr>
<th>UKI-LT2-QMUL</th>
<th>No comment</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>UKI-LT2-RHUL</td>
<td>OK (problems internal)</td>
<td>Move of cluster IC-&gt;RHUL leads to 1 Gbit/s contended link</td>
<td>Important issue but not a JANET one as its internal to RHUL and LMN</td>
</tr>
<tr>
<td>UKI-LT2-CENTRAL</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKI-LT2-HEP</td>
<td>No comment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKI-NORTHGRID-LANCS-HEP</td>
<td>No comment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKI-NORTHGRID-LIV-HEP</td>
<td>Shared University 1Gbit/s link limiting transfers</td>
<td>Important issue but not a JANET one as its internal to LIV and NNW</td>
<td></td>
</tr>
<tr>
<td>UKI-NORTHGRID-MAN-HEP</td>
<td>No comment</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKI-NORTHGRID-SHEF-HEP</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKI-SCOTGRID-DURHAM</td>
<td>8 breaks in last year – 10 hours outage</td>
<td>1 Gbit/s link shared with all other users.</td>
<td></td>
</tr>
<tr>
<td>UKI-SCOTGRID-ECDF</td>
<td>1 Gbit/s connect to WAN, not shared with 2 Gbit/s campus connection</td>
<td>No foreseen problems</td>
<td></td>
</tr>
<tr>
<td>UKI-SCOTGRID-GLASGOW</td>
<td>OK</td>
<td>Arrangements OK for 2009</td>
<td></td>
</tr>
<tr>
<td>UKI-SOUTHGRID-BHAM-HEP</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UKI-SOUTHGRID-BRIS-HEP</td>
<td>OK</td>
<td>Link to SWERN being upgraded to 2.5 Gbit/s</td>
<td></td>
</tr>
<tr>
<td>UKI-SOUTHGRID-CAM-HEP</td>
<td>OK</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Summary

- The LHC OPN connection from RAL to CERN has been reliable and adequate for all traffic to date. There have been no major outages on our link in 08/09.

- Rates during tests have reached approximately 5 Gbit/s. The experiments forecast a modest increase in Tier-1 to Tier-0 and a larger increase in Tier-1 to Tier-1 traffic in 09/10. We expect to 10 Gbit/s link to be adequate in terms of capacity until the end of 2010.

- The OPN rate should be kept under close scrutiny in 2010 and if the rest of LCG moves to greater capacity for 2010+ we should be prepared to follow quickly.

- The experiments and GridPP are concerned at the Tier-1 to Tier-0 resilience.

- In respect of Tier-1 to Tier-2 traffic no major change is expected in 09/10 and specifically there is no case today for considering any dedicated links from Tier-1 to any Tier-2 site.

- A compendium of Tier-2 site comments on networking as seen locally is included. The only emergent issue is that sites limited to a 1 Gbit/s connection to their WAN (particularly if shared with the rest of campus traffic) should review their situation and draw up a plan plan for what would happen if their rates start to push towards this limit.

<table>
<thead>
<tr>
<th>EFDA-JET</th>
<th>OK</th>
</tr>
</thead>
<tbody>
<tr>
<td>UKI-SOUTHGRID-OX-HEP</td>
<td>2 Gbit/s Janet connection. Cluster shares 1 Gbit/s link but could be upgraded.</td>
</tr>
<tr>
<td>UKI-SOUTHGRID-RALPP</td>
<td>No comment</td>
</tr>
<tr>
<td>Grid Ireland csTCDie</td>
<td>No comment</td>
</tr>
</tbody>
</table>

