Title
Machine learning-based Analysis of COVID-19 Pandemic Impact on US Research Networks

Journal
ACM SIGCOMM COMPUTER COMMUNICATION REVIEW, 51(4)

ISSN
0146-4833

Authors
Kiran, Mariam
Campbell, Scott
Wala, Fatema Bannat
et al.

Publication Date
2021

Peer reviewed

Machine learning-based Analysis of COVID-19 Pandemic Impact on US Research Networks

Mariam Kiran*
mkiran@es.net
Energy Sciences Network (ESnet)
Berkeley, CA, USA

Scott Campbell*
scottc@es.net
Energy Sciences Network (ESnet)
Berkeley, CA, USA

Fatema Bannat Wala*
fatemabw@es.net
Energy Sciences Network (ESnet)
Berkeley, CA, USA

Nick Buraglio
buraglio@es.net
Energy Sciences Network (ESnet)
Berkeley, USA

Inder Monga
imonga@es.net
Energy Sciences Network (ESnet)
Berkeley, USA

ABSTRACT

This study explores how fallout from the changing public health policy around COVID-19 has changed how researchers access and process their science experiments. Using a combination of techniques from statistical analysis and machine learning, we conduct a retrospective analysis of historical network data for a period around the stay-at-home orders that took place in March 2020. Our analysis takes data from the entire ESnet infrastructure to explore DOE high-performance computing (HPC) resources at OLCF, ALCF, and NERSC, as well as User sites such as PNNL and JLAB. We look at detecting and quantifying changes in site activity using a combination of t-Distributed Stochastic Neighbor Embedding (t-SNE) and decision tree analysis. Our findings bring insights into the working patterns and impact on data volume movements, particularly during late-night hours and weekends.

CCS CONCEPTS

• **Networks** → **Network performance evaluation**;

KEYWORDS

research networks, COVID-19, clustering, network performance

1 INTRODUCTION

The current coronavirus pandemic has caused economic, educational, and industrial disruption across the world [10]. The World Health Organization declared an emergency on January 30th, 2020 and a global pandemic on March 11th, 2020, triggering numerous intervention strategies across many countries in attempts to control the virus spread. Within the science and research domain, many scientists were left stranded with limited access to lab facilities and scheduled experiments, particularly where experiments had to be conducted in situ. Limits on international travel, constrained resources, and long periods of closed stay-at-home restrictions, also create potential repercussions to long term productivity. As staff began working from home, the increased use of VPNs and remote communication tools like Zoom created challenges for commercial networks like AT&T in addressing this surge of network demand [1, 2].

The R&E (Research and Education) networks saw the same sort of changes as these commercial and mobile network providers. The U.S. Department of Energy (DOE) Energy Sciences Network (ESnet), shown in Figure 2, provides international high-speed connectivity to over 55 research sites and universities. Similar to many other wide area networks (WAN) networks, ESnet employs multiple connections to other organizations which provide performance access to commodity resources, experiments, and other R&E networks like Internet2, JANET, and GEANT. ESnet is responsible for carrying regular user traffic as well as science data generated from experimental facilities at National Labs, the LHC (Large Hadron Collider) at CERN, and universities and collaborators worldwide. To ensure reliability and availability, service is provisioned to account for unexpected traffic surges [7] which allows experiments to grow, while keeping costs down [6]. Figure 1 shows the total yearly traffic volumes from 2018-2020 over the high performance computing facilities.

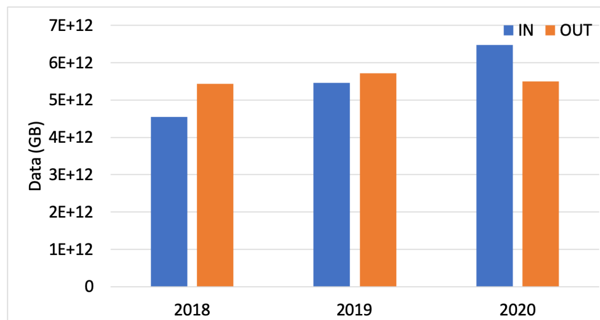


Figure 1: Total Traffic in/out DOE Compute Facilities - 2018, 2019, 2020.

Contrary to initial expectations that efforts to dampen the impact of COVID-19 would impact science negatively, these graphs show that the volume of science traffic kept growing during the stay-at-home orders. During our analysis, we also found that a large number of DOE science experiments have been automated, allowing scientists to operate instruments remotely through specialized software and intuitive processes developed to streamline their science experimentation.

In this paper we study network traffic patterns across the DOE networks during the transition period before, during, and after the period of time when the WHO declared the global health emergency a Pandemic (March 11th, 2020). Specifically, we look at three

*Authors contributed equally to this research.

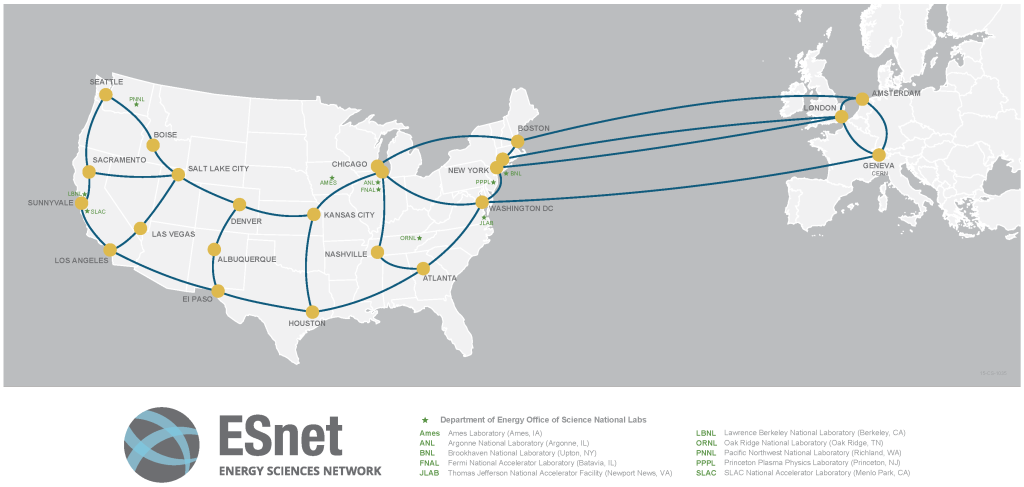


Figure 2: ESNet Network Map.

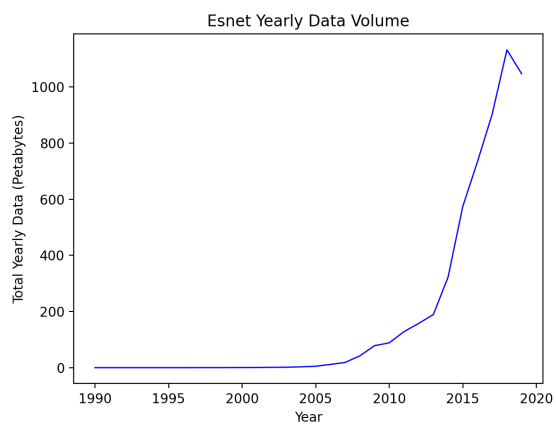


Figure 3: Yearly traffic totals since January, 1990.

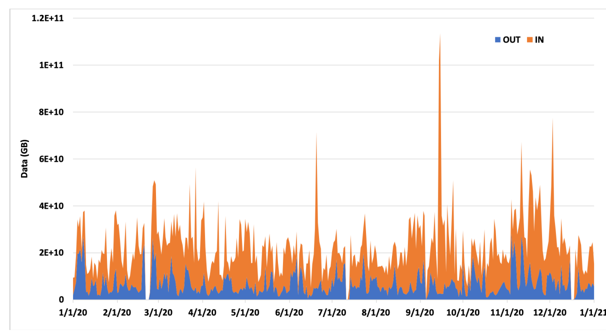


Figure 4: 2020 Year SNMP Traffic Traces on NERSC.

high-performance computing facilities and two user-dependent science facilities (Details: https://science.osti.gov/User-Facilities) as a proxy for larger scale behavior. In our network analysis, we study multiple sources such as VPN access logs, Netflow records, and SNMP interface counters to study changes in network use during the months January to June in 2020. We use both statistical and machine learning algorithm to find inherent properties and clustering in the data analyzed. Our main findings show that (1) there was no significant change in the network traffic behaviors of the HPC sites, (2) user facilities demonstrated new traffic profiles after March 11th, and (3) significant changes in users working patterns both in terms of day of the week and hour of the day. We witness ample evidence that remote work has changed the distribution of network traffic in commodity peerings - significantly more traffic between home and sites as well as sites and cloud providers. Some of these changes will be explored in some detail later in the paper. Significantly, with the onset of working from home, we observed a change in typical working hours away from a "normal" 9-5 morning and afternoon to more very early and very late times. This does raise questions about how these working patterns are impacting home and family lives, as well as what long-term effects this will have on staff stress and burnout.

The rest of this paper is organized as follows. Section 2 describes the measurement data set and network statistics taken from ESnet consisting of SNMP, Netflow, and VPN login data. Section 3 describes overall traffic observations, including methods, and detailed analysis of sites, user facilities and commercial peerings. Section 4 focuses on observations around changes in working patterns. Finally, we present related work, lessons learned, and conclude the paper with dominant patterns observed.

2 DATASETS

The ESnet network backbone consists of a high-performance international fiber optic infrastructure carrying IP traffic to sites, facilities, and instruments worldwide (Figure 2). Unlike most commodity providers, the ESnet network has been engineered to provide high-performance data movement for experimental facilities, specializing in the delivery of high-bandwidth and deadline-driven traffic with minimum packet loss. The capacity of network links is measured in terms of data transfer rates (e.g. megabits/second (Mbps) or gigabits/second(Gbps)) (Figure 4). To validate expectations around performance, these links are monitored for excess utilization and errors which can create packet loss, jitter, and latency. This is similar in cloud network providers to measure performance [11, 12].

SNMP Statistics. The Simple Network Management Protocol (SNMP) generates statistics that provide data about network interface usage in terms of bytes, packets, and errors. It is considered a core component of network monitoring and is ubiquitous across most networking and compute systems. At ESnet, values are gathered every 5 minutes from router interfaces as simple counters [3]. Because SNMP statistics focus on countable items, they are good for volume measurements and error detection but not to provide protocol information about individual flows. SNMP collects both device and application-level data from the perspective of the system creating the logs.

Netflow Records. While SNMP excels at counting measurements, protocol details are normally captured by Netflow records. There records provide a convenient way to summarize network flows in the form of standardized records [8]. A flow is defined as a unidirectional sequence of packets with some common characteristic that passes through a network device. Flow records can be generated on a sampled basis as well to address issues with performance on large networks. For some of our analysis, we used a summarizing heuristic to generate bidirectional (connection) data from raw flow records.

VPN Connections. As described in the Human Interactions section, we use VPN logins to measure changes in work behavior such as changes in the day of the week or hour of the day that an account is normally seen. Our data set includes a summary count of all logins and unique accounts per hour to the production LBNL VPN service between Jan 1, 2020 and Dec 31, 2020. No information was used that could identify an individual user.

3 TRAFFIC EVOLUTION DURING 2019, 2020

As suggested in the introduction, the first thing that we will examine is changes in overall network traffic characteristics during the period before, during, and after the transition to working from home. In order to better understand the details of specific sites, we look at a handful of locations rather than the aggregate measurements from the network at large. This proxy for behavior is done to simplify the data analysis and to be able to look at traffic in terms of in/out of sites.

The following sections provide a detailed overview of the techniques used to analyze the HPC and experimental user facilities.

3.1 Decision Trees to Identify Dominant Features

From machine learning techniques, we can use decision trees to identify dominant features and their differences in the two clusters formed - Pre and Post COVID. A decision tree is a classifier approach to recursively partition the data sets into rule space. The tree constructs a root, that follows binary rules, either true or false, to determine the path taken along to internal nodes and leaves. As we explore this path, rules are extracted, which lead to leaves as terminal or decision nodes. The objective is to partition the samples based on the data into the purest sample possible. This measurement of purity is determined by the Gini index or entropy function.

During classification, a lower Gini index means that most samples belong to one class label. We label the NetFlow records as pre and post covid, to construct decision trees for each location. We also prune the trees using a greedy approach to find optimum tree depth, number of leaf nodes, and number of samples per leaf. This allows only the important rules to be found in earlier tree branches, rather than letting it grow to less important leaf nodes.

3.2 t-SNE Analysis

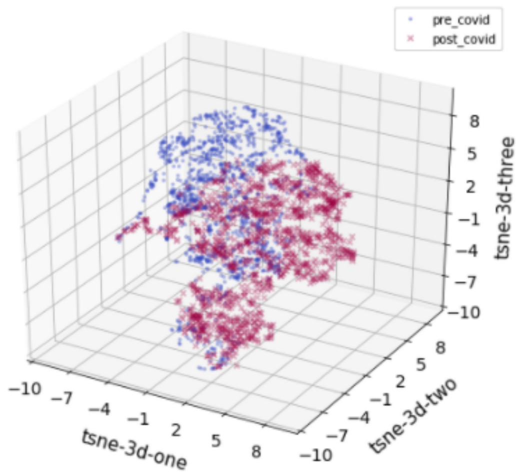
As a visualization approach, t-SNE (t-Distributed stochastic neighbor embedding) is a mathematical method for dimensional reduction in complex data sets, and is mainly used for better visualization of related components. The more familiar Principal Component Analysis (PCA), identifies correlations between high and low dimensional data and tries to provide a minimum number of variables that keeps the maximum amount of variation or information about how the original data is distributed. For t-SNE instead of relying on linear algebra, it uses a probabilistic technique to take a set of points in a high-dimensional space and find an equivalent representation of those points in a lower-dimensional space. The algorithm is non-linear and adapts to the underlying data, performing different transformations on different regions in such a way as to preserve locality as defined by the perplexity variable.

There are some interesting problems with using t-SNE which are more subtle than PCA - axes in the low dimensional space don't have a particular meaning. In fact, one could arbitrarily rotate the low dimensional points and the t-SNE cost function wouldn't change. Furthermore, t-SNE doesn't construct explicit mappings relating the high dimensional and low dimensional spaces.

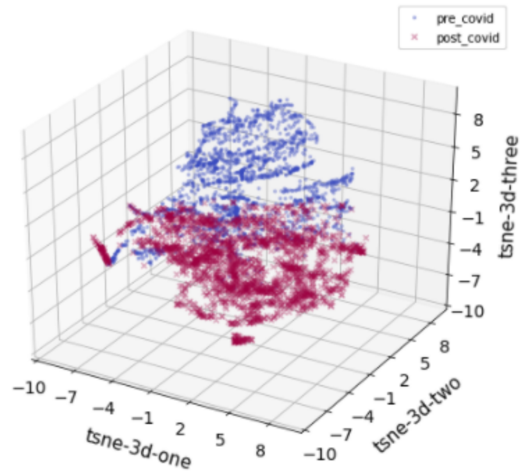
In generating the t-SNE images found here, we used the following values: perplexity=40; steps=300.

3.3 High-Performance Computing Centers

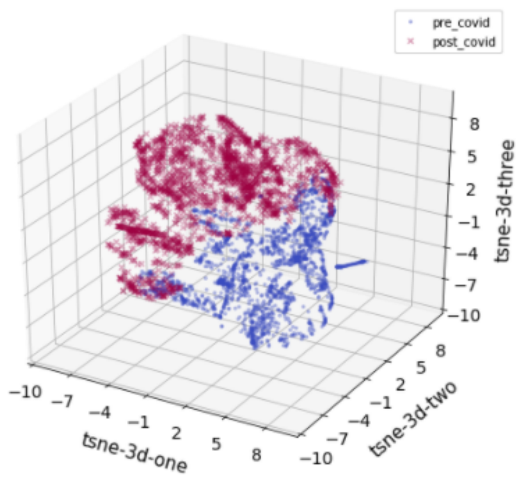
The HPC facilities are usually dominated by the data transfers that occur over the network verses the normal user behavioural network traffic. Hence, detailed study of Netflow records provides more information about protocols, addresses, port numbers, and packet/byte counts available which would not be apparent in the SNMP statistics. The following sub-sections cover the analysis of Netflow traffic for HPC and interesting network traffic trends we observed by visualizing the data collected over the period of pre/post covid months.



(a) OLCF.



(b) NERSC.



(c) ALCF.

Figure 5: t-SNE visualization of Pre and PostCovid Traffic across Supercomputing Facilities.

3.3.1 t-SNE Results. Figure 5 shows t-SNE plots to visualize how the network traffic looks across the two groups of pre/post-Covid months in the computing facilities. Table 1 shows the features being captured in the Netflow data. When all of the features listed in the table were used to visualize the data, we got mixed clusters that weren't showing any specific traffic pattern change in pre and post covid months. Since the HPC facilities traffic is heavily biased by the data transfers, by filtering out the bytes and packet counts and only using the fields that show network connection statistics, i.e. unique server addresses and ports, we got very distinctive clusters for network connections pre/post covid groups, as shown in Figure 5. This t-SNE analysis on sub-set of network connections fields captured in Netflow data for HPC facilities clearly shows that there were distinctive changes in the network traffic during pre/post covid period. The decision tree analysis of various fields show what caused this significant distinction between the pre/post covid groups, which is discussed in the next sub-section.

3.3.2 Decision Tree Results. The decision tree analysis shows which of the features in the two clusters as shown in the t-SNE visualization in Figure 5 were distinct. Particularly, there is a significant change in the number of distinct Server logins (such as in the case of NERSC (Figure 12b)). The features identified in the decision tree show that there are distinct changes in TCP and ICMP traffic profiles in the various sites. For example in Figure 12b, the decision tree shows that 'bytesinbound:1', 'bytesoutbound:1' show clear rules being identified in the clusters 0 (precovid) and 1 (postcovid). The left and right branches in the decision tree show the dominant feature properties that push the samples into the corresponding cluster.

3.4 User Facilities

The user facilities network traffic show the user behaviour during the pre/post covid months, which is expected to be different than the characteristics of the network traffic of HPC facilities. Hence, it would be interesting to see if there were any distinctive changes observed in the network traffic for the user facilities and if there is any difference observed as compared to the HPC facilities, if at all. A detail classification of Netflow data was also done on user facilities using t-SNE, and decision trees to show the more prominent changes in network traffic that happened during the pre/post covid months.

3.4.1 t-SNE Results. Figure 6a and Figure 6b show Netflow data mapped during the pre-covid and post-covid months and the changes in user behavior during the same. The feature set used is shown in Table 1. Unlike HPC facilities, the network traffic isn't biased by data transfers and hence the complete feature set was used without any specific filtering of fields. These data sets have a larger proportion of on-site users than HPC facilities, so traffic from these sites would map the users' behavior during the months. In the case of user facilities, we observed very distinctive clusters for user activity during pre/post covid months for the complete feature set, which shows that the user behaviour changed considerably during those months. The features that were seen changed predominantly were analyzed using decision tree analysis and discussed in the next sub-section.

3.4.2 Decision Tree Results. Figure 6 shows the overall traffic patterns during the pre-covid and post-covid months for user facilities. There is a significant change seen in the two clusters. The corresponding decision trees (Figure 14) show distinct behaviors in unique server logins, the packet sizes and number of bytes in both clusters are similar, showing the same kind of experiments were running, but people were logging in from home more after March 11th, 2020. For example, the dominant feature rules 'ServerPortIn:6' for JLAB and 'ServerPortIn: 1' for PNNL show dominant rules followed in the two clusters formed. This feature reflects unique servers logging in for TCP and ICMP traffic correspondingly.

Type of feature (TCP(:6), UDP(:17), ICMP(:1))	Feature description
Bytes Inbound	Integer
Bytes Outbound	Integer
Packet Count Inbound	Integer
Packet Count Outbound	Integer
Unique Server IP Inbound recorded hourly	Integer
Unique Server IP Outbound recorded hourly	Integer
Unique Server Port Inbound recorded hourly	Integer
Unique Server Port Outbound recorded hourly	Integer

Table 1: Features from hourly NetFlow summaries used in t-SNE and decision tree analysis.

3.5 Network Performance Statistics and Peering

ESnet is an international network with a rich and robust collection of connectivity peers. These peers serve to provide access, redundancy, and resiliency to the various labs, experiments and collaborators who represent our customers. Since the stay at home orders came out in the United States, there has been a very clear increase in the volume of traffic between ESnet and our various commercial peerings, as seen in Figure 7 and Figure 8. This increase in traffic to commercial and residential internet providers can be identified as beginning at the start of stay-at-home orders and continuing throughout. While overall traffic load to national, regional, and local residential internet providers rose dramatically across ESnet, the difference in increase between Internet protocol version 4 (IPv4) and internet protocol version 6 (IPv6) was apparent. With IPv4 the increase was on the order of 5x, while for IPv6 it was closer to 2x which also provides an inference about modern implementation practices in the commodity space around IPv6.

4 CHANGES IN WORKING PATTERNS DURING 2019, 2020

While a detailed analysis of SNMP counters and flow records suggest that some changes are happening in the way that scientists and lab employees are doing there work, there are a number of other data sources that can be used to identify the details here.

For example, one way of measuring changes in the work habits of users and staff is to look at how remote access usage changes over time. Remote access normally uses a virtual private network (VPN) which is an access device used to securely provide access to sensitive data from an off-site location. This access can be examined by either directly examining remote access authentication logs (highest value in terms of identity and access timing), or looking at patterns from the network traffic itself. These patterns provide some insight into behavior without having to look at logs but at the cost of precision and individual login granularity.

In this section, we will look at anonymized LBNL VPN login records from Jan-Dec of 2020 as well as IPSec traffic across ESnet as transited via various commodity peerings.

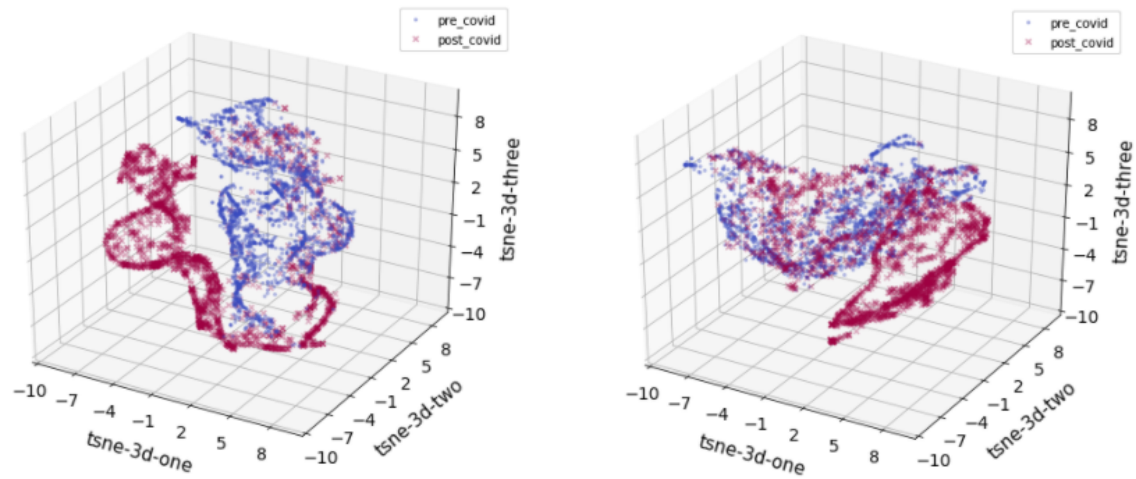
4.1 LBNL VPN Login Records

As already suggested, VPN login records provide high-resolution data for identity, start time, and duration for remote access usage. For this part of the study, we requested anonymized login records for the LBNL institutional VPN, summarized by the number of unique logins and the total number of logins per hour for the entire 2020 calendar year. This data set should contain most of the off-site interactions with the Lab by employees and staff. Summary statistics provide a good idea of behaviors without having to address login anonymization. We chose to look at LBNL data rather than ESnet since the total staff count (less than 100) is not large enough to strongly generalize changes in work habits.

4.2 Login Time of Day

In looking at the VPN data, we hope to determine if there have been changes in the *time of day* and the *day of the week* for staff remote access between January and December of 2020. To reduce the effects of noise and to amplify trends we did two things. First, we took the 24-hour day and broke it into 4 blocks of 6-hours starting at midnight (00) in order to create an implicit structure loosely characteristic of a "normal" working day. In addition to binning, the graphs represent a 20-day rolling average to reduce the impact of outlier values.

The three marked dates in Figure 9 represent a number of specific events which are important in understanding the Covid response



(a) JLAB. **(b) PNNL.**

Figure 6: t-SNE visualization of Pre and PostCovid traffic across User Facilities.

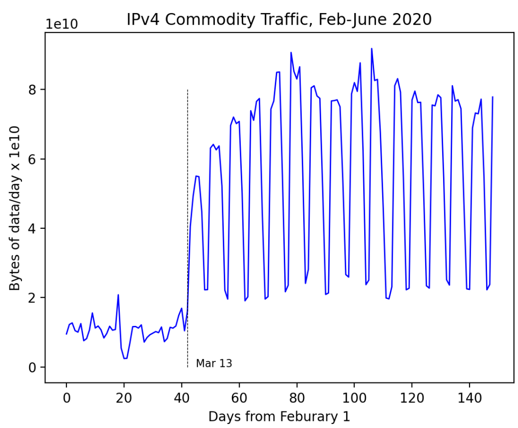


Figure 7: IPV4 traffic between ESnet and commodity residential internet providers.

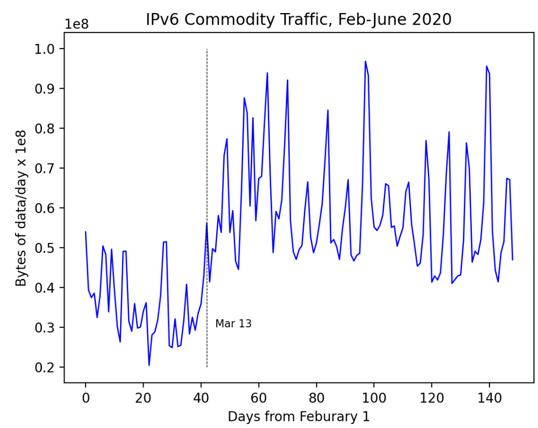


Figure 8: IPv6 traffic between ESnet and commodity residential internet providers.

at LBNL. These are the combined transition to curtailed operations (March 13th) and Bay Area shelter in place (March 16th), the Oct 5th pilot where 300 workers let back into work, and the Nov 2nd pilot where 250 additional workers were allowed to return. These time periods are named Phase 1 through Phase 4 in Table 2.

Starting with Figure 9 we look at distribution of VPN logins broken down by time of day. The graph is showing the percent of traffic in each of the four time blocks rather than an absolute count of logins. The most significant takeaways from Fig. 9 are the reduction in logins during the regular workday morning and afternoon, and the steady increase in logins in the evenings and very early mornings. Specifically in the transition between Phase 1 and Phase 2 the proportion of logins between midnight and 6AM increased by 3.04%. During the same time period, the proportion of logins between 6AM and noon decreased by 5.15%. When the transition period between Phase 1 and 2 is removed from the averages, the increase is 4.27% and the decrease is 5.93%. Some of these changes to working hours could be understood by the effects of public schools being closed during the summer months, but the overall trends start before the normal end of the school year. What we can say is that on average VPN login times moved away from a more stable and expected distribution (more daytime work) and toward evening and very early morning activity. The afternoon logins seem fairly steady which could be understood in terms of both scheduled/inflexible activities and persons not directly affected by changes in family dynamics. Since the data we used was a set of summary statistics, we could not look into the question of how evenly the change in work patterns was distributed across gender, income, etc.

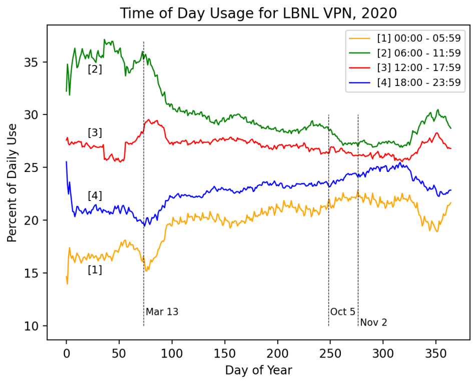


Figure 9: Distribution of LBNL VPN login times broken out into 6 hour blocks using 20 day rolling average. Y values represent the percent of that days traffic in each category.

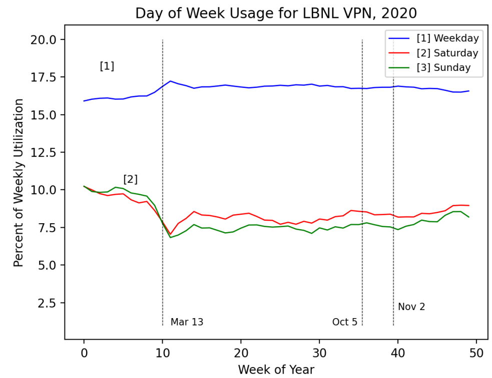


Figure 10: Distribution of LBNL VPN login days using a 3-week rolling average. Y-values represent the percent of that weeks traffic represented by the day. Weekday numbers are averaged together.

4.3 Login Day of the Week

Besides the time of day, looking at which day(s) of the week for staff logins can be telling as well. For Figure 10 we look at the day of the week when logins happen, consolidating weekdays as a single average and weekends as another. The graph is showing the percent of traffic over a 1-week period rather than an absolute count of logins. There are two interesting takeaways from this graph. First, the percentage of weekend logins consistently drops from around 10% on average to around 7.5% between Phase 1 and 2. The second is that the most common weekend day moved from Sunday to Saturday.

Type	Phase1	SD1	Phase2	SD2	Phase3	SD3	Phase4	SD4
Hours 00-06 [1]	16.82	0.85	19.86	1.50	21.81	0.41	12.24	1.02
Hours 06-12 [2]	35.26	1.32	30.11	1.90	27.85	0.63	28.13	1.14
Hours 12-18 [3]	26.79	0.69	27.51	0.70	26.50	0.24	26.53	0.72
Hours 18-00 [4]	21.13	0.93	22.53	1.06	23.84	0.40	24.10	0.95
Weekday [1]	16.06	0.10	16.87	0.19	16.80	0.06	16.72	0.13
Saturday [2]	9.73	0.34	8.10	0.41	8.42	0.21	8.50	0.28
Sunday [3]	9.95	0.18	7.56	0.58	7.60	0.12	7.90	0.38

Table 2: Average values and standard deviation for percent of time spent in each category type for Figure 9 and Figure 10.

4.4 Distribution of IPSec Traffic

While looking at login records can provide a high degree of accuracy in terms of accounts and date/time, accessing those logs can be difficult or impossible outside of internal resources or well established relationships. To get a better idea about the mechanics of VPN traffic on a larger scale, we can look at the behaviors of IPSec traffic across the WAN. Using NetFlow records from our routing infrastructure, we can look at characteristics of IPSec traffic over the first part of the year to see if there are identifiable changes. Since this measurement is retrospective, we use IP type 50 (ESP, encapsulating security payload [4]) as a proxy for traditional tunnels. There are other types of tunnels used for VPNs (for example SSL/TLS based), but they tend to be more difficult to measure from a NetFlow perspective. As well the flow data we have available is sampled 1:1000 so measurements will need to focus on aggregate behaviors and trends.

Because ESnet carries a great deal of routinely encrypted traffic for our various sites, we had to come up with a way to identify the human traffic in that general sea of background data. To do this we chose to look at ESP traffic to and from several different commodity network providers typically used for individual internet access. They will be identified as ISP1 through ISP8.

Measurement Type	Phase 1	StDev 1	Phase 2	StDev 2
Hours 00-06 [1]	11.83	1.46	10.97	3.11
Hours 06-12 [2]	6.50	0.68	13.38	2.28
Hours 12-18 [3]	35.57	3.27	45.05	2.84
Hours 18-00 [4]	46.10	2.91	30.61	3.97

Table 3: Distribution of IPSec traffic broken out into 6-hour blocks using 20-day rolling average from Figure 11. Y-values represent the percent of that days traffic in each category.

Like the measurements in LBNL VPN login data, the time of day observations are broken up into the proportion of traffic per day found in 6-hour bands. The observed changes represent what you would expect from a rapid increase in the number of remote staff - a steady increase in the proportion of traffic between 12 PM-6 PM and a mirrored decrease between 6 AM-12 PM. Without more detail in terms of IP source and destination, packet sizes, and data transferred it would be difficult to make a stronger correlation argument. In terms of this survey we did not look into the traffic details, but that might make for interesting future work. Not all the observed traffic displayed the degree of change seen in Figure 10, but there was an observable change in four out of the eight.

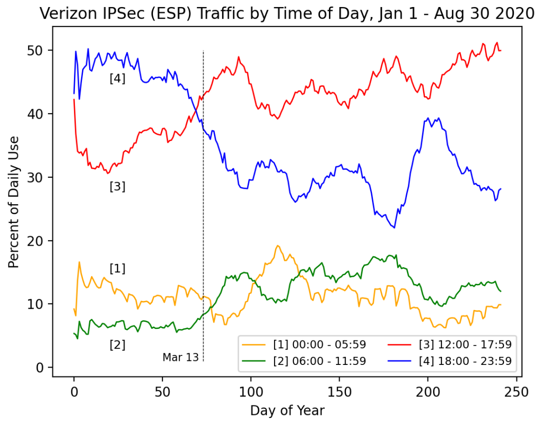


Figure 11: Verizon Unique Connections.

5 RELATED WORK

Digital technologies like sensors, data monitoring, and network changes have witnessed a surge in need during the COVID era [13]. Changes in mobility networks and their impact on the cellular networks were also analyzed in the areas of London and geo-temporal granularity of UK areas showing a significant drop in data volume during lockdown [9]. Other internet studies show a 15-20% peak on remote work but do not increase more than 5% compared to normal working hours [5].

Research and experimental networks have also stepped up their game, with Geant and ESnet studying the number of virtual peering points and showing the network could well adapt to the high demand. Overall, while economies reset, we do see a major shift in the working patterns of users and which may have an impact in the long run on mental health and work-family life divisions.

6 LESSONS LEARNED

Given the duration of the continuing global pandemic, the observation period presented in this paper is not as well placed to discuss longer term observations and conclusions, but better to document the complicated and chaotic transition period between the emergence of COVID-19 and the few months after when the first set of decisions and adjustments were beginning to take effect. Even with that caveat, there are a number of things that we, as a international network service provider, can point out.

While large scale experiments are important in terms of long term Open Science research, without sufficiently flexible and robust infrastructure (like remote access), there will be critical service problems. Cloud based solutions for collaborative document sharing and telecommuting introduce new data usage patterns that we, as service providers, need to be flexible enough to change with.

Our results are able to show that the working patterns of scientists and staff logging into facilities and performing their experiments has changed. Even with the shift in work-from-home patterns, science productivity has increased especially as a result of the additional COVID analysis taking place on DOE facilities.

7 CONCLUSIONS

The material fallout from safety measures introduced to combat the worldwide spread of COVID-19 in 2020 is significant and will remain a topic of study for years. In this study, we look at a combination of Netflow, SNMP, and VPN access logs from 3 HPC centers (ALCF, OLCF, NERSC) and 2 user sites (PNNL, JLAB), as well as VPN logs from LBNL and show that while HPC related traffic was not significantly affected, real differences in user behavior and application composition could be observed. This is not unexpected and reflects new work patterns necessitated by our changing environments.

As the COVID-19 virus continues to mutate and stress social and public health infrastructure, we can be fairly certain to see continuing work within our various research facilities in support of solutions. For us, this takes the form of both immediate support for the physical infrastructure used for data sharing as well as making sure that infrastructure is flexible enough to address the changing needs of the working staff and scientists. While science productivity seems not to have been greatly impacted, we see significant changes in working hours with a heavier pattern working in the evening and early morning hours and weekends compared to previous patterns. This may warrant further study on the impact on sociological and health to prevent odd working hours and over-exhaustion if continued for longer periods in the future.

ACKNOWLEDGMENTS

This work was supported by the U.S. Department of Energy, Office of Science Early Career Research Program for 'Large-scale Deep Learning for Intelligent Networks' Contract no FP00006145.

REFERENCES

[1] A. Affinito, A. Botta, and G. Ventre. 2020. The impact of covid on network utilization: an analysis on domain popularity. In *2020 IEEE 25th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*. 1–6. https://doi.org/10.1109/CAMAD50429.2020.9209302

[2] Mary Branscombe. 2020. The Network Impact of the Global COVID-19 Pandemic. https://thenewstack.io/the-network-impact-of-the-global-covid-19-pandemic/. (2020). Online; accessed 14 April 2020.

[3] J. Case, M. Fedor, M. Schoffstall, and J. Davin. 1990. A Simple Network Management Protocol (SNMP). *Internet Engineering Task Force [IETF]* (1990).

[4] B. Claise. 2004. Cisco Systems NetFlow Services Export Version 9. *Internet Engineering Task Force [IETF]* (2004).

[5] Anja Feldmann, Oliver Gasser, Franziska Lichtblau, Eric Pujol, Igmar Poese, Christoph Dietzel, Daniel Wagner, Matthias Wichtlhuber, Juan Tapiador, Narseo Vallina-Rodriguez, Oliver Hohlfeld, and Georgios Smaragdakis. 2021. Implications of the COVID-19 Pandemic on the Internet Traffic. In *Broadband Coverage in Germany; 15th ITG-Symposium*. 1–5.

[6] C. P. Guok, D. W. Robertson, E. Chaniotakis, M. R. Thompson, W. Johnston, and B. Tierney. 2008. A User Driven Dynamic Circuit Network Implementation. In *2008 IEEE Globecom Workshops*. 1–5. https://doi.org/10.1109/GLOCOMW.2008.ECP.14

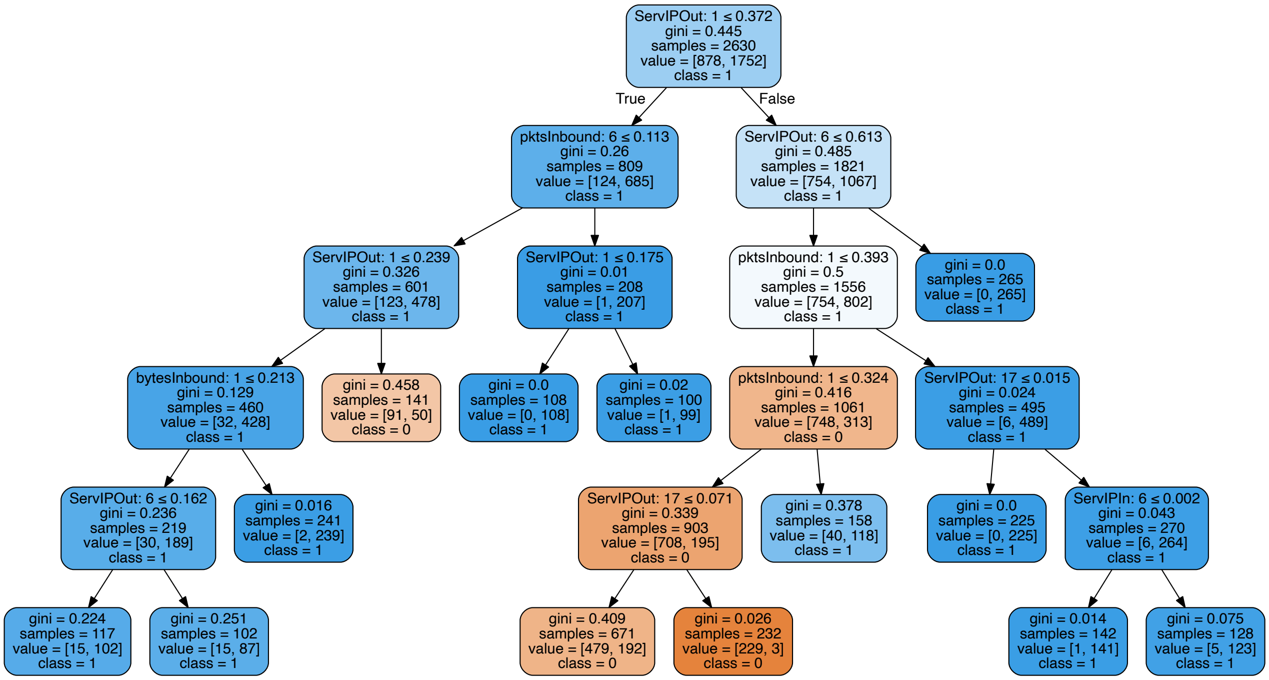
[7] William Johnston, Evangelos Chaniotakis, Eli Dart, Chin Guok, Joe Metzger, and Brian Tierney. [n. d.]. The Evolution of Research and Education Networks and their Essential Role in Modern Science. In *Trends in High Performance and Large Scale Computing, Lucio Grandinetti and Gerhard Joubert editors, to appear.* IOS Press publisher.

[8] S. Kent. 2005. IP Encapsulating Security Payload (ESP). *Internet Engineering Task Force [IETF]* (2005).

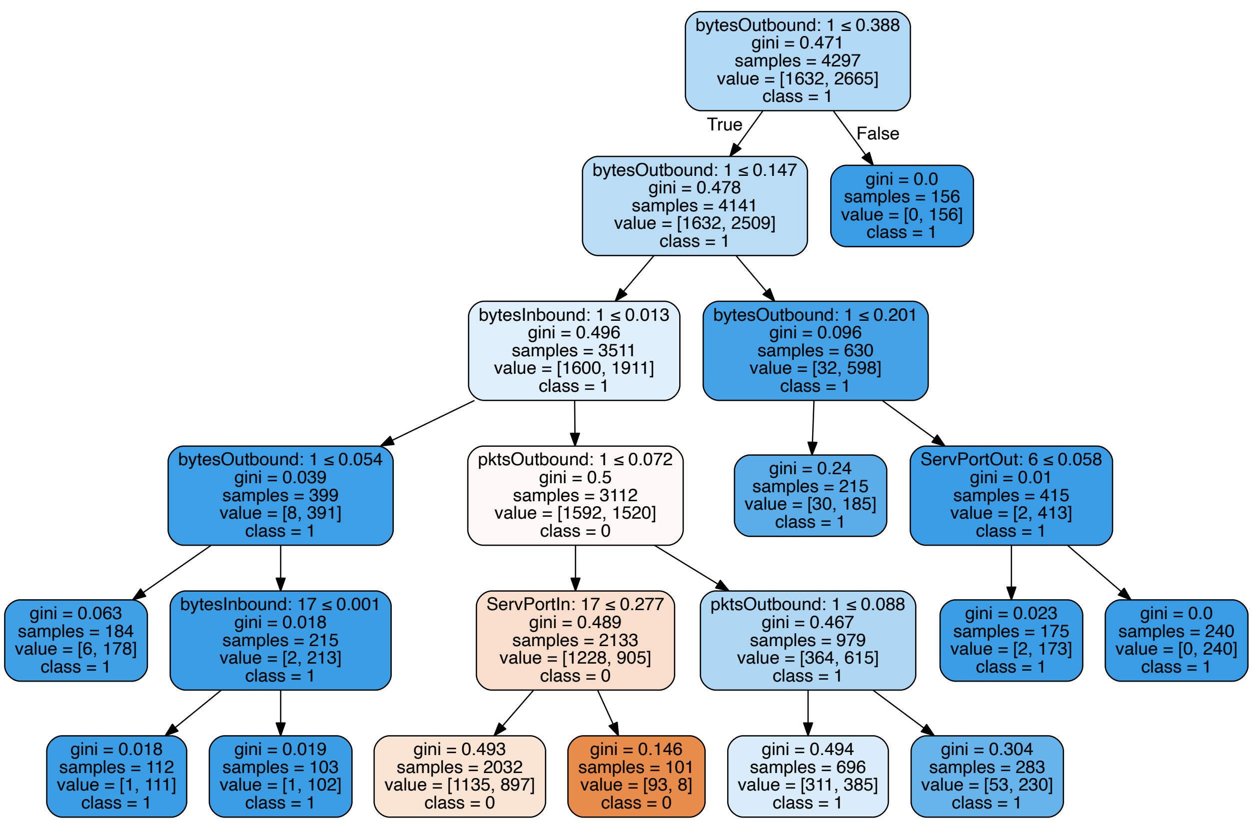
[9] Andra Lutu, Diego Perino, Marcelo Bagnulo, Enrique Frias-Martinez, and Javad Khangosstar. 2020. A Characterization of the COVID-19 Pandemic Impact on a Mobile Network Operator Traffic. In *Proceedings of the ACM Internet Measurement Conference (IMC '20)*. Association for Computing Machinery, New York, NY, USA, 19–33. https://doi.org/10.1145/3419394.3423655

[10] Lucy Meakin. 2021. Remote Working's Longer Hours Are New Normal for Many. https://www.bloomberg.com/news/articles/2021-02-02/remote-working-s-longer-hours-are-new-normal-for-many-chart. (2021). Online; accessed 1 February 2021.

[11] Jeffrey C. Mogul and Lucian Popa. 2012. What We Talk about When We Talk about Cloud Network Performance. *SIGCOMM Comput. Commun. Rev.* 42, 5 (Sept. 2012), 44–48. https://doi.org/10.1145/2378956.2378964
[12] R. Prasad, C. Dovrolis, M. Murray, and K. Claffy. 2003. Bandwidth estimation: metrics, measurement techniques, and tools. *IEEE Network* 17, 6 (2003), 27–35. https://doi.org/10.1109/MNET.2003.1248658
[13] Daniel Shu Wei Ting, Lawrence Carin, Victor Dzau, and Tien Y. Wong. 2020. Digital technology and COVID-19. *Nature Medicine* 26, 4 (2020), 459–461. https://doi.org/10.1038/s41591-020-0824-5



(a) OLCF.



(b) NERSC.

Figure 12: Decision Tree of Pre and PostCovid Traffic across Compute Facilities: OLCF and NERSC.

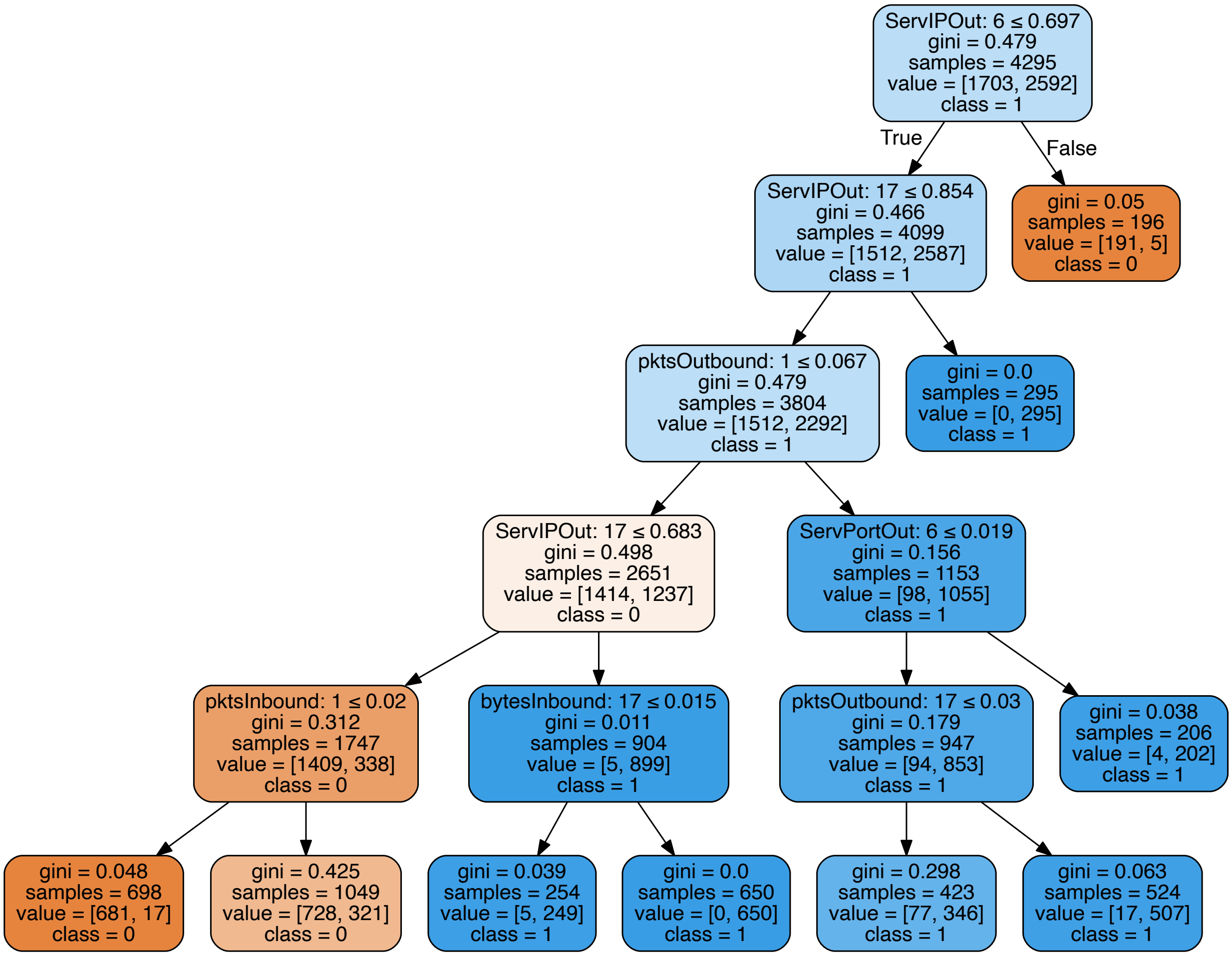
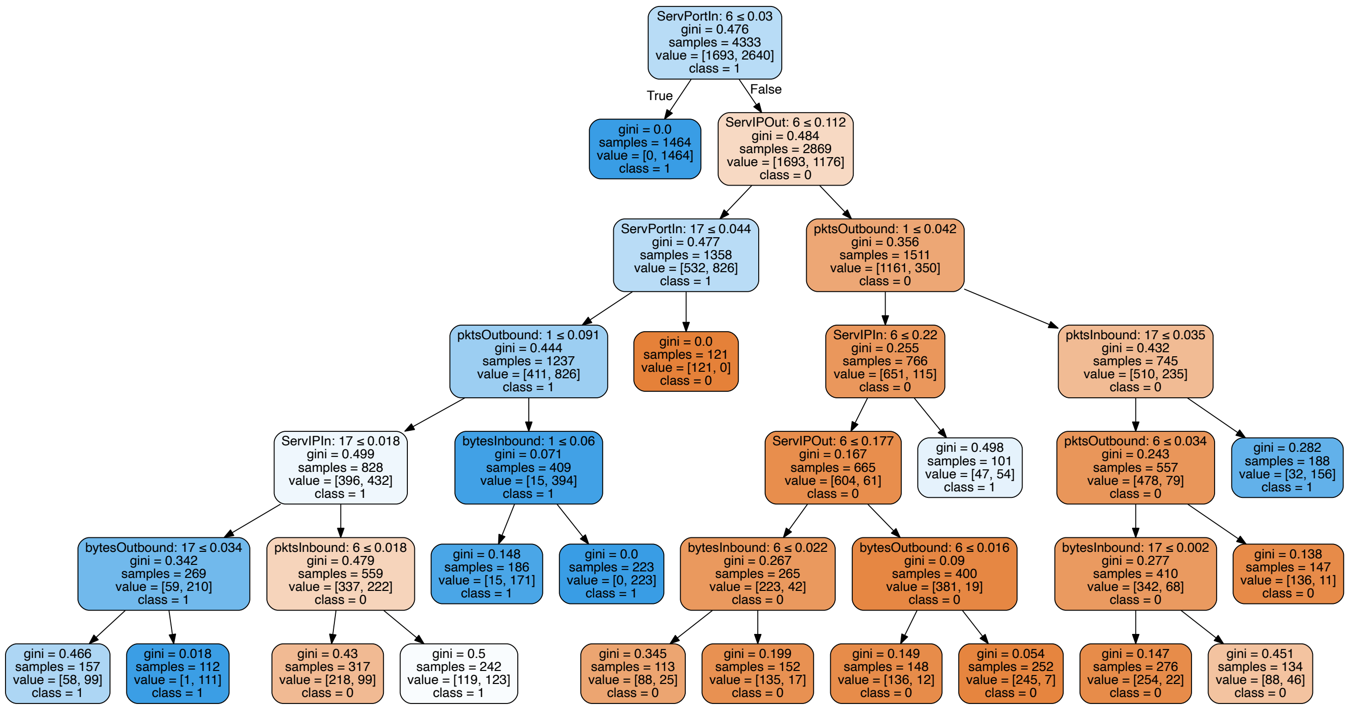
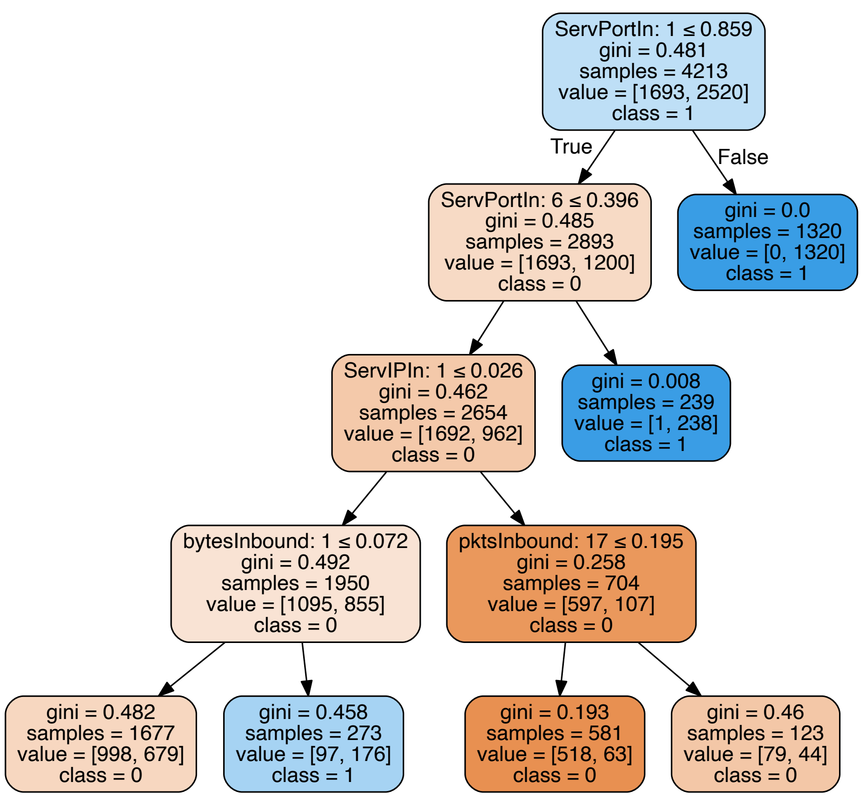


Figure 13: Decision Tree of Pre and PostCovid Traffic across Compute Facilities: ALCF.



(a) JLAB.



(b) PNNL.

Figure 14: Decision Tree of Pre and PostCovid Traffic across User Facilities.