Network in Belle II

Silvio Pardi[1,*] *on behalf of Belle II Computing Group*

[1]INFN – Napoli Unit – Via Cintia. 80126, Napoli Italy

Abstract. Belle II has started the Phase 3 data taking with a fully equipped detector. The data flow at the maximum luminosity is expected to be 12PB of data/year and will be analysed by a cutting-edge computing infrastructure spread over 26 Countries. Several of the major computing centres for HEP in Europe, USA and Canada will store the second copy of RAW data. In this scenario, the international network infrastructure for research plays a key role in supporting and orchestrating all the activities of data analysis and replication. The large-scale network data challenge will also take advantage from LHCONE VRF service and the support of network experts of KEKCC, Belle II sites and NREN. The program of major upgrade in 2019 empowered the connection among Japan, Europe and USA over a 100Gb geographic ring. In this work, we summarize the network requirements needed to accomplish all the tasks provided by the Belle II computing model. We also highlight the status of the major network links that support and advance Belle II. Lastly, we present the results of the last Network Data Challenge campaign performed between KEK and the main RAW data centres with the additional usage of the Data Transfer Node service provided by GÉANT.

1 Introduction

Belle II collaboration [1] involves 26 countries/regions, around 120 institutions and more than 1000 people distributed all over the world. In 2019 the experiment is entered in the Phase 3, starting data taking with the fully equipped detector and performing the first scientific runs. The collaboration is now working hard to ramp up to the maximum luminosity, which is expected to be achieved in 2026. Then the data taking will continue up to 2028.

According to the current estimation [2], Belle II will collect more than 12PB of RAW data per year at the maximum luminosity, and the data will be processed over a distributed computing infrastructure. One full copy of the collected RAW data will be stored in Japan at the main site Kō Enerugī Kasokuki Kenkyū Kikō (KEK) [3] hosting the Tier-0 in the KEK Computing Centre (KEKCC), while a second copy will be distributed to five countries with shares of which 30% to USA, 20% to Italy, 20% to Germany, 15% to France and 15% to Canada. The sites responsible for data replication are called RAW Data Centres for Belle II, some of them are already host WLCG Tier1 [4] such as CNAF (Bologna Italy), KIT (Karlsruhe, Germany) and CC-IN2P3 (Lyon, France).

* Corresponding author: spardi@na.infn.it

According to the current assumptions, at the maximum luminosity the average throughput of the RAW data replication is estimated around 42TB/day outbound KEK resulting in different inbound traffic to RAW data centres depending on their RAW shares. In addition to the RAW data management, the Belle II Computing Model provides several activities that will produce traffic among all sites of the collaboration which compose the distributed infrastructure: i.e. skimming, analysis and MC production.

In order to accomplish all tasks, Belle II takes advantage of network services and support offered by the national research and education network (NREN) organisations and the sites. In last years several efforts have been spent by the experiment to estimate the network traffic, test the already existing links, and provide feedback to the international community. Belle II has also joined the main forums and working groups involving LHC experiments, NRENs, and all the stakeholders with the aim to be an active actor in the development of the network that can be used for the experiment.

In this paper we will present the current status of the network on which Belle II relies, the activities carried out for testing and monitoring the network, and the recent achievement in term of performance and tools put in place. In section 2 we describe the main aspects of the international network on which Belle II relies, followed by Section 3 and 4, in which we present the results of the Network Data Challenge activity and we describe the monitoring tools currently used. Finally, in Section 5 we summarize our work and discuss the next steps.

2 Belle II Network

The main strategic network infrastructure of Belle II experiment is represented by a set of international links connecting Japan to other continents.
Early 2019 the Japanese academic backbone network SINET[5] completed a major upgrade of its international links to a 100Gbps ring which globally connects Japan through the path Tokyo, Amsterdam, New York and Los Angeles (Figure 1). In addition another 100Gbps link to Singapore has been established which may allow to have a secondary link to European countries in the future.

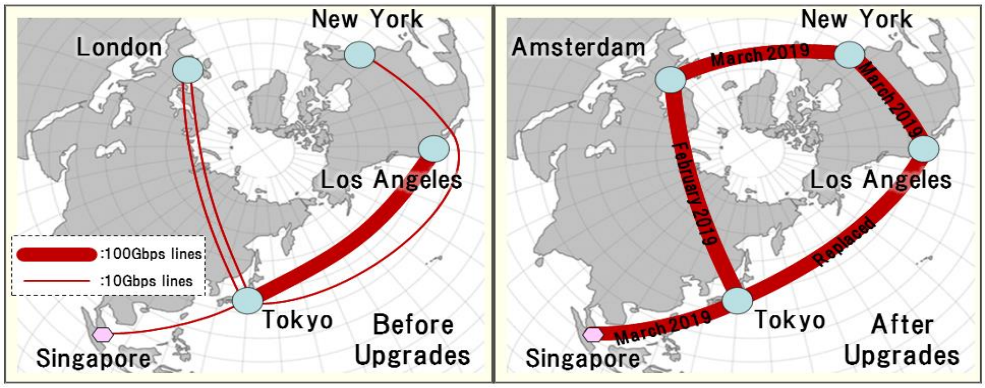

Figure 1 – The previous SINET global network connection is shown on the left side and the new 100Gbps global ring currently in place is shown on the right [6].

Note that the SINET 100Gbps global ring is not dedicated to Belle II experiment and it is shared with other communities including LHC experiments.

In addition to this key infrastructure, Belle II can take advantage of a part of the high speed links provided by Large Hadron Collider Optical Private Network (LHCOPN) [8].

Indeed, 5 out of 14 sites in LHCOPN, such as IT-INFN-CNAF, FR-CCIN2P3, DE-KIT, KR-KISTI, US-T1-NBL, are considered as the Belle II data centres (Figure 2). Thanks to an agreement reached among WLCG Tier-1 sites participating in LHCOPN in 2017, Belle II can use the LHCOPN links for the internal traffic among those data centres, unless jeopardizing LHC operations. This agreement also helps to simplify the site network configuration.

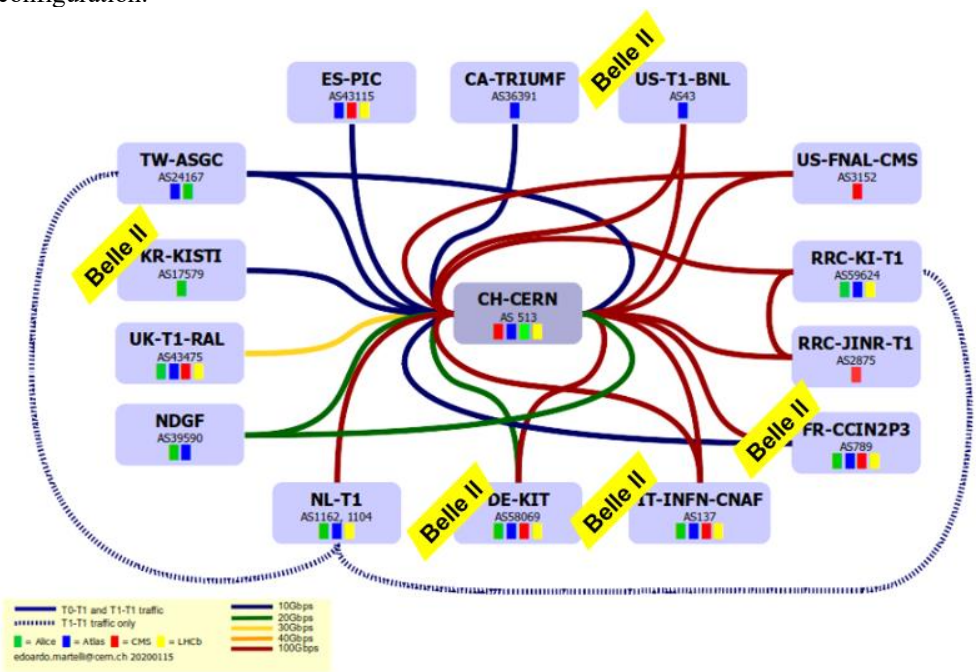

Figure 2 – LHCOPN network connectivity scheme highlighting the Belle II Data Centres [7].

As regards the layer 3 service, Belle II has joined the Large Hadron Collider Open Network Environment (LHCONE) [8] network which connects several sites in the collaboration including all RAW Data Centres and KEK, the latter connected at 40Gbps in the current setup.

Today the distributed infrastructure provides around 13PB of disk space from about 60 active sites. Among them 30% of sites, which actually own the more than 80% of Belle II storage and computing resources, are running over LHCONE. The rest of sites are working on general internet provider (IP) via their NREN.

The main Belle II computing challenges from the network point of view, is to send a large amount of data, which in size is similar order of LHC experiment RUN1/RUN2, over a high latency environment, without reserved resources.

3 Network Data Challenge

3.1 KEK vs EU Data Centres

For the several years Belle II has run Network Data Challenge campaigns, focussed on measuring the maximum achievable throughput over the available links connected to KEK from the main data centres of the collaboration. Tests have been performed by sending massive transfer jobs through the File Transfer Service (FTS), which is one of the most

used file transfer service in High Energy Physics, and monitoring in the meanwhile the bandwidth, both peaks and average, failure frequency, timeout etc.

This activity has allowed to highlight bottlenecks, issues related WAN and LAN and has contributed to improve the overall performance by giving feedbacks to the NREN and sites.

In 2019 a massive test campaign has been performed with the aim to measure the throughput over the KEK 40Gbps link to the main European data centres, reached through the new SINET 100Gbps global ring, with 170ms of RTT on average. The participating sites were KEK, CNAF, DESY, IN2P3, KIT, NAPOLI and SIGENT.

From 15th May to 18th May 2019 more than 40TB has been sent from KEK to EU data centres and vice versa using FTS jobs, which each FTS job composes of 100 files and each file has 10GB.

Results, summarized in Table 1, show that we were able to reach a peak greater than the 89% of the maximum available bandwidth, sending data from KEK storage to storages at the European sites, and 87% in the opposite direction.

Table 1 – Result of Network Data Challenge 2019

Source	Destination	#files	#streams	Peak (Gbps)	Average (Gbps)
KEK	CNAF, DESY, IN2P3, KIT, NAPOLI and SIGENT	100	16	35.8	13.5
CNAF, DESY, IN2P3, KIT, NAPOLI and SIGENT	KEK	100	16	34,9	11,8

Network usage has been monitored using multiple tools, such as: the dashboard of FTS server in BNL, the GÉANT's CACTI [9] portal which shows the peering in Amsterdam between the pan-European data network for the research and education community GÉANT[10] and SINET, and the internal KEK Grafana dashboard which shows the traffic over the 40Gbps link through LHCONE (Figure 3).

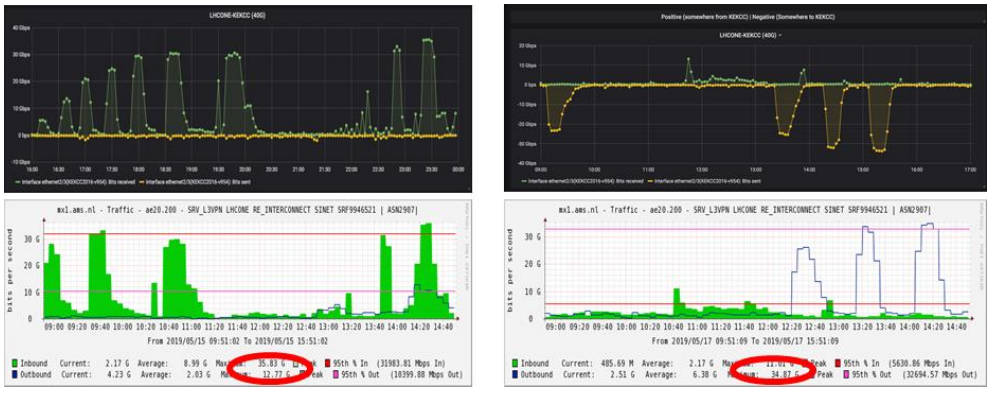

Figure 3 – Network Monitoring during Network Data Challenge 2019

3.2 GÉANT DTN Service

The Network Data Challenge campaign previously described, has been performed testing end-to-end data transfer using grid Storage Resource Manage (SRM) protocol [11]. In order

to measure the network speed trying to be as much as possible free from site effects (i.e. storage limitation due to file systems, grid protocol, LAN configuration, etc), we decide to double-check the archived results using the Data Transfer Node service (DTN) provided by GÉANT[12]. The DTN service consists of a pair of servers connected in strategic points on the network, one in London and the other one in Paris, and optimized for 100G transfer. Each server is already configured to perform a series of network tests in particular using iperf3 tool.

The test scenario is described in Figure 4. For Europe, the London server with a single 100G card was chosen, while at KEK the production storage connected with 4x10G cards was used. The test was carried out in the KEK-> EU direction, which will be the route used for the replication of the second copy of the RAW data. The Round Trip Time measured between the two sites was equal to 161 ms.

Figure 4 – Setup of the test performed between KEK Storage and DTN services in London.

Figure 5 shows the results obtained by starting 10 and 20 iperf3 sessions respectively, with 16 parallel streams each. In particular the graphs show that we can reach the maximum peak of 37Gbps saturating over 92% of the total band with 10 concurrent flows.

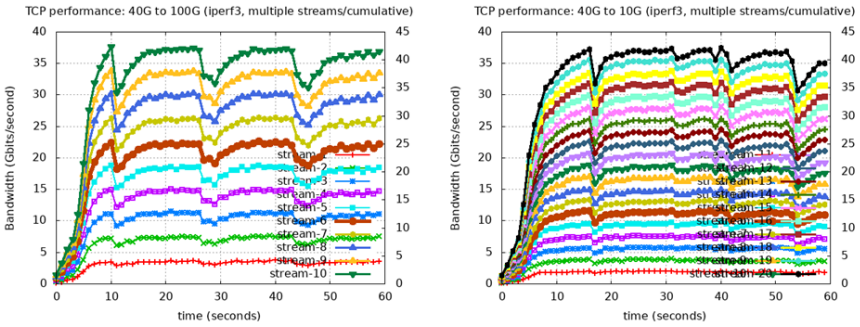

Figure 5 – The graphs in the figure show the performances achieved with iperf3, each line represents the contribution of a single transfer to the total band. On the left the test carried out with 10 parallel transfers, on the right the test result carried out with 20 parallel transfers. The plots were generated using a python script collection [13].

4 Monitoring Tools

In order to monitor the Belle II network traffic, several tools and platforms have been considered for collecting information from different points of observation. Some of them have been largely used during the Data Challenge campaign and the normal operation activities. The main tools currently in place are the following:

- Tools provided by data centers
- Graphs provided by network operators
- Perfsonar Mesh[14]
- FTS monitoring
- Internal Tools integrated in DIRAC[15]

Considering that more than 80% of compute and storage resources, the Tier-0 and all the raw data centers are connected to LHCONE, in the next paragraphs we focus on the specific traffic in LHCONE for our analysis.

4.1 LHCONE Network Traffic Monitor

From the local Grafana monitoring service in KEKCC where the Tier-0 of Belle II experiment is hosted, we can monitor the 40Gbps link to LHCONE. The graph in Figure 6 shows the network activities in the last quarter in which we appreciate some peaks higher than 15 Gbps in both direction and a bursty pattern of network usage, with some small period of sustained traffic of around 10Gbps.

Figure 6 - Monitoring of the 40Gbps LHCONE link of the KEKCC data centre

Belle II traffic from KEKCC to major European data centres can be monitored through GÉANT's CACTI portal. In particular, it is possible to monitor the traffic on the LHCONE SINET-GÉANT peering in the Tokyo-Amsterdam section.
In Figure 7 we have the last three months of statistics as of writing, currently it is not possible to discriminate the traffic related to KEK, however comparing with what is shown in Figure 6 we can deduce that at present it is not dominant.

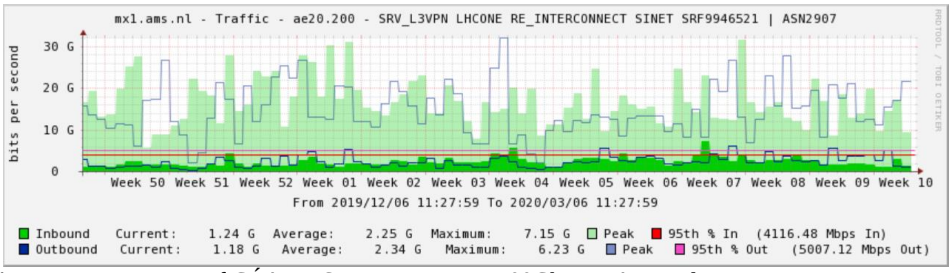

Figure 7 - Monitoring of GÉANT-SINET peering at 100Gbps in Amsterdam

Another traffic observation point is provided by the CANARIE network provider which also provides traffic discrimination on incoming/outgoing of KEK in LHCONE. In Figure 8 in light green and blue it is possible to appreciate even a moderate traffic mostly due to the activities of MC, analysis and testing.

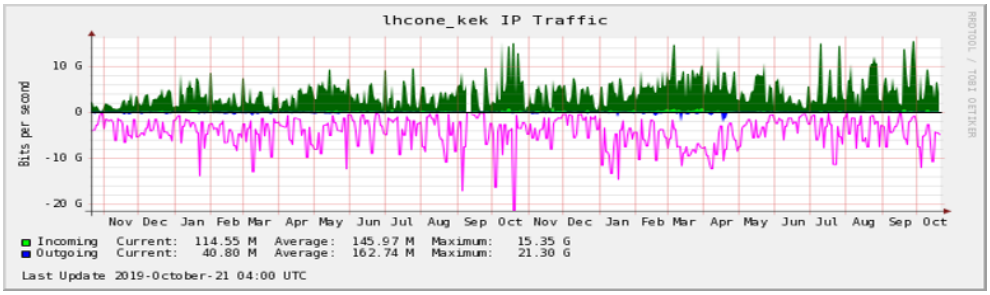

Figure 8 – KEK Traffic on the CANARIE LHCONE Link

4.2 Perfsonar Mesh

The Belle II experiment uses the perfsonar service to monitor site reachability and the fundamental parameters of the global network that connects the main computing centres of the collaboration. In the last year, the mesh representing the connection status of all raw data centres has been consolidated (Figure 9). In addition, an IPv6 map helps keep track of sites that have implemented the new version of the IP protocol.

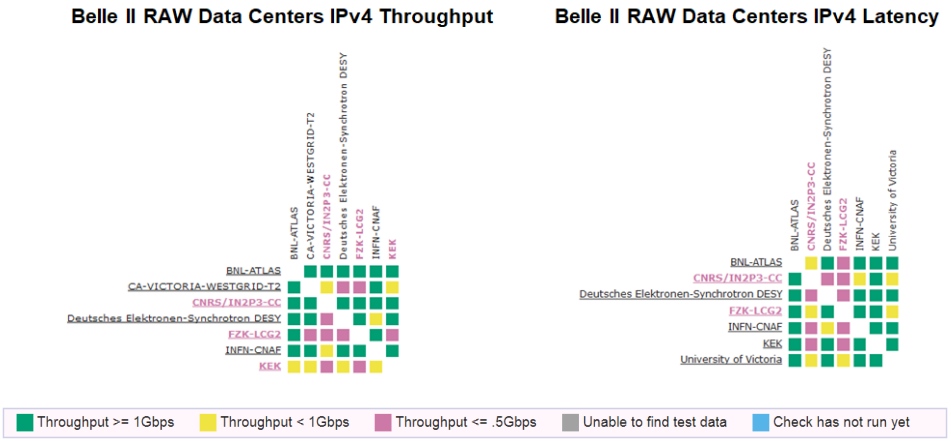

Figure 9 – Belle II Maddash [14] of all RAW Data Centres

By using the perfsonar historic data reports, we can appreciate the 10ms latency decrease from KEK to the KIT data centre in Germany (Figure 10). The result was obtained as a follow-up of the update of the Japan-Europe connection made by SINET in the first half of 2019, as described in section 2.

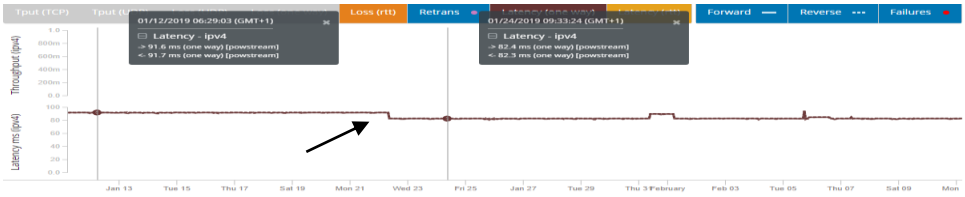

Figure10 – Latency between KEK Perfsonar and KIT Perfsonar in early January. The arrow in the picture shows the effect of activation of the 100G link between SINET and GÉANT

5 Conclusions

Belle II actively participates in network activities in order to contribute to monitor and improve the performance of one of the experiment's greatest assets.

In 2019 several important results have been achieved, and the new 100G ring has opened new opportunities and new connectivity scenarios.

The Data Challenge activity run vs the European Data Centres and double-checked with the usage of new tools such as the GÉANT DTN, made it possible to confirm that the main Network requirements to copy and reprocess RAW Data are archived on the international links. This makes possible to concentrate next tests on the connection site-to-site finalized to optimize LAN, storages configuration and tape systems.

Network monitoring now becomes one of the key aspects for controlling activities, for optimizing transfers and for improving troubleshooting. Some steps have been taken that allow the Belle II community to globally check the traffic on the main links, and the health of the network through perfsonar. Other activities related to the extension of monitoring and the study for the recognition of traffic will be followed in the context of the main international working group.

Acknowledgments

Author wish to acknowledge the assistance and the large efforts spent by all members of the Belle II network mailing list, the staff team of KEKCC, site administrators, the GÉANT Team for providing the DTN service.

References

1. Belle II Official Web Page - https://www.belle2.org/
2. S. Pardi et al. *"Computing at Belle II"* - Nucl.Part.Phys.Proc. **273-275,** 950-956 (2016)
3. KEK web page - https://www.kek.jp/en/
4. WLCG web page https://wlcg.web.cern.ch/
5. SINET web page https://www.sinet.ad.jp/en/aboutsinet-en

6. Original Picture of SINET 100G Global Ring
   https://www.nii.ac.jp/en/news/release/2019/0301.html
7. Original Picture of LHCOPN
   https://twiki.cern.ch/twiki/bin/view/LHCOPN/OverallNetworkMaps
8. E Martelli and S Stancu 2015 J. Phys.: Conf. Ser.664 052025
9. GÉANT's CACTI Portal
   https://tools.geant.org/portal/links/p-cacti/graph_view.php?action=tree&tree_id=14&leaf_id=710&select_first=true
10. GÉANT web page https://www.geant.org/
11. F Donno et al 2008 J. Phys.: Conf. Ser.119 062028
12. DTN Reference page - https://wiki.GÉANT.org/display/DTN/
13. M. Babik - Script collection for plot iperf3 results
    https://gitlab.cern.ch/mbabik/perfsonar-100g-testing
14. Maddash web page https://psmad.opensciencegrid.org/maddash-webui/
15. A Tsaregorodtsev and the DIRAC Project 2014 J. Phys.: Conf. Ser.513 032096