Major Project 2 - XYZ Telecom – Performance Sustainability

Abhijeet Bhardwaj

Summary

It is essential to characterize the performance of the passive optical network in order to have sustainable and cost effective deployment. XYZ Telecom plans to build high performance fiber network connectivity across the country. In order to have sustainable and efficient networks, they would like to learn how the performance of the network varies due to several network parameters and its complexity. In order to better understand the variability in their network performance various questions are answered in this report, which provide:

- Overview of the passive optical network (PON) dataset and its characterization
- Quantification of network performance by several parameters of the passive optical network (PON technology)
- Prescriptive analysis to identify issues in network performance which help to detect onset of faults in the network
- Provide recommendation for improving network performance

Methodology: The dataset provided contains large volume of inspection data collected by XYZ Telecom. The inspection frequency for most of the records is around 30 minutes. The PON network follows a hierarchical structure having having Optical Line Terminal (OLT) at the top of the network and Optical Network Termination (ONT) at the bottom of the network. A stationary customer/subscriber is associated with a ONT unit. Multiple parameters of the PON network could affect its performance. The uptime variable measures the time an ONT unit has been working without any interruption. If there is an interruption in the uptime variable it means the system has been reset. *In this report we assume that resetting of an ONT unit corresponds to its failure.* To study the effects of network parameters over ONT failures we use stratified parametric (weibull) survival model. This model helps us to identify the hazards (risk of failure) of a ONT unit with different PON parameters adjusting for idiosyncrasies induced by each ONT unit. Prescriptive methods suggesting faulty network operations are also provided by using Residual Control Charts. The control charts help to club network performance parameters in a efficient manner and make use of a hierarchical linear mixed effects model, which is motivated by the design of PON network.

Results: Sanity check on data helps to identify issues in the data provided. The parameters that were relevant to the study includes 'Product Code', 'Model', 'Current Software version', and PON Port parameters ('Shelf', 'Slot', 'Port'). The parameter 'clei' have a one-to-one correspondence with 'Model' and is thus excluded, while the parameters like 'Profile' seem to have similar but erroneous entry as 'Model'. The different stratum in survival study is made using Port parameters ('Shelf', 'Slot', 'Port') and a significant difference in survival probability (system failures) were observed. 'Model' and 'Current Software version' also have statistically significant effect over failures of ONT while 'Product Code' have insignificant effects

in ONT failures. Power variables (including 'Received level at OLT' and 'Optical Signal Level') measures the power loss in fiber for upstream and downstream signals. These variables are combined to generate a degradation signal by observing trends in them. Control charts provided for the generated degradation signal helps to identify faulty operation of ONT as shown in the report.

1 Introduction

XYZ Telecom aims at providing highly reliable fiber connections to its subscribers. The goal of the current study is to help XYZ Telecom identify some key factors that effect the network parameters. An initial thought is that the network quality degrades due to complacency of the network, the load on the network and engineering material of the fiber optics. The data for a single vendor (manufacturer) is provided. The PON networks consists of an OLT which have multiple shelfs in it. Each shelf could have multiple slot which could have multiple ports. The columns indicating information about shelf, slot and port in the raw data contains error and thus this information is extracted from the pon_port column (shelf#/slot#/port#). The bottom layer of the PON network consists of ONT. The data provides unique ONT and customer ids ('subscriber id'). Ethernet addresses (denoted by onu_mac / mta_mac) for different customers are also present. The column of serial numbers should essentially have a one-to-one correspondence with ONT id, however, it contains errors and is thus not considered for analysis in this report. Several reasons for violation of one-to-one relation in 'subscriber id', 'ont id' and 'onu_mac' are identified and reported in Figure 9. Most of the reasons are justifiable from business perspective however many a times a subscriber is allotted a shared mac address initially and then transferred to another one (for example subscribers 11148351, 11881466). Also we notice a change of pon-port for subscriber 14921185.

The values provided in network 'profile' have same entries as 'model' of ONT with slight misalignment. Profile seems to be untrustworthy as it does not have a one-to-one correspondence with 'clei' while 'model' has that. After discussions with client the column 'alternate software versions' seems to convey minimal information for the present study as failure in software upgrade seems to effect network performance. The voip_config_file_versisons seems to have high number of Null values and hence is droped from analysis. It is thus safely assumed that the network performance may vary by parameters like 'Product code', 'model', 'current software version'. The performance of the hierarchical PON network will also depend on the individual shelf, slot and port ofg the OLT unit from much multiple ONT units are connected. The Design of the PON network is illustrated in Figure 1. We demonstrate the variation in uptime for a random subscriber in Figure 2. It can be seen that during the operation of the ONT the uptime variable is reset multiple times suggesting that the ONT device has been reset. We assume that this resetting of ONT device indicates system failure. This infromation is further used to develop stratified parametric survival models where different stratums are provided by different grouping variables of Figure 1 (shelf, slot and port). We also include the random effects of each ont unit while generating the survival model. For developing a health indicator metric we focus on the observable performance parameters which include transmitted and received power levels of PON network. Bit error rates for upstream and downstream propagation are also provided along with 'transmit optical power levels' and 'response time' variable. These variables however, have minimal variability and thus are excluded from this analysis. As indicated by the client, range length is a derived variable

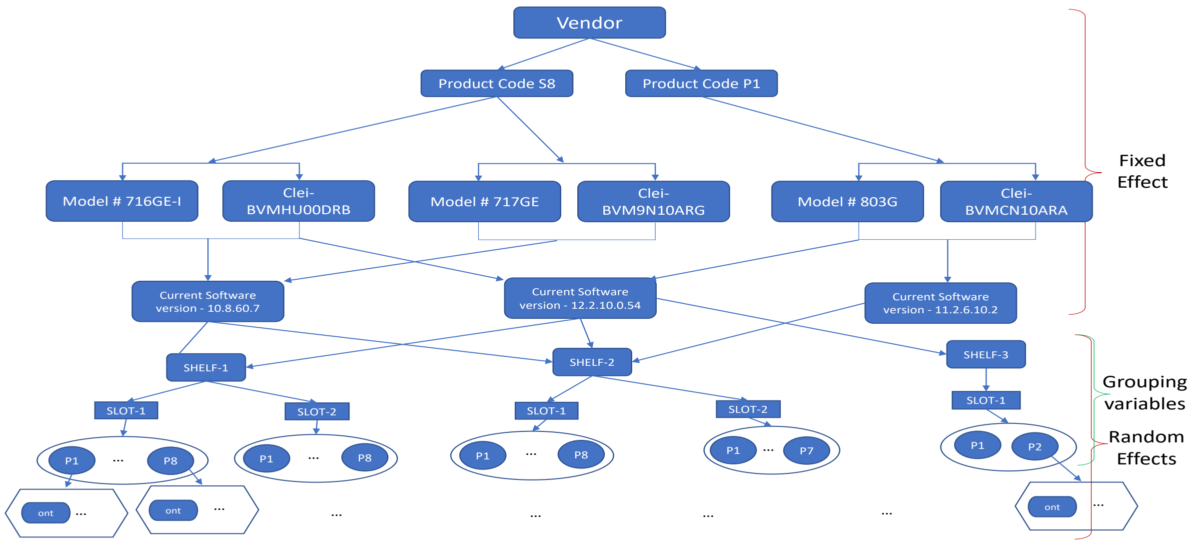

Figure 1: PON network design

and thus can be safely excluded. Promising variability is observed in 'Receive level at OLT' and 'Optical Signal Level' as shown in Figure 3. This prompt us to focus on these parameters for developing control charts for prescriptive analysis.

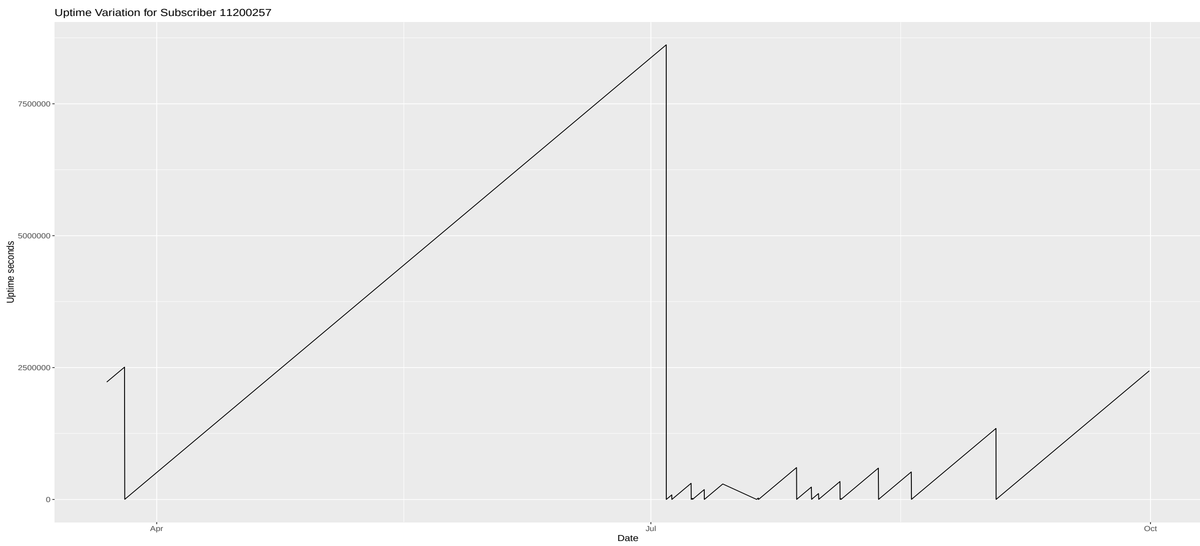

Figure 2: Uptime variation for Subscriber 11200257

2 Survival Analysis- Quantifying network reliability by parameters

ONT units are inspected regularly at an interval of 30 minutes. As can be seen in Figure 2 the uptime variable measures the time for which the unit has been up and running without any interruption. A reset of ONT unit could be estimated by calculating the difference in uptime variable and its lagged (shifted value) Whenever this difference is negative it could be safely assumed that the ONT unit has been reset (Please note that

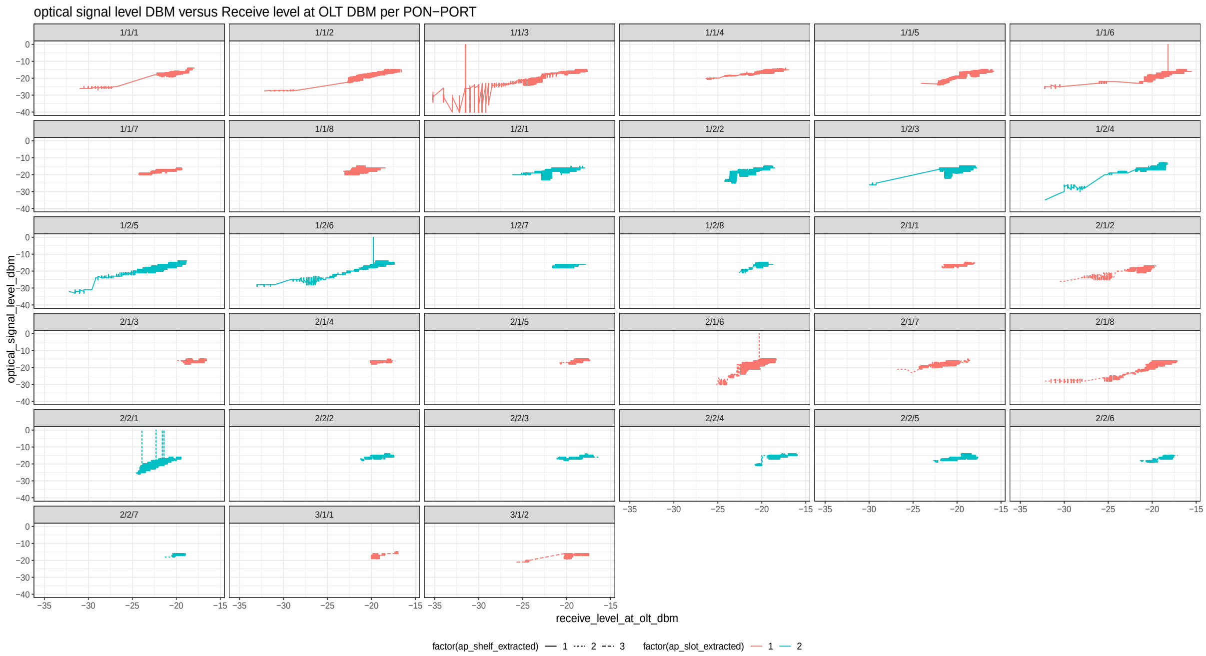

Figure 3: Variation in Optical signal level and received signal level

there were erroneous entries in up-time values for some subscriber like subscriber id 11128858) Such values were removed from the analysis as they occurred due to system error. Figure 4 provides a rough overview of how different number of failures are related to different model number and pon-port. Although uptime variable provides us the number of instances at which ONT units have been reset, it does not provide us the exact time at whicht the ONT was reset. This is because the inspection is made at a fixed interval of 30 minutes and the uptime values just after the reset always lie between (0-30 minutes). Thus we have a interval where the ONT unit is reset. This leads to the problem of interval censoring in survival data.

In order to simplify the analysis we assume that the reset of ONT units are independent of each other (i.e., the reset at one particular time should not affect the reset ot some other time). This assumption is reasonable as for the digital devices like fiber optics material degradation with age is minimal and the failures would depend upon external environmental conditions like rain/wind. Thus for a given subscriber we can assume that multiple independent failure events occur which have the failure times to be interval censored. The last data point however is right censored. This is represented in Figure 5 for a sample subscriber 11200257.

In order to model this relationship we choose parametric weibull distribution as our failure time distribution and regress it with the parameters of interest ('Product code', 'model', 'current software version'). As each pon-port component is assumed to have its own impact over the network performance we stratified the model with 'shelf', 'slot' and 'port'. The ont units in each pon-port could have their own idiosyncrasies and thus to achieve better estimates of variance we also include the frailty by clustering on each ont unit. The syntax of the model as run in R is shown in Appendix Figure 10. The statistically significant factors are shown in Table 1. It can be seen that the parameter 'model' and 'current software version' have statistically significant difference in the survival probabilities of their baseline levels. The effect of stratification is that we achieve a seperate intercept and scale parameter for each stratum (which is formed by

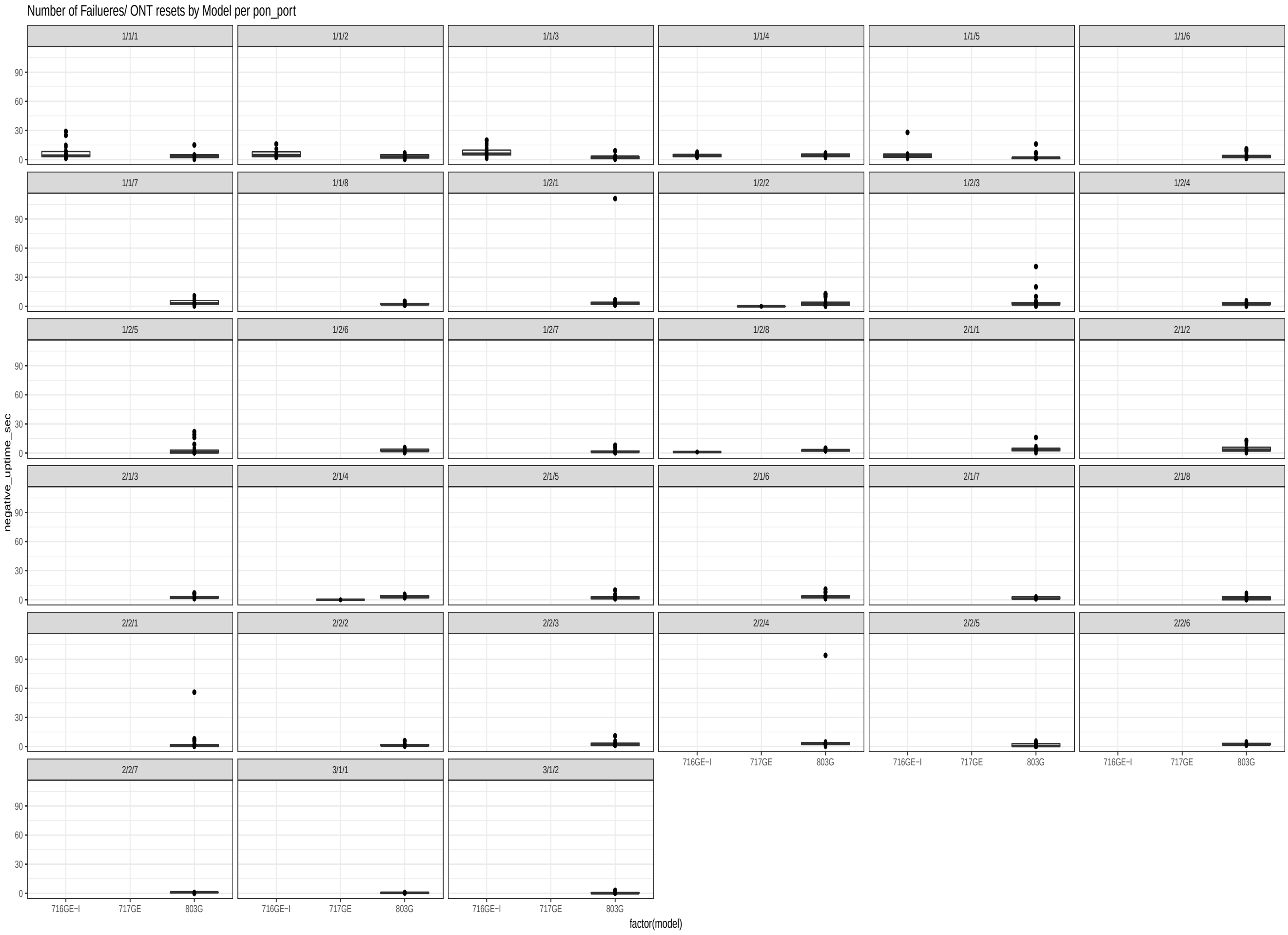

Figure 4: Number of Failures by model per pon-port

the combination of shelf, slot, port). In order to evaluate the hazard from the predicted coefficients we us the formula $HZ_{i,j} = \exp(-\frac{\beta_i}{scale_j})$, where β_i is the coefficient of the i^{th} variable and $scale_j$ is the scale for the j^{th} stratum (pon-port). We evaluate and provide the base line hazard for each stratum (pon-port) in Appendix in Table 4. As can be seen in Table 4 the pon-port 1,2,5 have a significantly large baseline hazard as compared to other pon-ports. For calculating the hazards of each stratum, we use the equation $HZ_{0j} = \exp(-\frac{\beta_0+\beta_j}{scale_j})$. Larger the hazard would mean larger the probability of failures. However, only the baseline hazards for each stratum does not necessarily provide the complete picture. To show this consider Table 2, which gives us the hazard of each parameter for three categories of stratum (stratum with highest, moderate and lowest hazard). Using Table 2 we show in Figure 6 that by keeping the shelf at 1, slot at 2 and port at 7 (green curve) we can guarantee a high long term survival probability for ONT with product code P1, model number 716GE-I and current software version of 12.2.10.0.54. For smaller time values the shelf 2, slot 2 and port 7 with product code S8, model number 716GE-I and software version 12.2.10.0.54 gives best survival probability.

Table 1: Parameter significance for the model

Factors	**Value**	**Std .Error**	**z**	**P-value**
(Intercept)	10.61022989	0.468016776	22.670619	8.74E-114 ***
product_code (S8 versus P1)	0.101939994	0.471618666	0.2161492	0.828871456
model_number (717GE versus 716GE-I)	26.64884669	$\sim$0		0 *****
model_number (803G versus 716GE-I)	-0.130624681	$\sim$0		0 *****
current_sw_version 11.2.6.10.2 versus 10.8.60.7	-2.487779756	0.526775484	-4.722657	2.33E-06 ***
current_sw_version 12.2.10.0.54 versus 10.8.60.7	0.24028112	0.459412477	0.5230183	0.600961537

Table 2: Hazard rates of each covariate by different stratum of high/medium and low hazard

Parameters	**High.Hazard**	**Medium.Hazard**	**Low.Hazard**
product_code S8 versus P1	9.65E-01	9.22E-01	6.21E-01
model_number 717GE versus 716GE-I	9.24E-05	6.87E-10	8.97E-55
model_number 803G versus 716GE-I	1.05E+00	1.11E+00	1.84E+00
current_sw_version 11.2.6.10.2 versus 10.8.60.7	2.38E+00	7.17E+00	1.11E+05
current_sw_version 12.2.10.0.54 versus 10.8.60.7	9.20E-01	8.27E-01	3.26E-01
pon_port	1, 2, 5	1, 2, 7	2, 2, 7

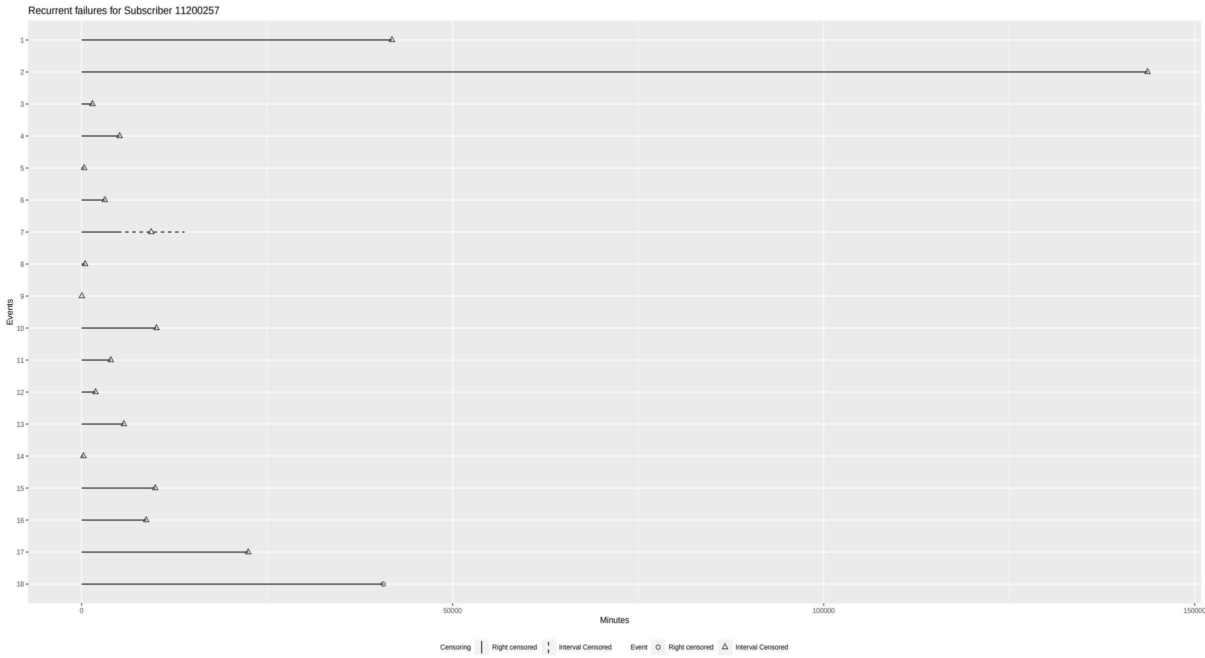

Figure 5: Multiple Interval censored ONT Failures for subscriber 11200257

3 Prescriptive Analysis and Residual Control Chart

As discussed power loss occurred in the fiber can be a useful source of to monitor network health at any given time. As shown in Figure 3 a linear relation could be observed between received power at olt and optical signal level measured at onu. As guided by domain knowledge the received power should lie with the limits of -8dB, -27dB to ensure optimal flow of information. As optical signal levels and received power levels are suffer power losses after passing through the fiber optics in upstream and downstream flow a linear relation between them should tell us the overall health of the network. Keeping this in mind we create a surrogate variable called power difference which we calculate by taking the difference between power received at OLT and optical signal levels measured at ONTs. Figure 7 helps us to visualize the different power variables for subscriber 11200257. The red dotted lines indicates ONT restarts for this subscriber as shown in Figure 2. We can see that the surrogate variable of power difference shown in the bottom right part of Figure 7 shows promising results in terms of identifying ONT resets/failure.

In order to develop a method for monitoring the power level we develop a residual control chart for the power difference variable. Essentially we assume that for the points which are in-control (normal operation) the power difference variable would have a linear dependence on the different fixed parameters shown in Figure 9. The pon-port parameters (shelf, slot and port) are shown to behave differently in the survival analysis and thus it becomes essential that we allow the linear model to have random intercepts for each pon-port. Also as onts are hierarchically cross classified in the pon-ports, it becomes essential that we specify the correct hierarchical relation. We thus use a linear mixed effect model which is specified by equation 1. Here δP refers to the surrogate variable that we have created.

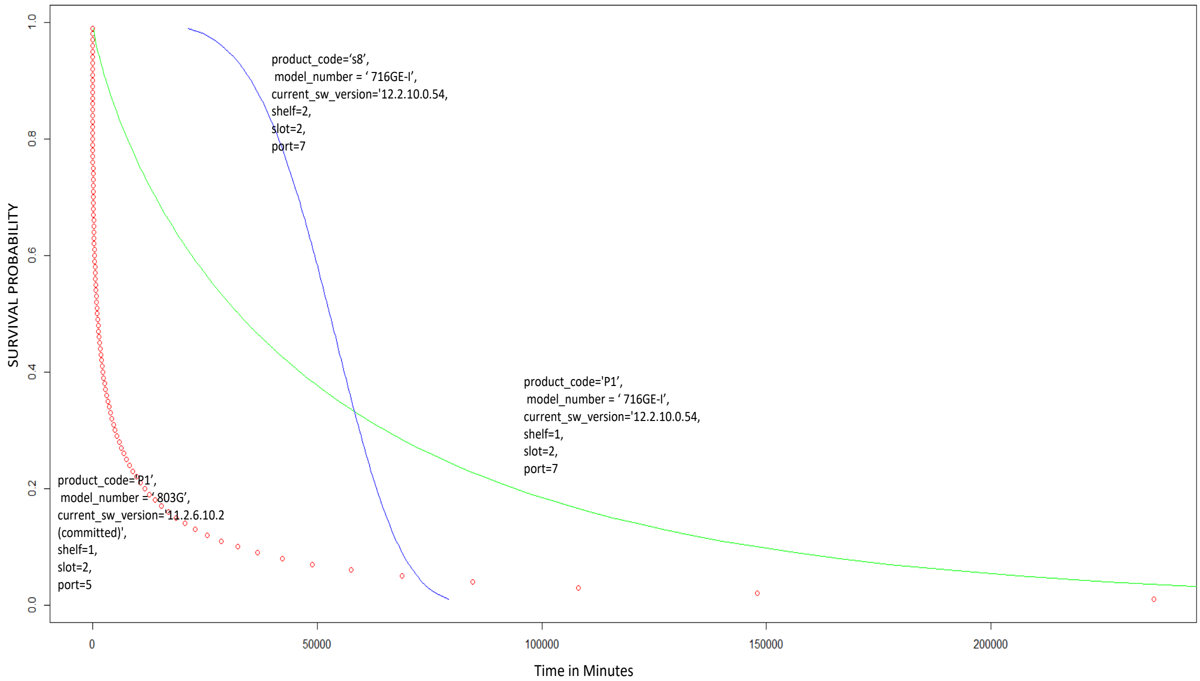

Figure 6: Survival curves for different combinations of shelf, slot, port and covariates.

$$\begin{aligned} &Model\ 1: \\ \delta P_{i,pon,ont} = &\ \beta_0 + \beta_1 ProductCode_i + \beta_2 Model_i + \beta_3 CurrentSoftware_i + \\ &P_{0pon} + P_{0pon} : O_{0ont} + +\epsilon_{i,pon,ont} \end{aligned} \tag{1}$$

To estimate the parameters of the model in equation 1, we subset the data for which received level at olt lies between -8dB to -27dB. As we have huge amount of data convergence of model takes very long. Thus we randomly sample datapoints from this subset to get a smaller training set for the linear model. By using the syntax as shown in Figure 11 we get the estimates of model parameters for the in-control data. For this particular case we find that all fixed effects are statistically significant except for model which is marginally significant (but this could be due to smaller data sample that we have generated.) Please note that the inference of effects here are not of much significance to network performance as the main goal that we have here is to demonstrate a methodology that can help us signal bad network health. We also find that the variation induced by random intercepts of pon-ports and ont units is statistically significant, thus affirming the model formulation.

After we have the in-control model developed we predict residuals for the in-control dataset. The residuals are the difference between the predicted power difference and the true power difference. We now generate the upper ontrol limit (UCL) and lower control limits (LCL) for these residuals by a simple 3σ method. The UCL $= Residual\bar{I}ncontrol+ 3\Sigma_{ResidualIncontrol}$ while the LCL$=$ $Residual\bar{I}ncontrol - 3\Sigma_{ResidualIncontrol}$. These UCL and LCL values becomes the mean for us to classify future data points to lie in control or out of control. Using the model fit, we now predict difference in power for all the

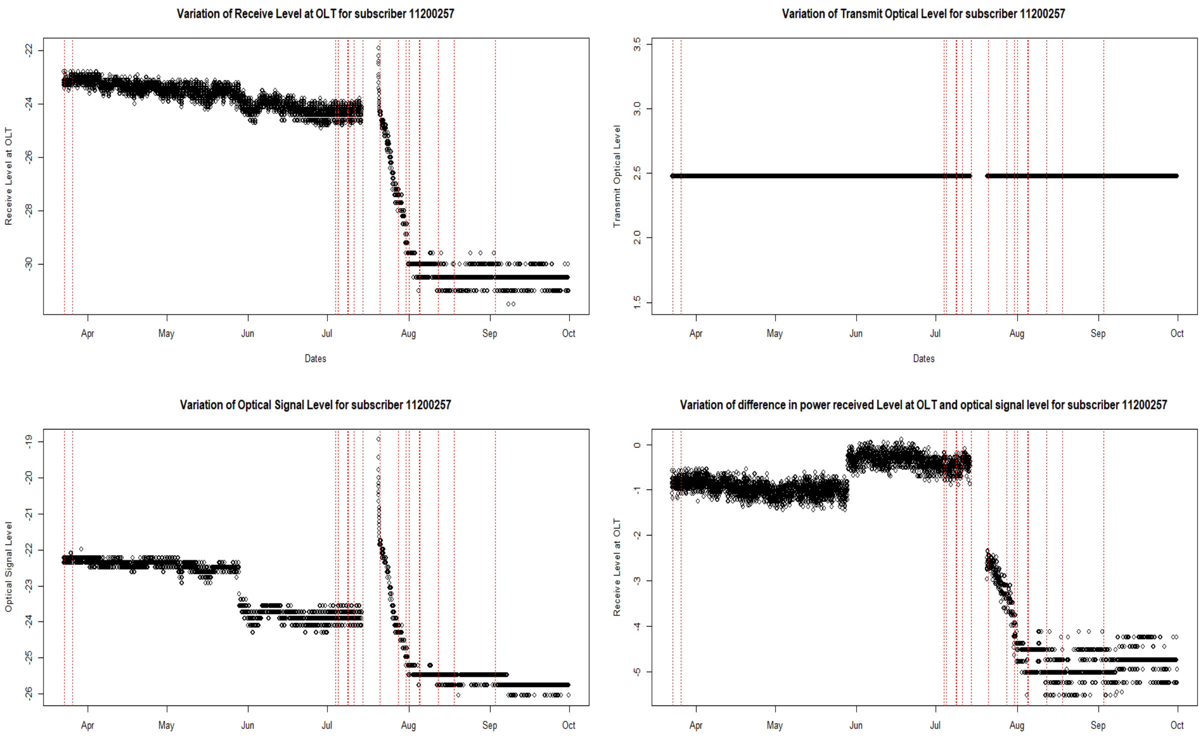

Figure 7: Variation in attenuation of measured power variables for subscriber 11200257

data points $(\hat{\delta P})$. We also calculate the true power difference between received power levels at olt and optical signal level at ont (δP). The difference between the two gives us the residuals which help us identify the incontrol and out of control points. Thus $Residual = \delta P - hat\delta P$. Now Residuals which lie above the UCL or lie below the LCL are potential out-of-control points and must be flagged to indicate network performance issue. We carry out the analysis for subscriber 11200257 in order to identify the points which lie out of the control limits of the residual chart in Figure 8. We can see that the residual chart successfully identifies most of the points which lie outside the control limit and indicate ONT restarts. The results look promising in order to construct a health index for PON network using the power variables. The engineering and science behind the linear relation in received power levels and optical power levels will provide us better insights for reducing the false alarm rate of the proposed control chart.

Table 3: Fixed effect estimates for Model in equation 1

Variables	**Estimate**	**Std. Error**	**df**	**t value**	**Pr(>—t—)**
(Intercept)	-3.96E+00	1.73E-01	7.61E+01	-22.87	<2e-16 ***
product_code S8 versus P1	6.88E-01	1.02E-01	6.10E+05	6.732	1.68E-11 ***
model_number 717GE versus 716GE-I	9.85E-01	5.28E-01	5.51E+02	1.865	0.06267 .
current_sw_version 11.2.6.10.2 versus 10.8.60.7	1.14E+00	3.23E-01	2.12E+04	3.522	0.00043 ***
current_sw_version 12.2.10.0.54 versus 10.8.60.7	1.23E+00	1.02E-01	6.10E+05	12.035	<2e-16 ***

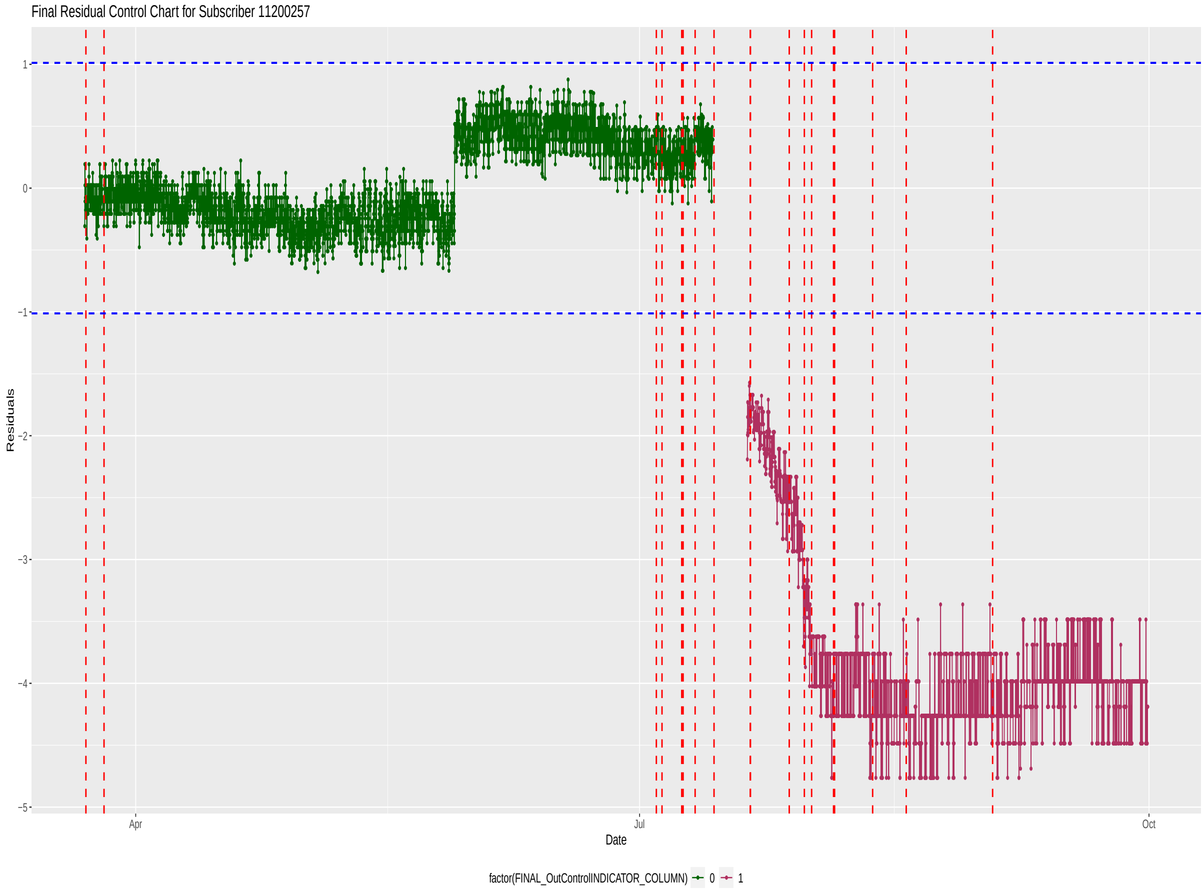

Figure 8: Final control charts with UCL and LCL (blue dots) for subscriber 11200257

4 Discussion

In this report we try to identify the parameters that affect the network performance of the PON network. We made use of survival analysis to identify how hazard of the failing varies with different pon-ports for different covariates. In the discussion we would also like to provide our thoughts on the important question, about how does the network performance change when the network densify. The top 5 dense pon-ports are 1/2/3, 2/2/1, 1/2/1, 2/1/6, 1/1/1. The base lines hazards for these pon-ports as indicated by Table 4 is 0.0037, 0.000086, 0.003, 0.0005, 0.0005. While the highest baseline hazard is for the pon-port 1/2/5 and 2/2/4 which have modest network density of 21 and 14 subscriber. Thus by this analysis we could get a sense that poor performance of a pon-port may be caused by various factors associated with the port like manufacturing defects of shelf, slot, port; model of ONT; software versions etc.,. Thus it should be taken into consideration that causality of poor performance of network has to be thoroughly inspected for defects in manufacturing and incompatible software types. ANOVA analysis could also be considered to identify the difference in defects between different pon-ports. For real-time identification of problems in PON network physics behind the linear relation of received power at olt and optical signal level at ont should be considered to generate a similar looking control chart that can help falg issues in network in real-time.

5 Appendix

Table 4: Base hazard for each stratum of shelf, slot, port (pon-port)

Factors	**Value**	**P-value**	**Significance**	**Weibull Statum Scale**	**Base Hazard each stratum**
(Intercept)	10.61022989	8.74E-114	***		
1, 1, 1	0.370694	3.34E-08	***	1.4487396	5.11E-04
1, 1, 2	0.427411	2.72E-04	***	1.5332820	7.48E-04
1, 1, 3	0.363167	6.45E-06	***	1.4378765	4.85E-04
1, 1, 4	0.121017	2.69E-01		1.1286439	7.42E-05
1, 1, 5	0.426405	1.05E-08	***	1.5317417	7.43E-04
1, 1, 6	0.060447	6.25E-01		1.0623115	4.34E-05
1, 1, 7	0.186236	1.80E-03	***	1.2047071	1.28E-04
1, 1, 8	0.250628	1.18E-03	***	1.2848316	2.13E-04
1, 2, 1	0.675009	5.92E-04	***	1.9640514	3.20E-03
1, 2, 2	0.039740	6.78E-01		1.0405402	3.59E-05
1, 2, 3	0.705315	3.16E-04	***	2.0244850	3.74E-03
1, 2, 4	-0.130029	2.07E-01		0.8780697	6.55E-06
1, 2, 5	1.053870	4.95E-14	***	2.8687305	1.71E-02
1, 2, 6	0.030990	8.52E-01		1.0314748	3.31E-05
1, 2, 7	0.233559	1.10E-01		1.2630873	1.87E-04
1, 2, 8	-0.238323	1.02E-01		0.7879483	1.92E-06
2, 1, 1	0.375443	5.73E-02	.	1.4556360	5.28E-04
2, 1, 2	0.124343	2.97E-01		1.1324039	7.64E-05
2, 1, 3	0.375722	7.22E-02		1.4560423	5.29E-04
2, 1, 4	0.047427	7.99E-01		1.0485696	3.85E-05
2, 1, 5	0.052528	7.15E-01		1.0539324	4.04E-05
2, 1, 6	0.372496	1.44E-02	**	1.4513524	5.17E-04
2, 1, 7	0.130861	2.21E-01		1.1398091	8.08E-05
2, 1, 8	0.688302	6.07E-08	***	1.9903339	3.42E-03
2, 2, 1	0.138162	6.95E-01		1.1481613	8.60E-05
2, 2, 2	0.164535	2.07E-01		1.1788450	1.07E-04
2, 2, 3	0.313495	2.01E-02	**	1.3681991	3.41E-04
2, 2, 4	0.989131	9.83E-09	***	2.6888959	1.34E-02
2, 2, 5	0.400041	2.98E-03	***	1.4918862	6.24E-04
2, 2, 6	-0.219116	2.66E-01		0.8032283	2.41E-06
2, 2, 7	-1.541145	4.69E-08	***	0.2141357	4.04E-19
3, 1, 1	0.517954	2.95E-01		1.6785892	1.32E-03
3, 1, 2	0.835398	8.69E-04	***	2.3057323	6.99E-03

subscriber_id	onu_mac	ont	serial_num	first	last	Comment
11113212	48-77-46-C6-0F-37	1010526	7B0968	3/23/2021 11:11	4/15/2021 8:41	ONT Transferred
11128858	48-77-46-29-ED-EA	1010138	70FB3A	3/23/2021 11:40	9/30/2021 21:10	
11128858	48-77-46-D6-39-84	1010138	70FB3A	3/23/2021 12:10	3/25/2021 4:10	Went to a new mac address and returned
11129424	48-77-46-C5-AF-27	1010157	711AF4	3/23/2021 11:40	9/10/2021 13:10	
11129424	60-DB-98-45-0F-AE	1010157	A88B22	9/10/2021 23:10	9/30/2021 14:40	Changed to new mac and remained on it Location change
11139923	48-77-46-A4-78-14	1010413	71F031	3/23/2021 4:11	6/21/2021 16:11	ONT Transferred
11148351	48-77-46-CA-95-7F	1010321	7B9A31	3/23/2021 10:10	9/30/2021 15:40	
11148351	D0-76-8F-49-4B-31	1010321	7B9A31	3/25/2021 18:10	3/25/2021 18:10	Changed to new mac and remained on it Location change
11213244	48-77-46-D6-39-84	1010136	7EFC3F	3/23/2021 2:10	9/30/2021 21:10	A customer visited to this address
11226547	48-77-46-DC-D1-B6	1010120	7FC6F1	3/23/2021 16:10	9/30/2021 19:10	A customer visited to this address
11310833	D0-76-8F-20-BA-FA	1010437	87D7FF	3/23/2021 15:10	6/23/2021 23:41	ONT Transferred
11446113	48-77-46-C5-BB-07	1010121	7AFEE2	3/23/2021 16:10	9/30/2021 5:10	
11446113	48-77-46-DC-D1-B6	1010121	7AFEE2	3/26/2021 4:40	3/26/2021 4:40	Went to a new mac address and returned
11881466	D0-76-8F-49-4B-31	1010319	8DBF1B	3/23/2021 11:10	9/30/2021 23:41	A customer visited to this address
12129206	D0-76-8F-49-50-17	1010809	8DC0BD	3/23/2021 16:11	9/30/2021 22:13	A customer visited to this address
12131465	D0-76-8F-A5-A8-9F	1010758	96073F	3/24/2021 22:11	3/26/2021 4:41	Went to a new mac address and returned
12131465	D0-76-8F-AD-52-1D	1010758	96073F	3/23/2021 15:11	9/30/2021 23:12	
12201045	D0-76-8F-29-EB-09	1010652	890ADE	3/23/2021 16:11	9/30/2021 15:41	
12201045	D0-76-8F-71-D8-D8	1010652	890ADE	3/23/2021 11:11	3/26/2021 4:41	Went to a new mac address and returned
12215465	D0-76-8F-71-CB-9A	1010815	94DEEF	3/23/2021 5:41	3/26/2021 5:42	Went to a new mac address and returned
12215465	D0-76-8F-A5-B1-93	1010815	94DEEF	3/23/2021 10:11	9/30/2021 15:42	
12215525	D0-76-8F-49-50-17	1010810	94DEF5	3/23/2021 2:11	3/26/2021 4:42	Went to a new mac address and returned
12215525	D0-76-8F-A5-B1-A5	1010810	94DEF5	3/23/2021 10:11	9/30/2021 19:42	
12218285	D0-76-8F-AD-56-73	1020233	9608B1	3/23/2021 14:12	3/28/2021 16:43	ONT Transferred
12237911	D0-76-8F-AD-53-D9	1010823	9607D3	3/23/2021 11:11	9/30/2021 23:12	
12253607	60-DB-98-1D-6C-2F	1010804	A1AB62	3/23/2021 11:11	9/30/2021 21:42	
12279845	D0-76-8F-A5-A8-9F	1010755	94DBF3	3/23/2021 16:11	9/30/2021 21:12	
12288916	D0-76-8F-71-CB-9A	1010813	8EA647	3/23/2021 11:11	9/30/2021 15:42	
12295025	D0-76-8F-A5-A3-E6	1020204	94DA60	3/23/2021 2:42	9/30/2021 23:43	
12295025	D0-76-8F-AD-54-2A	1020204	94DA60	3/24/2021 3:42	3/24/2021 3:42	Went to a new mac address and returned
12353805	D0-76-8F-AD-53-0A	1010824	96078E	3/23/2021 2:11	9/30/2021 23:43	
12353805	D0-76-8F-AD-53-D9	1010824	96078E	3/23/2021 21:41	3/25/2021 7:41	Went to a new mac address and returned
12366269	60-DB-98-15-ED-8A	1010805	A0A0AA	3/23/2021 2:11	9/30/2021 15:13	
12366269	60-DB-98-1D-6C-2F	1010805	A0A0AA	3/23/2021 14:41	3/26/2021 5:41	Went to a new mac address and returned
12388366	60-DB-98-1D-68-90	1010852	A1AA2D	3/23/2021 11:41	6/17/2021 13:12	ONT Transferred
12392126	D0-76-8F-AD-54-2A	1020161	9607EE	3/23/2021 3:12	9/30/2021 23:13	
12400285	D0-76-8F-71-D8-D8	1010648	8EAAB1	3/23/2021 15:11	9/30/2021 23:42	
12561766	60-DB-98-15-EE-F2	1020214	A0A122	3/23/2021 11:42	8/11/2021 0:12	ONT Transferred
12876205	60-DB-98-28-3B-D3	1020233	A3FBAF	4/7/2021 22:42	9/30/2021 23:43	
13082545	D0-76-8F-AD-53-F7	1020719	9607DD	5/17/2021 23:12	9/21/2021 9:14	ONT Transferred
13185865	60-DB-98-2B-80-E1	1020544	A484A7	7/22/2021 18:13	7/22/2021 18:13	NEED LESS ENTRY
13185865	60-DB-98-30-68-A1	1020544	A52F64	7/22/2021 19:44	9/30/2021 19:13	
13193865	60-DB-98-28-3C-06	1010526	A3FBC0	5/14/2021 19:41	9/30/2021 22:11	
13215565	60-DB-98-1B-22-8D	1020309	A0FDAC	6/2/2021 23:12	8/31/2021 9:14	ONT Transferred
13447966	48-77-46-E9-08-2A	2010217	80C2A9	6/18/2021 23:15	7/21/2021 8:14	ONT Transferred
13573565	60-DB-98-2B-91-A3	2010802	A48A3D	6/22/2021 8:14	9/30/2021 23:16	
13573565	60-DB-98-2B-91-B2	2010802	A48A42	6/21/2021 23:45	6/22/2021 0:16	Changed to new mac and remained on it Location change
13644405	60-DB-98-30-68-38	2020102	A52F41	6/23/2021 21:17	7/3/2021 1:18	Changed to new mac and remained on it Location change
13644405	60-DB-98-45-33-C9	2020102	A8972B	7/3/2021 19:46	9/30/2021 23:16	
13661605	60-DB-98-2B-93-80	2010834	A48ADC	6/24/2021 19:17	8/5/2021 13:16	ONT Transferred
13879405	D0-76-8F-20-BA-FA	1010437	87D7FF	7/1/2021 20:41	9/30/2021 21:11	
13916327	60-DB-98-44-FC-A9	2020111	A884CB	7/13/2021 23:17	9/30/2021 23:45	
13922506	60-DB-98-1D-68-90	1010852	A1AA2D	6/18/2021 22:43	9/30/2021 22:13	
13949605	48-77-46-A4-78-14	1010413	71F031	6/22/2021 18:41	9/30/2021 23:41	
13961965	60-DB-98-30-67-1B	2020418	A52EE2	7/16/2021 22:46	7/16/2021 22:46	Have to change ONT unit
13961965	60-DB-98-30-67-1B	2020424	A52EE2	7/16/2021 23:48	9/30/2021 1:48	
14446045	60-DB-98-2B-93-80	2010834	A48ADC	8/5/2021 23:45	9/30/2021 15:16	
14471485	60-DB-98-45-2B-29	2010217	A8944B	8/12/2021 21:15	9/30/2021 23:44	
14596845	60-DB-98-30-69-01	2020148	A52F84	8/25/2021 23:45	9/30/2021 15:46	
14621725	60-DB-98-30-69-01	3010227	A52F84	8/25/2021 17:48	8/25/2021 19:46	Changed to new mac and remained on it Location change
14621725	D0-76-8F-A5-B1-A8	3010227	94DEF6	8/25/2021 20:46	9/30/2021 23:48	
14841205	60-DB-98-45-36-24	1020214	A897F4	9/15/2021 23:12	9/30/2021 22:43	
14844805	60-DB-98-83-52-B0	1020719	B2C85E	9/22/2021 22:14	9/30/2021 16:44	
14844805	D0-76-8F-AD-53-F7	1020719	9607DD	9/21/2021 15:44	9/22/2021 13:43	Changed to new mac and remained on it Location change
14888765	60-DB-98-44-FD-4B	1020309	A88501	9/16/2021 16:14	9/30/2021 21:43	
14921185	60-DB-98-83-56-B5	2010150	B2C9B5	9/25/2021 18:14	9/25/2021 20:45	
14921185	60-DB-98-83-56-B5	2010150	B2C9B5	9/25/2021 22:16	9/30/2021 13:14	Customer changed to a new pon port

Figure 9: Violation of one-to-one relationship in 'subscriber id', 'ont' and 'mac address'.

```
model_survreg2 = survreg(Surv(time= INTERVAL_TIME_t1, time2 = INTERVAL_TIME_t2,
                              type = c("interval2")) ~factor(product_code)+ factor(model_number)+
                         factor(current_sw_version)+strata(factor(ap_shelf_extracted))+
                         strata(factor(ap_slot_extracted))+
                         strata(factor(ap_port_extracted))+
                         cluster(ont),
                       data =TDS_survival_data_subset_2_copy,dist="weibull", control = list(maxiter=200))
```

Figure 10: Syntax for Survival Model in R stratified by shelf, slot, port and clustered on ONT

```
> full_model_lmer = lmer(power_difference~ factor(product_code)+ factor(model_number)+
+                          factor(current_sw_version)+(1|pon_port/ont),
+                        data= tds_processed_correct_subset_incontrol_final_sample,
+                        control=lmerControl(optimizer="optimx",optCtrl=list(method="nlminb")), verbose = TRUE)
```

```
Random effects:
 Groups       Name        Variance Std.Dev.
 ont:pon_port (Intercept) 0.5150   0.7176
 pon_port     (Intercept) 0.6103   0.7812
 Residual                 0.1139   0.3374
Number of obs: 610141, groups:  ont:pon_port, 582; pon_port, 33
```

```
ANOVA-like table for random-effects: Single term deletions

Model:
power_difference ~ factor(product_code) + factor(model_number) + factor(current_sw_version) + (1 | ont:pon_port) + (1 | pon_port)
                  npar  logLik     AIC    LRT Df Pr(>Chisq)
<none>               8 -205420  410855
(1 | ont:pon_port)   7 -693820 1387654 976801  1  < 2.2e-16 ***
(1 | pon_port)       7 -205600  411213    360  1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Figure 11: Syntax and output for linear mixed effects model