M3 Challenge 2021

Team 14624

2/28/2021

Contents

1 The Cost of Connectivity

1.1 Problem Summary

In terms of cost per unit of bandwidth, the price for internet has gotten cheaper as a result of better infrastructure and rapid advancements in digital technology. In this section, we hope to predict the cost per unit of bandwidth in dollars or pounds per Mbps(megabits per second) over the next 10 years for consumers in the United States and the United Kingdom. With the rise of 5G in recent years, we hope to project the price per unit bandwidth for U.S. and U.K. consumers.

1.2 Assumptions

1. Data plans are fairly constant for every person across the U.S. (and similar for the U.K.)
   - Given that both countries operate under a free market, market competition should drive prices to be around similar levels (for the most part.)
2. There is an exponential increase in both supply and demand for bandwidth
   - Technological progress for the electronics and the internet has grown exponentially over the past few decade, so it is reasonable to assume that the amount of bandwidth provided by companies will continue to grow exponentially over time.
   - In the same manner, future media will increase in bandwidth requirement as the technology to produce them improves as well, resulting in a similar exponential growth for demand as well.
3. The data used for analysis is the average **peak** download speed in megabits per second.
   - Data provided by Ookla's Global Connectivity Report in recent years only contains data for average peak download speed.
4. The level of growth of bandwidth usage in the cities is comparable to the level of growth of bandwidth usage in the suburbs
   - The data we use to establish a rate of growth relies mostly on cities with use much higher than the national average, however it is reasonable to assume that even though the magnitudes of use vary, the rate of growth remains mostly the same throughout the country due to the same internet providers providing for majority of the country.
5. There exists a base cost for providing bandwidth independent of the amount of bandwidth due to fixed costs such as setting up infrastructure and maintenance. We assume that this cost will remain fairly constant over the next 10 years.
   - Our analysis of the data, which will be covered below, extrapolates that there is an intrinsic non-zero cost even when the amount of bandwidth provided is 0. Although there is some slight variance in the base costs we calculate, we believe that it is reasonable to model it as a constant since we do not expect the cost of the infrastructure itself to change significantly over the next decade.
6. The data on peak bandwidth usage provided by Akamai and Ookla differ by a constant scale factor due to differences in data collection.
   - We can note that there is a sharp difference in the values provided by the two data collectors. Since both providers give different values for the 2017 data, we attempt to reconcile their different methods of data collection by scaling the Akamai data, which precedes 2017, by a constant factor such that it agrees with the more modern Ookla's data which continues past 2017.

7. The demand for bandwidth is approximately the same between America and the UK
   - Due to increasing globalization, it is reasonable to assume that consumers in well-developed first world countries have similar standards for the type of content that they consume. As such, they will require similar amounts of bandwidth to go through their daily lives.

1.3 Variables

We define t as the number of years after 2020.

Symbol	**Definition**	**Units**	**Value**
$S(t)$	Increase in Monthly Price per Additional Unit Bandwidth	Dollars/Mbps	?
B	Base Cost	Dollars	?
$D_{US}(t)$	Average Peak Download Speed(US)	Mbps	?
$D_{UK}(t)$	Average Peak Download Speed(UK)	Mbps	?

1.4 Developing the Model

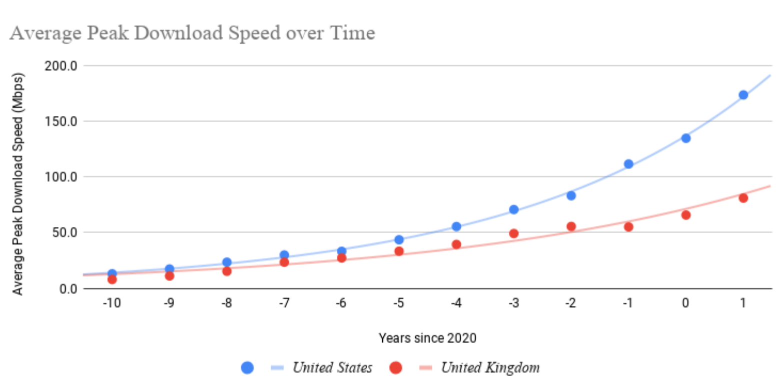
Our model relies primarily on creating values for each of the variables and functions above, as once we do so, we can calculate the price per Mbps in 10 years to be

$$\frac{B}{D(t)} + S(t)$$

We first recall that both supply and demand of bandwidth are assumed to grow exponentially. As such, we get that the price per unit bandwidth should decrease exponentially over time as it becomes easier for companies to distribute it to customers, while the average usage of bandwidth by consumers will also increase exponentially. As such, both $S(t)$ and $D(t)$ can be modeled by exponential functions of the form.

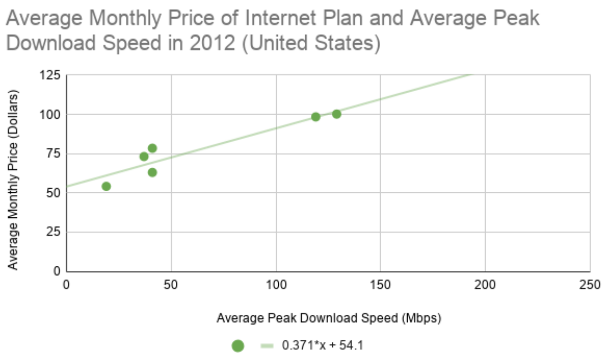
$$S(t) = a_1 r_1^t \qquad D(t) = a_2 r_2^t$$

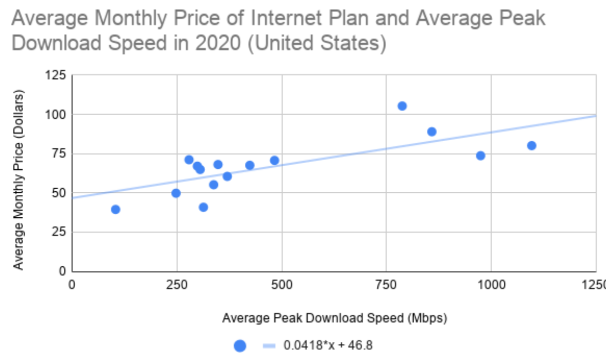
In order to calculate these functions, we look to the historical data on average download speeds in America and the UK[9]. Below, we provide the graph of average peak download speed in America over the last decade. (Note: The data collected by Akamai prior to 2017 has been scaled due to assumption 6). We make use of exponential regression to create a line through the data points.



The exponential regression yields that $D_{US}(t) = 137(1.255)^t$ with an R^2 value of 0.999, and that $D_{UK}(t) = 71.4(1.188)^t$ with an R^2 value of 0.963.

The next step is to create a function for $S(t)$. We will solely use data for America[7][8] for this function, as we have assumed that the demand for bandwidth should be approximately equal between highly developed countries in assumption 7.





We compare the graphs of Average Monthly Price of Internet Plan against Average Peak Download Speed in 2020 and 2012 to gain an idea of the rate of growth. Since we had assumed earlier that S grows exponentially, we can use the slopes of these two graphs to deduce r_1.

The ratios of the slopes, which indicate the price per additional Mbps, is equal to $\frac{0.0418}{0.371} = 0.113$. Given that there is a difference of 8 years between these two sources of data, we can see that $r_1^8 = 0.113$. Solving yields $r_1 = 0.761$.

We then note that a_1 is also equal to the slope of the 2020 graph, as $S(0) = a_1(1)$, so we get the equation

$$S(t) = 0.0418(0.761)^t$$

We can also extract the base cost by looking at the intercepts of both graphs. Note that although the slopes vary drastically, and the mean value of the average peak download speeds are also significantly different between the two years, both graphs have very similar intercepts, lending credence to our assumption that the fixed costs are relatively constant. We will approximate the fixed cost as 50 dollars for the time being.

Now, we have created formulas for all 3 equations and variables. Plugging them in to our formula for the overall cost per Mbps(and using the 2020 conversion for dollars to pounds[6]), we get the following expressions for America and UK respectively:

$$\frac{B}{D_{US}(t)} + S(t) = \frac{50}{137(1.255)^t} + 0.0418(0.761)^t = 0.365(0.797)^t + 0.0418(0.761)^t$$

$$\frac{£1}{\$1.40} \cdot \frac{B}{D_{UK}(t)} + S(t) = \frac{35.7}{71.4(1.188)^t} + 0.0299(0.761)^t = 0.500(0.842)^t + 0.0299(0.761)^t$$

1.5 Predictions and Results

We can now apply these formulas to extrapolate the prices for the next 10 years.

Year	**Price(US, \$)**	**Price(UK, £)**
2021	0.323	0.444
2022	0.256	0.372
2023	0.203	0.312
2024	0.161	0.261
2025	0.128	0.219
2026	0.102	0.184
2027	0.081	0.154
2028	0.064	0.130
2029	0.051	0.109
2030	0.040	0.092

1.6 Strengths and Weaknesses

Our model works well in that it accounts for the changes in both the suppliers capabilities, in Britain and America, as well as the demands of the customers. Furthermore, the exponential model very closely matches the demand curve as evidenced by our R^2 values of 0.999 and 0.963 for America and Britain respectively.

One of the main weaknesses is that we extrapolated the formula for the price per additional Mbps from solely two years of data, and the data in those two years had a fairly significant deviation from the line of best fit that we used to approximate the slope, which could result in significant changes for the actual function $S(t)$.

Another weakness is that the formula for the demand was derived from two sets of data that varied in their method of data collection, resulting in a discrepancy between the data. Though we accounted for this weakness with the analysis in the previous section, it does not completely mitigate this issue and in the future it would be desirable to have a set of continuous data from the same source sourced using the same methods.

2 Bit by Bit

2.1 Problem Summary

Create a flexible model to predict the minimum bandwidth needed to support 90% and 99% of a household's internet needs based on basic information about the household inhabitants.

2.2 Assumptions

1. People do not use internet while asleep
   - This is reasonable since people would be unable to use internet while asleep
2. People under 11 generally get at least 9 hours of sleep per day, people between 12 and 17 generally get at least 8 hours of sleep per day, and anyone older than 17 generally gets at least 7 hours of sleep per day [3]
   - This is reasonable based on the recommended sleep guidelines for the respective age groups. We conservatively chose the lower bounds so that internet usage could only be lower than that predicted by our model.
3. People generally wake up at similar times
   - This is reasonable per household: people in the same household likely wake up at similar times due to performing similar activities at similar times. As such, the main variance in sleep schedules among those in the same household is the sleeping time.
4. Weekly bandwidth usage is a good representation of a person's bandwidth usage over the course of a year
   - This is reasonable since people generally perform similar activities on a weekly basis
5. Average hours per week of media consumption after the COVID pandemic will be in between the average hours per week of media consumption for the first quarter of 2019 and the first quarter of 2020. Specifically, we assumed the future average hours per week of media consumption of any type to be the average of the average hours per week of media consumption in the first quarter of 2019[4] and the first quarter of 2020[4].
   - This is reasonable since after the COVID pandemic, we expect that media consumption will drop as people will perform more non-internet related activities, but we do not expect it to immediately return to the pre-pandemic levels as people will take time to adjust to the changing lifestyle.
6. An undergraduate student generally refers to someone around the age of 21
   - This assumption was based on data collected by Southeastern Louisiana University regarding the number of enrolled students in a given age group.[10] As such we assume that the people mentioned in situation 3 fall in the age group $18-34$.
7. We assumed that a person of the age range $12-17$ has similar internet usage for activities $4-9$(see chart below) as a person of the age range $18-34$, while a person of the age range $2-11$ has negligible internet use of those categories.
   - This assumption is due to data not being available for the first two age groups. We feel that it is reasonable as many teenagers spend similar times on the given activities as the adults in the age range immediately after them due to similar time commitments with jobs vs. school.

- Children of the youngest age category typically consume these media in lower quantities, although not negligible, which makes setting them to negligible a shortcoming of our model due to lack of data.

8. People choose an action each 30 minutes independently of the actions they do before and after, except for the time period in which they are asleep. The probability that they do a certain action is proportional the the average amount of time spent on that activity per week.

    - This assumption oversimplifies the actions of a person, however it is necessary in order to simplify the problem to reasonably computational levels. Furthermore, there is likely some correlation between a person's own actions from time slot to time slot, as well as between different members of the household's activities, but we choose not to account for those to keep the model simpler.
    - By keeping the sleep time separate, we ensure that there is still some semblance to a normal schedule as people typically have similar sleep schedules from day to day.

9. Heavy downloads occur infrequently enough that we do not have to account for them within the bottom 99% of data usage.

    - Most people do not download files for a significant portion of their time, so we will assume that the average person does not spend a significant portion of their time downloading large files.

2.3 Variables

Note: We conservatively took the maximum required bandwidth for each class of activities as to ensure a household's requirement would be met.

Symbol	**Definition**	**Units**	**Value**
G	Age Group	...	$0-5$
A	Activity Number	...	$0-10$
$M(A)$	Megabit Usage for A	Mbps	see charts below
$T(A,G)$	Average time spent by a member of G on A	hours/week	see charts below
$S(G)$	Average time spent member of G sleeping	hours/week	see charts below
$P(A,G)$	Average percent of time awake spent on A by a member of G	...	$T(A,G)/(168\text{-}S(G))$

Table 1: Variables

Number	Activity	Mbps cost
1	Watching Traditional Television	4
2	TV Connected Game Console	3
3	TV Connected Internet Device	8
4	Internet on a Computer (not including video)	1
5	Video on a Computer	4
6	Total App/Web on a Smartphone	1
7	Video Focused App/Web on a Smartphone	4
8	Total App/Web on a Tablet	1
9	Video Focused App/Web on a Tablet	4
0	Awake non-internet	0

Table 2: Activity Numbers and Megabit Usage[5]

Number	Age Group
1	2-11
2	12-17
3	18-34
4	35-49
5	50-64
6	65+

Table 3: Age Groups[4]

T(A,G)	Age Group					
Activity	**2-11**	**12-17**	**18-34**	**35-49**	**50-64**	**65+**
1	12.825	8.185	12.375	24.935	40.425	50.7
2	2.82	4.13	3.68	1.675	0.46	0.16
3	6.635	3.975	6.14	5.91	4.175	2.685
4	0*	3.835*	3.835	4.51	4.715	2.97
5	0*	1.3*	1.3	0.955	0.785	0.365
6	0*	24.31*	24.31	25.62	21.2	15.47
7	0*	2.765*	2.765	1.95	1.18	0.775
8	0*	3.705*	3.705	5.585	5.98	6.92
9	0*	1.015*	1.015	0.995	0.715	0.625
$S(G)$	63	56	49	49	49	49

Table 4: Values for T(A,G)[4] (*Assumed Values)

2.4 Developing the Model

We decided to use a Monte-Carlo simulation to solve for the expected 90th and 99th percentiles of bandwidth usage in a household. We begin by computing the values for $P(A,G)$ for each activity and age group using the formula, where $P(A,G)$ is defined as the average percent of time spent awake by a member of G spend doing activity A. We intend to use $P(A,G)$ as the probability of that member doing action A in an arbitrary interval, and calculate it with the formula:

$$P(A,G) = \frac{T(A,G)}{168 - S(G)}$$

P(A,G)	Age Group					
Activity	**2-11**	**12-17**	**18-34**	**35-49**	**50-64**	**65+**
1	12.21%	7.31%	10.40%	20.95%	33.97%	42.61%
2	2.69%	3.69%	3.09%	1.41%	0.39%	0.13%
3	6.32%	3.55%	5.16%	4.97%	3.51%	2.26%
4	-	3.42%	3.22%	3.79%	3.96%	2.50%
5	-	1.16%	1.09%	0.80%	0.66%	0.31%
6	-	21.76%	20.43%	21.53%	17.82%	13.00%
7	-	2.47%	2.32%	1.64%	0.99%	0.65%
8	-	3.31%	3.11%	4.69%	5.03%	5.82%
9	-	0.91%	0.85%	0.84%	0.60%	0.53%
0	78.78%	52.48%	50.32%	39.38%	33.08%	32.21%

Table 5: Values for P(A,G)

In order to code the simulation, we follow the steps below. The full code is included in the appendix.

1. Split a week up into 338 30-minute slots.
2. The final $2S(G)$ slots for each day are allotted to sleeping, where no bandwidth is used. Note that all people in the same household are assumed to wake at the same time denoted by $t = 0$, though they may sleep at different time.
3. For all other slots, we pick a random activity with probability $P(A,G)$.

4. Repeat this for each member of the household for each time slot, and aggregate the total bandwidth required across all household members for each slot.

5. Finally, calculate the 90th and 99th percentile of the 338 time slots in order to calculate metrics for a given week. Repeat 10,000 times and graph results.

2.5 Predictions and Results

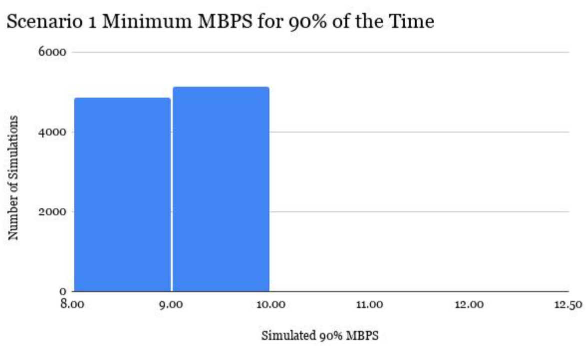
We applied our model to the three scenarios given:

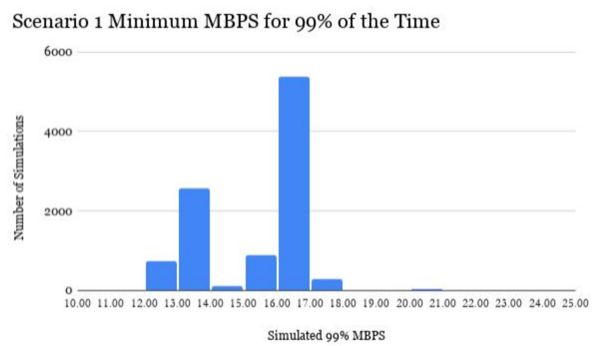
1. A couple in their early 30's (one is looking for work and the other is a teacher) with a 3-year-old child.

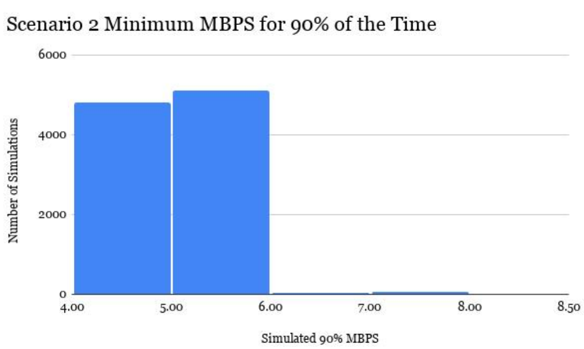
2. A retired woman in her 70's who cares for two school-aged grandchildren twice a week.

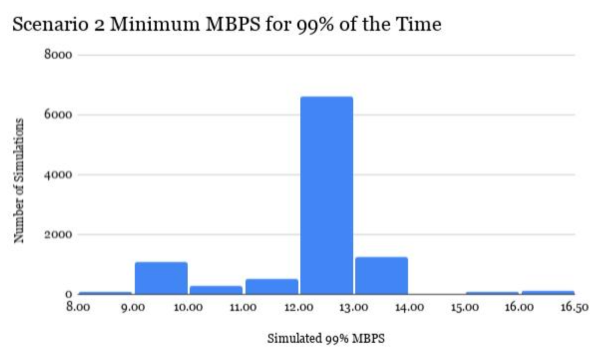
3. Three former M3 Challenge participants sharing an off-campus apartment while they complete their undergraduate degrees full-time and work part-time.

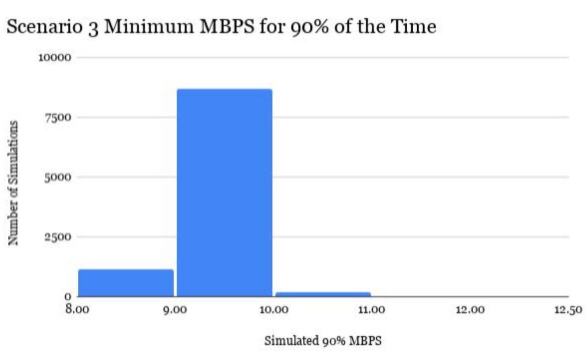
Our model was used to generate the following histograms below.

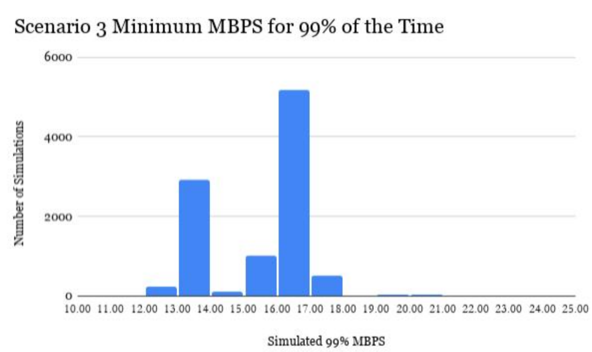












Note that we used a slight modification of the above procedure for the second scenario as two of the household members are only present two days out of the seven. We implement this by returning 0 Mbps for bandwidth usage for all times after 48 hours for both of the children in that scenario.

Based on the histograms that were collected, we would recommend the use of the smallest bandwidth that is an upper bound for at least 98% of the simulations in order to avoid extremely large and unlikely outliers that occurred in the simulations. Specifically, using our model we predict that the following minimum bandwidths will provide adequate coverage 90 and 99 percent of the time for each scenario:

Scenario	90%	99%
1	9 Mbps	17 Mbps
2	5 Mbps	13 Mbps
3	9 Mbps	17 Mbps

2.6 Strengths and Weaknesses

Using a Monte Carlo simulation helps to ensure that our models are robust and able to simulate a wide variety of combinations of internet-consuming activities. Our model also takes into account the sleep schedules of the household members which makes the distribution of internet consumption more realistic as most people do not use the internet while they sleep.

One shortcoming of our model is that it does not account for the occupation or income of the members of a household. Given additional time, this could potentially be improved by adding a scaling factor to the probabilities of each internet-usage activity based on the percent of time per week spent on it by average members of a given income category.

More specifically, we could develop a chart $P(A, G, I)$ for a person of income bracket I as opposed to $P(A, G)$. One possible method of formulating could be

$$P(A, G, I) = P(A, G)\frac{T(A, I)}{T(A)_{\text{avg}}}$$

where $T(A, I)$ is the average time spent on A by members of income bracket I and $T(A)_{\text{avg}}$ is the average time spent on A across all people. Both of these can be sourced from $D4$ as well. However, we also do not have concrete information on all of the household incomes of those in our sample cases, so we would have to make further assumptions to apply this modification.

Another potential weakness of the simulation as a whole lies in the assumption that a person's schedule can be broken into independent 30 minute blocks. As mentioned in the assumptions section, a real person's schedule will often have some degree of correlation between consecutive activities, as well as with the other members of the household. However, our model is still versatile in the sense that it allows for a large degree of potential variance among the different activities, testing out various possible combinations of activities that might occur in a household.

3 Mobilizing Mobile

3.1 Problem Summary

Improvements in cellular network technologies, particularly in 5G, requires infrastructure to deploy and utilize, which costs money to produce and build. To give everyone the maximum possible bandwidth would require expensive towers to be placed everywhere, an impractical feat. Therefore in this section, we hope to produce an optimal plan of the distribution/placing of cellular nodes in a region to meet the demands of the population in a region in an efficient manner.[1]

3.2 Assumptions

1. Populations in a subregion are largely homogeneous
   - Subregions are small enough for the living conditions to be relatively similar
2. People value a bandwidth provided lower than their demand very poorly
   - If the internet is too slow to meet the demand then people will be frustrated and annoyed
3. People value a bandwidth provided higher than their demand slightly poorly
   - If the internet is much faster than they need then people would think of it as a waste of money, but would not feel the frustration and annoyance than if it were lower
4. Cost of setting up a tower is similar among low, mid, high towers
   - The majority of the cost from building a tower comes from construction of the base and civil installation, the specific circuitry is far less expensive.[2]
5. Higher bandwidth will cost higher for a consumer than lower bandwidth
   - Typical ISPs will charge a higher cost for a high bandwidth than a lower bandwidth

3.3 Variables

Symbol	**Definition**	**Units**	**Value**
P	population	people	0-5000
A	area	mi^2	0-10
I	median income	dollars/year	0-1000000
M	median age	years	0-100
R_x	radius of low (1), mid (2), high (3) towers	mi	0-100
D_x	download speed of low (1), mid (2), high (3) towers	mbps	0-10000

Table 6: Variables

3.4 Developing the Model

We first model the demand for cellular data as $D(M)$. We can assume that income has no bearing on the demand as regardless of income, people use similar streaming services (i.e. Netflix, Hulu, Disney+), however there is a significant difference of usage between age groups due to generational trends. We model

Band	R	D
Low	10-20 mi	30-250 mbps
Mid	2-3 mi	100-900 mbps
High	0.5-1 mi	1000-3000 mbps

Table 7: Towers[11]

Band	Monthly Cost of Band
C_1	\$40
C_2	\$80
C_3	\$120*

Table 8: Cost for each type of Band[12] (*Extrapolated from other two values)

this as a piece-wise linear function peaking at $M = 20$ at about the data required for heavy usage of streaming.

To gauge the effectiveness of each type of tower in different scenarios, which we measure as S(x), we consider the marginal utility of the data provided compared to the marginal cost expended by the customer. The former is modeled by a log function of the ratio between data provided and the demand for cellular data of the population to accurately address the diminishing returns of excess data. The marginal cost is modeled by the cost of a plan relative to the median income, which properly reflects the marginal utility of a dollar. This model matches key traits of the dollar, including diminishing utility and an unbounded total utility for the buying power of a dollar. This is then multiplied by the total number of people who are provided service by the given tower to accurately reflect the total satisfaction provided by any single tower. We therefore optimize by picking towers with the highest satisfaction value in each subregion.

$$D(M) = \begin{cases} 5M & M \leq 20 \\ 120 - M & M > 20 \end{cases}$$

$$S(x) = \frac{\pi R_x^2 P}{A} \left[\frac{\log\left(\frac{D_x}{D(M)}\right)}{\frac{C_x}{I}} \right]$$

3.5 Predictions and Results

We evaluate the function for each type of tower in each subregion. However, you may happen to notice that all of the regions have 1 instance of the low tower, as it happened that it was unexpectedly always the optimal tower for maximizing the satisfaction of the inhabitants of the region.

Subregion	Tower Type	number of towers
1	low	1
2	low	1
3	low	1
4	low	1
5	low	1
6	low	1

Table 9: Region A

Subregion	**Tower Type**	**number of towers**
1	low	1
2	low	1
3	low	1
4	low	1
5	low	1
6	low	1
7	low	1

Table 10: Region B

Subregion	**Tower Type**	**number of towers**
1	low	1
2	low	1
3	low	1
4	low	1
5	low	1
6	low	1
7	low	1

Table 11: Region C

3.6 Strengths and Weaknesses

Our model takes into account how satisfaction changes with income and age, taking into account the diminishing returns of both the value of the dollar and the value of excess data provided. This allow us to observe the various advantages and disadvantages of certain towers due the limitations of their range and speed.

One shortcoming of our model is that the gauge for satisfaction is inherently difficult to measure to the subjectivity of the issue. Furthermore, the use of the log function does not satisfaction of matching demand, and even the satisfaction of people who receive mediocre service. Given more time, we would find a better suited function.

Another potential weakness is assumed homogeneity of subregions. To address this, it would require more detailed data of each subregion as well as updated to the model to take into account of various factors like population distribution and the irregular shape of land.

One other significant weakness of the model is that it always outputs that a single small tower is the best option. To be fair, even the small tower gives a significant amount of bandwidth if we compare it's values to the requirements explored in problem 2. At least in the way we set up the cost function, the value of additional bandwidth is typically not worth the price as you can do pretty much anything you would normally need to do with the low tower. This could be different if people used higher bandwidth frequently, as the more powerful towers would then be more useful, however, as things stand at the moment, we stand by our conclusion that simply using low bandwidth 5G towers throughout any of the given regions serves as the optimal rollout in terms of customer satisfaction as well as cost of implementation. Why else do you think 5G is being popularized everywhere.

We realize the log function is perhaps one of the least apt functions we could choose however, in the moment it seemed like a good idea. Due to time constraints, we will refrain from any further revisions for this paper.

References

[1] https://www.omnisci.com/technical-glossary/5g-infrastructure

[2] http://wirelessestimator.com/content/industryinfo/184

[3] https://www.sleepfoundation.org/how-sleep-works/how-much-sleep-do-we-really-need

[4] https://www.nielsen.com/us/en/insights/article/2020/the-nielsen-total-audience-report-hub/

[5] https://www.nerdwallet.com/article/finance/how-to-decide-what-internet-speed-you-need

[6] https://www.poundsterlinglive.com/best-exchange-rates/best-us-dollar-to-british-pound-history

[7] https://www.newamerica.org/oti/reports/cost-connectivity-2020/

[8] https://www.newamerica.org/oti/reports/cost-connectivity-west-virginia/

[9] https://www.akamai.com/us/en/resources/our-thinking/state-of-the-internet-report/archives/state-of-the-internet-security-reports-2017.jsp

[10] http://www2.southeastern.edu/Administration/Inst-Research/Student/data.cgi?stuage.txt

[11] https://venturebeat.com/2019/12/10/the-definitive-guide-to-5g-low-mid-and-high-band-speeds/

Code for Bit by Bit Model (C++)

```
#include <fstream>
#include <vector>
#include <algorithm>
#include <ctime>

const double arr[9][7] = { // The data table
    // 2-11     12-17     18-34   35-49   50-64   65+     MBPS
       {0.1221, 0.0731, 0.1040, 0.2095, 0.3397, 0.4261, 4},
       {0.0269, 0.0369, 0.0309, 0.0141, 0.0039, 0.0013, 3},
       {0.0632, 0.0355, 0.0516, 0.0497, 0.0351, 0.0226, 8},
       {0.0000, 0.0342, 0.0322, 0.0379, 0.0396, 0.0250, 1},
       {0.0000, 0.0116, 0.0109, 0.0080, 0.0066, 0.0031, 4},
       {0.0000, 0.2171, 0.2043, 0.2153, 0.1782, 0.1300, 1},
       {0.0000, 0.0247, 0.0232, 0.0164, 0.0099, 0.0065, 4},
       {0.0000, 0.0331, 0.0331, 0.0469, 0.0503, 0.0582, 1},
       {0.0000, 0.0091, 0.0085, 0.0084, 0.0060, 0.0053, 4}
};

const int sleep[6] = { 9, 8, 7, 7, 7, 7 }; // Hours of sleep for each age group

int getAgeGroup(int age) // Get the age group based on the actual age
{
    int agegroup = 0;
    if (age < 11)
        agegroup = 0; // indexing at 0
    else if (age < 17)
        agegroup = 1;
    else if (age < 34)
        agegroup = 2;
    else if (age < 49)
        agegroup = 3;
    else if (age < 64)
        agegroup = 4;
    else
        agegroup = 5;
    return agegroup;
}

int getMBPS(int age, int time)
{
    double random = (rand() % 10000) / 10000.0; // random number
    double count = 0;
    int agegroup = 0;
    agegroup = getAgeGroup(age);

    if (time % 24 >= 24 - sleep[agegroup])
        return 0; // Asleep
```

```
    //if (age < 50 && time > 48)
    //  return 0; // children not at the household for scenario 2

    // determine activity based on random number and probabilities
    for (int j = 0; j < 9; j++)
    {
        count += arr[j][agegroup];
        if (random <= count)
            return arr[j][6]; // return the Mbps for that activity
    }
    return 0; // for awake but not using internet
}

int * performAnalysis(std::vector<int> age)
{
    int * mbpsData = new int[168 * 2];
    for (int i = 0; i < 168 * 2; i++) // simulate a week by half-hour inteverals
    {
        mbpsData[i] = 0;
        for (int j = 0; j < age.size(); j++) // Find the total Mbps over household
            mbpsData[i] += getMBPS(age[j], i / 2);
    }
    std::sort(mbpsData, mbpsData + 168 * 2); // sort Mbps; easily find 90 and 99%

    int * data = new int[2];
    // 90th percentile
    data[0] = mbpsData[302]; // 302 is the index for the 90th percentile
    // 99th percentile
    data[1] = mbpsData[332]; // 332 is the index for the 99th percentile
    return data;
}

int main()
{
    int  r;
    std::srand(time(0)); // set a seed for random number generation
    r = rand();
    int * histogram90 = new int[10000];
    int * histogram99 = new int[10000];
    std::vector<int> ages; // Below is for Scenario 3
    ages.push_back(21); ages.push_back(21); ages.push_back(21);
    for (int i = 0; i < 10000; i++)
    {
        int * data = performAnalysis(ages); // Do one simulation
        histogram90[i] = data[0]; // Store the 90th percentile
        histogram99[i] = data[1]; // Store the 99th percentile
    }

    std::ofstream fout("data90.csv"); // Store the data into a .csv file
```

```
    for (int i = 0; i < 10000; i++)
        fout << histogram90[i] << std::endl;
    fout.close();
    fout = std::ofstream("data99.csv"); // Store the data into a .csv file
    for (int i = 0; i < 10000; i++)
        fout << histogram99[i] << std::endl;
    fout.close();
    return 0;
}
```