Bandwidth and the Future of High Speed Internet

Team 14684

February, 2021

1. Executive Summary

At the end of 2019, the COVID-19 pandemic set in. Since its beginning, the world has changed drastically in an attempt to adapt to the realities and constraints of limiting its spread. As a result of the virus's threat to global human life, many governments have created restrictions to slow the spread of the COVID-19 disease while vaccines continue to be developed. In the time since the end of 2019, the distribution of vaccines has begun, but the pandemic has left a permanent mark on the world in more ways than one. [1]

One of the most interesting, and perhaps unexpected, result of the COVID-19 pandemic is the turn to life at home. Activities that were once exclusively in-person, such as schooling, various types of employment, and even sports, music, and the functions of government, have been forced to turn to virtual alternatives in an attempt to slow the spread of the coronavirus disease. Where once children crowded into buildings and workers into offices, now stands only the internet. Whether through text, email, or voice call, the entire global economy has gone virtual, and for better or worse, will likely remain as such. [2]

Whether the pandemic accelerated a process that was already in motion or caused only a temporary increase in internet demand, is a debate for another time. One fact that is undeniable, however, is that demand for the internet is growing, and whatever the reasons, must be dealt with. Governments would be prudent to investigate this matter and work towards ensuring adequate broadband internet access for all, especially during pandemic times.

Addressing the cost of broadband in the future may be a first step. For the first problem, we performed a linear regression to compare average peak download speed to average download speeds in the US, allowing us to model future download speeds. Adjusting the data to newer data to avoid conflict, we could predict values for 2020 with an exponential regression. Data for the median monthly price paid for plans with different download speeds, but not the overall average download speed or overall average monthly price, so we used a scatter plot with a linear regression.

For the next part of the question, we modeled the minimum amount of required bandwidth needed to cover 90% and 99% of the total internet needs of three distinct households. While approaching this problem, we took into consideration the ages of members of the household and their typical internet usage activities. Age is often an indicator of the time spent online, such as the youngest age groups spending less time on smartphone devices than an older demographic that spends more time on streaming services and social media that require greater bandwidth. Based on the average bandwidth required per age category, we were able to model the total bandwidth required in a household per year by dividing members of the household by age category and applied them to three separate scenarios.

Finally, we determined the broadband demand in the geographical locations across the three regions using the equation we derived in Part 2. 5G cellular nodes were preferred due to their greater connectivity density and lower cost. Each subregion was determined to be urban or rural based on population density, where urban areas have higher demand and lower node range. Cell node placement about Region C was determined using a computational model. We created a Python program that maps Subregion C onto a 2-D coordinate system and computes the optimal cell node distribution about the plane for 5,000 nodes using a Monte Carlo simulation. The optimal node distribution was defined as the distribution that best addressed the bandwidth needs of each subregion based on the reception provided by each node.

2. Contents

3. Global Assumptions

1. It is assumed that inflation will not play a significant role in changing the cost per unit bandwidth per Mbps over the next decade.
    - Justification: In order to ensure stable prices and consistent demand for broadband internet, we will assume that there will not be any inflation significant enough to alter any of the data we examine in this paper.
2. It is assumed that there will be no major revolutions in bandwidth technology (eg. 6G) within the next decade. [10]
    - Justification: In order to ensure that broadband demands are met, we are assuming that no significant changes will occur in broadband distribution technology, so as not to alter our findings.

4. Part 1: The Cost of Connectivity

4.1 Restatement of the Problem

- The problem asks us to create a mathematical model to predict the cost (in dollars) per Mbps of bandwidth over the next 10 years in the United States and the United Kingdom.

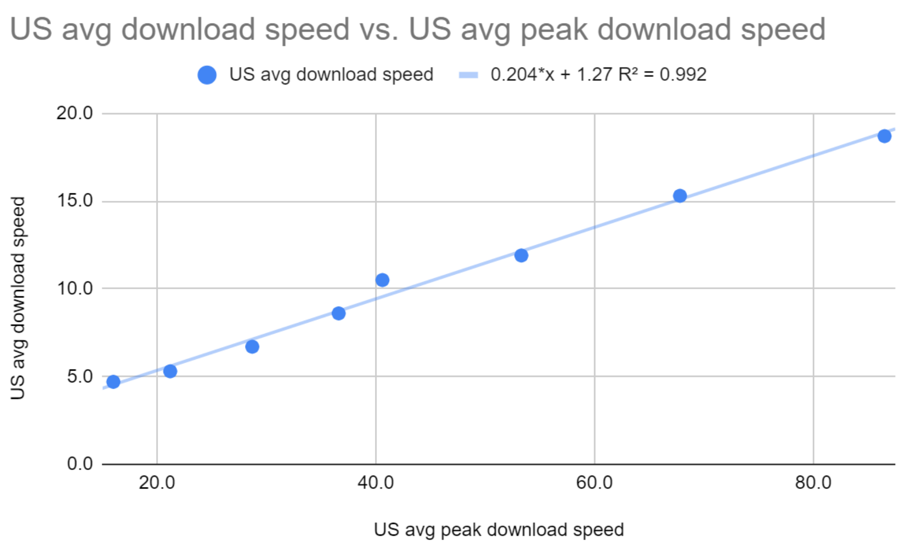
4.2 Assumptions

1. It is assumed that US average peak download speed is directly proportional to US average download speed
    - Justification: See strengths in section 4.4.
2. It is assumed that average peak download speed solely determines average download speed (in other words, year does not play a factor in the model).
    - Justification: Because we are looking at 10 years of information, we do not want annual changes like seasons/weather and natural events to interfere with our information. We assume that change of seasons does not affect our information.
3. It is assumed that the newer data (Ookla) is more accurate than the older data (akamai), to better fit trends found in Ookla data, values in Akanami data were reduced by 3.
    - Justification: Technology has improved, and as a result, Ookla's measurements have become more accurate as compared to those of Akamai.

4. For the data from D2: we're using the median values bc average is higher than the median, meaning the distribution of speed is skewed right. To represent the average user, we use median-- we can speculate that these outliers are huge companies that don't represent the average consumer.
    - Justification: Because outliers disproportionately influence our data, we are discounting them, as they do not affect our findings except the skew the data.

4.3 Developing the Model

We're given the average download speeds and average peak download speeds (Mbps) for the US from 2009 to 2021 [D1], however, the average download speeds for 2017 onwards are missing. Plotting the US average peak download speed versus the US average download speed and performing a linear regression yields the following:



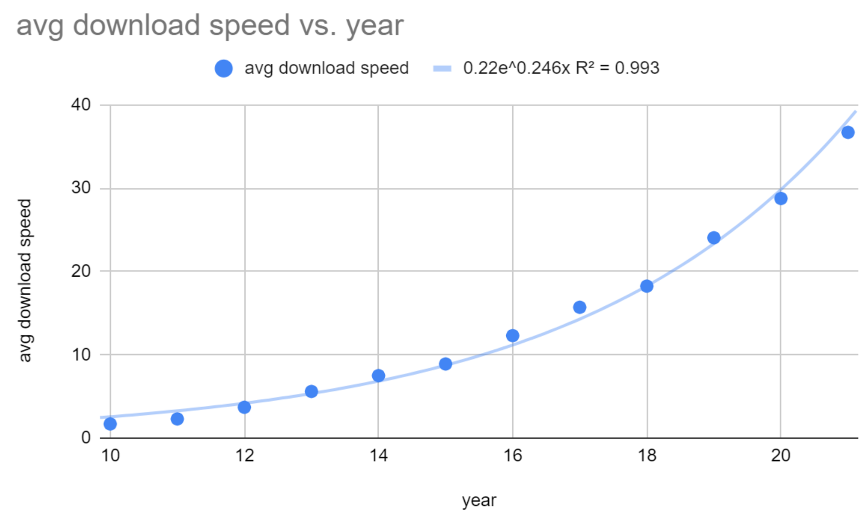
The R^2 value 0.992 means 99.2% of the variability in the average download speed is explained by changes in the average peak download speed, meaning that we can reasonably model future download speeds D as a function of peak download speeds D_P with the function:

$$D(D_P) = 1.27 + 0.204*D_P,$$

We can then use this equation and numbers for the peak download speeds to predict the average download speeds from 2017 onwards, which would be (in Mbps) 15.7 in 2017, 18.2 in 2018, 24.0 in 2019, 28.8 in 2020, and 36.7 in 2021.

However, we notice a conflict in the values for 2017. Assuming that the newer values from Ookla are more accurate, we can adjust the past data from Akamai and make the values for 2017 "mesh" by decreasing all old (Akamai, pre-2017) values by 3 Mbps, which is valid because the association still yielded a high R^2 value and good predictive power later on.

We could then plot the years after 2020 against the average download speed and perform an exponential regression to yield the following:

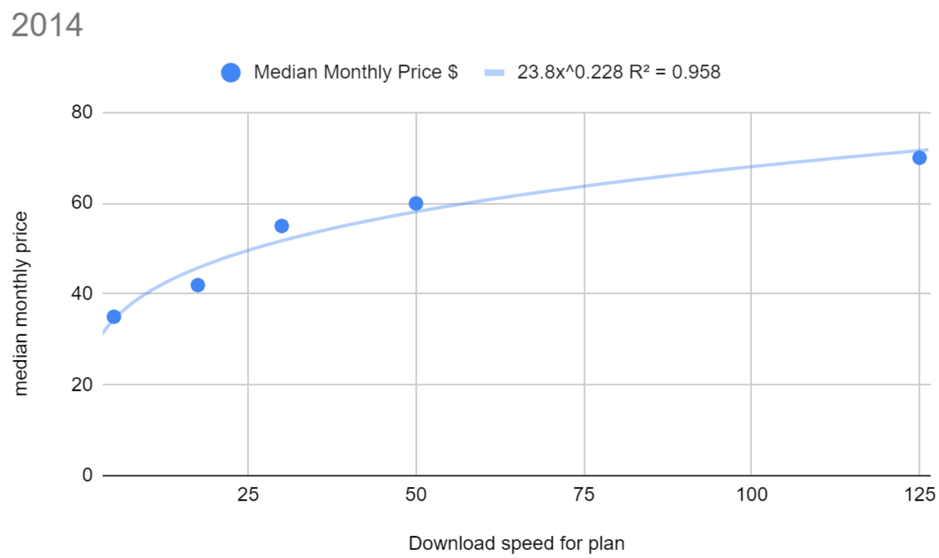


The R^2 value 0.993 means 99.3% of the variability in the average download speed is explained by the change in year, meaning that we can reasonably model future download speeds *D* as a function of *t* years after 2000 with the function:

$$D(t) = 0.22 * e^{0.246t}$$

This equation can be used to make predictions for the average download speeds after 2021, resulting in speeds of 49.3 in 2022, 63.05 in 2023, 80.63 in 2024, 103.12 in 2025, 131.88 in 2026, 188.66 in 2027, 215.70 in 2028, 275.85 in 2029, 352.79 in 2030, 451.18 in 2031, and 577.02 in 2032. We know this is reasonable because if we use the prior equation we figured $D(D_P)$ to convert the average download speed in 2022 (49.30) and 2030 (352.70) to peak download speeds (2022: 235.43 and 2030: 1723.14) and compare that change in peak download speed over eight years (733% increase) to the average change in peak download speed from 2012 to 2020 in D2 (819% increase), we can see that such a huge jump isn't out of the ordinary, assuming past trends continue.

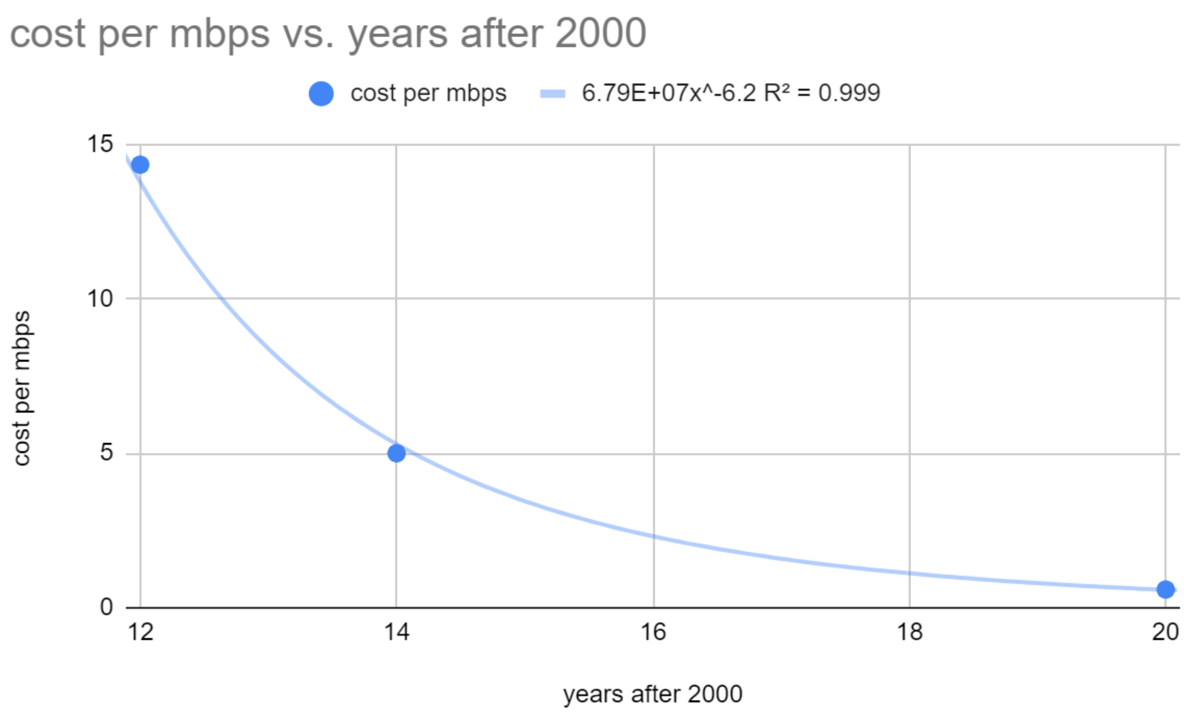
Data from 2014 [D2] lists the median monthly price paid for plans with different download speeds, but not the overall average download speed or overall average monthly price. So we can make a scatter graph with download speed for each plan on the x-axis (when a range is given, such as 4-6 Mbps, we'd use an average of 5 Mbps) and find a logarithmic regression, and then plug that value into the logarithmic equation above to find the the median monthly price in 2014 is 34.46.



We can then take the derivative of D2 data between median peak download speeds and median monthly price, and then fit in the median peak download speeds predicted by our first equation to find the price per Mbps in 2020, 2012, and 2014. Extrapolating this forward yields this data

years after 2000	cost per mbps
12	14.34
14	5
20	0.59
21	0.4306358157
22	0.3227372939
23	0.244994983
24	0.1881741988
25	0.1460971421
26	0.1145605307
27	0.09065973317
28	0.07235857075
29	0.05821061495
30	0.04717566111

Which, when plotted, yields this regression:



4.4 Validity of the Model

Strengths:

- Our prediction that the average US download speed would be 28.8 Mbps in 2020 is incredibly close to the actual value of 29.00 Mbps in the second quarter of 2020 [11], supporting the predictive power of our model. This also supports our assumption that the average peak download speed solely determines the average download speed, in other words year does not play a factor.

Weaknesses:

- The model is solely based on information from the United States. Due to time constraints, information for the United Kingdom was not listed. Despite this, the model's method can be used to predict the information for the UK, the results of that particular application are simply not shown in this paper.

5. Part 2: Bit by Bit

5.1: Restatement of the Problem

The problem asks us to do the following:

- Calculate the minimum amount of required bandwidth that covers the total internet needs of the following individuals 90% and 99% of the time:
  - Couple in their 30s (teacher & unemployed) with a 3-year-old
  - Retired woman in her 70s who looks after her 2 grandchildren twice a week
  - Three former M3 Challenge Participants sharing an off-campus apartment working a part-time job and full-time to complete their undergraduate degrees.

5.2 Assumptions

1. It is assumed that internet usage data from the first quarter of 2020 can be used as a representation of future average internet usage by category.
   - Justification: Quarter 1 of 2020 includes internet usage in March when more school and work transitioned to a remote system. Due to variations of remote students/workers based on infection rates and limited data on the increase of internet usage during this period, we assumed Quarter 1 could be representative for the next year.
2. It is assumed that a "school-age child" resides within the age bracket of 5-12 years old.
   - Justification: It is more likely than not that minors who require babysitting and are described as "children" fall into the first age range of 2-11 years old.

5.3 Developing the Model

We began by identifying the categories of internet usage compromising typical internet usage within a household by assigning each activity a value of n as shown below.

n=1	Watching Traditional Television	n=7	Streaming Audio on a Smartphone
n=2	TV Connected Game Console	n=8	Smartphone Misc.
n=3	TV Connected Internet Device	n=9	Streaming Audio on a Tablet
n=4	Internet on a Computer (not including video)	n=10	Video Focused App/Web on a Tablet
n=5	Video on a Computer	n=11	Tablet Misc.
n=6	Video Focused App/Web on a		

	Smartphone		

Figure 5.3.1: Internet Media Consumption (Activities)

Next, we resolved to find the average broadband requirements of a singular person in a household based on their age range as factors like schooling, work, and leisure time that affect internet usage are age-dependent. We defined the age ranges in alignment with data from the Nielsen Corporation [6], as shown below:

Age Range	Symbol
2-11	a
12-17	b
18-34	c
35-49	d
50-64	e
64 +	f

Figure 5.3.2: Age Distributions

Using this information, we calculated the average bandwidth required per year, or $P_{\text{age range}}$ for each age range, where h_{na} is the average time spent by age range in hours per week and B_{na} is the average bandwidth requirement for said activity[6][9]. A summation is used to find the total for all activities. Multiplied by 3600 seconds, this value is converted to Mbpw.

$$P_a = 3600 \sum_{n=1}^{11} h_{na} B_{na} \qquad P_d = 3600 \sum_{n=1}^{11} h_{nd} B_{nd}$$

$$P_b = 3600 \sum_{n=1}^{11} h_{nb} B_{nb} \qquad P_e = 3600 \sum_{n=1}^{11} h_{ne} B_{ne}$$

$$P_c = 3600 \sum_{n=1}^{11} h_{nc} B_{nc} \qquad P_f = 3600 \sum_{n=1}^{11} h_{nf} B_{nf}$$

P_a	P_b	P_c	P_d	P_e	P_f
167160 Mbpw	113400 Mbpw	360900 Mbpw	485460 Mbpw	619080 Mbpw	687540 Mbpw

Figure 5.3.3: Average Bandwidth Requirement Per Week (Mbpw)

Finally, to find the bandwidth requirement for the household per week, the number of members in a certain age range ($N_{age\ range}$) was multiplied by the average bandwidth requirement calculated in the previous step. To convert to Mbpy, H_W was multiplied by 52.1775 weeks.

$$H_W = N_a P_a + N_b P_b + N_c P_c + N_d P_d + N_e P_e + N_f P_f$$
$$H_Y = 52.1775(H_W)$$

Applying the Model:

	N	H_w (Mbpw)	H_Y(Mbpy)	90%	99%
Scenario 1	N_a= 1 ; N_c=2	888,960	46,383,710.40	41,745,339.36	45,919,873.30
Scenario 2	N_a=2 ; N_f=1	1,021,860	53,318,100	47,985,830.30	52,784,413.33
Scenario 3	N_c=3	1,082,700	56,492,579.25	50,843,321.33	55,927,653.46

After applying the model mentioned previously, our data yields the following data:

- Scenario 1: The couple and their child would need 41,745,339.36 megabits to cover their total internet needs 90% of the time. They would need 45,919,873.30 megabits to cover 99%.
- Scenario 2: The woman and her grandchildren would need 47,985,830.30 megabits to cover their total internet needs 90% of the time. They would need 52,784,413.33 megabits to cover 99%.
- Scenario 3: The former M3 Challenge participants would need 50,843,321.33 megabits to cover 90% of their total internet needs. They would need 55,927,653.46 megabits to cover 99%.

5.4 Validity of the Model

Strengths:

- The model finds accurate information for how much broadband service would be required to fulfill the internet requirements of the sample households. When looking at other information online about the average internet consumption in American households, the numbers produced by this model are within a very low margin of error as compared to actual values, indicating its accuracy.

- The model takes into consideration factors like age range and differentiates between different types of internet usage and their respective bandwidth requirements.

Weaknesses:

- The model does not take into account the income of the given individuals into consideration. The accessibility of certain devices and internet connections can vary across households of different income levels. However, our model relies solely on the assumption that the average hours spent per week for each range incorporates some of those differences. Factoring this into our model would have required an analysis of the accessibility to devices based on income and/or through educational organizations like schools and universities and organizations employing people.
- The model does not consider the occupations of the individuals and how their internet usage may be influenced by variations in remote/in-person learning or working environments. Analysis of these factors would have required modeling future changes in internet usage influenced by factors like vaccination rates, school reopenings, and government-level restrictions due to insufficient data on such trends after the first quarter of 2020.

6. Part 3: Mobilizing Mobile

6.1: Restatement of the Problem

The problem asks us to do the following:

- Develop a model that predicts the optimal placement of cellular nodes in a region.
- Ensure the model can meet the bandwidth requirements of a region.
- Prove the model meets the bandwidth requirements of a region by using it to determine if the bandwidth needs of given regions are covered.

6.2 Assumptions

1. It is assumed that there are currently no cellular nodes in Regions A, B, or C.
    - Justification: Because the purpose of the model is to determine the optimal placement of nodes, we will determine that no nodes are currently present, and all nodes will be placed in accordance with the model once it is developed.
2. It is assumed that users of the model will prioritize 5G over 4G cellular nodes.
    - Justification: While 4G and 5G are both mobile broadband internet services, we assume that those who would use this model, whether they be private corporations, governments, or other individuals, would be more interested in the placement of 5G cellular nodes because they are more efficient, and easier to build.
3. It is assumed that the model in Part 2 of this paper accurately predicts the broadband demand for a household.

- Justification: Question 2 of this paper addresses how much broadband internet is demanded by individual households. To determine how best to satisfy that demand, it is necessary to know what the values of that demand are. As such, we will assume that the model in Part 2 accurately predicts the demand for each household, so that we may use that information to help determine the placement of cellular nodes.

4. It is assumed that all 5G cellular nodes have the same broadband capacity of around 925 Mbit/s.
    - Justification: Sources [7] and [8] show that 5G microcells allow for speeds of between 50 Mbit/s and 1.8 Gbit/s. Because of the significant variation, we will assume that each 5G microcell has the same broadband capacity of approximately 925 Mbit/s, which is the average of these two extremes.
5. It is assumed that, for each subregion, all people use the internet in accordance with the age charts on page D4 of M3's statistics, in accordance with the median age of the population.
    - Justification: Age is a factor in the equation designed in Part 2. To use it, we must know the age of the population. The median, as the most commonly occurring age, represents a fair sample of the total population, and as such can be used in order to determine internet demand for each region.
6. It is assumed that all people will demand broadband internet if it is supplied to them.
    - Justification: To decide optimal broadband node placement, we will assume that all people will want internet access, and optimize our nodes based on the demand for broadband internet if all people were demanding access.
7. It is assumed all subregions are squares.
    - Justification: In order to use the optimization simulation with the greatest efficiency, all subregions will be squares.
8. It is assumed that any area with a population density greater than 10,000 people/sq mi is an urban area, and any area with a density less than or equal is rural.
    - Justification: In order to use the optimization simulation with the greatest efficiency, it is necessary to determine which areas are rural and which are urban, so as to be able to properly adjust the range of the 5G microcell. According to 5Gradar, the range of a microcell is dependent on its environment. [5]

6.3 Developing the Model

First, it is important to understand the differences between 5G, the latest broadband provider, as compared to previous generations of internet communications. Previous internet distribution methods, such as 4G, used large cellular towers to cover massive areas, with ranges of up to 30 miles. [4] However, 5G towers operate at higher radio frequencies, and as a result, are frequently blocked by many common construction materials. Because of this, 5G cellular nodes are designed to function as microcells, rather than large towers, with ranges of between 1,500 feet and 2 miles. [5] As a result, several dozens of 5G cellular nodes would be required to cover the same area of one 4G tower, but these nodes are much smaller, much easier to install, and provide greater connectivity density, which allows for a higher amount of data to be transmitted.

The usage of 5G microcells rather than macrocells (large towers) allows for a greater density of 5G connectivity, which is especially important for urban areas where demand is high. Furthermore, according to most information, the labor and time costs of installing a 5G microcell are low. Although the cost of the equipment may be slightly higher due to the higher standard of 5G equipment as compared with 4G, the overall cost is lower due to the ease of installment.

With this in mind, in order to determine the optimal placement of 5G microcells across the three regions, we must first determine the broadband demand in geographical locations. In order to do this, we will use the equation from Part 2 to find the demand for broadband internet (in Mbit/s) of each person in each subregion. The demand for broadband internet, in Mbit/s is as follows:

$$P_a = 3600 \sum_{n=1}^{11} h_{na} B_{na} \qquad P_d = 3600 \sum_{n=1}^{11} h_{nd} B_{nd}$$

$$P_b = 3600 \sum_{n=1}^{11} h_{nb} B_{nb} \qquad P_e = 3600 \sum_{n=1}^{11} h_{ne} B_{ne}$$

$$P_c = 3600 \sum_{n=1}^{11} h_{nc} B_{nc} \qquad P_f = 3600 \sum_{n=1}^{11} h_{nf} B_{nf}$$

Figure 6.3.1: Equations for Demand of Internet for a Person in an Age Group in Mbit/s per week

	Region A	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6
Population	6327	690	1422	1303	278	1243	1391
Median Age (Yrs)	35.5	28.2	30.2	40.7	64.3	37.8	36.9
Total Broadband Demand	4838.272 Mbit/s	385.907 Mbits/s	795.304 Mbits/s	996.407 Mbit/s	267.873 Mbit/s	950.525 Mbit/s	1063.701 Mbit/s
Demand Per Person	.764702 Mbit/s per	.559286 Mbits/s per	.559286 Mbits/ per	.764702 Mbit/ per	.963571 Mbit/s per	.764702 Mbit/s per	.764702 Mbit/s per

	person	person	person	person	person	person	person

Figure 6.3.2: Broadband Demand Data for Region A

	Region B	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6	Subregion 7
Population	15054	3873	2114	1253	1129	2493	2398	1794
Median Age (Yrs)	46.6	47.7	48.2	59	36.3	55.1	45.5	53.8
Total Broadband Demand	13828.871 Mbit/s	2961.691 Mbit/s	1616.580 Mbit/s	1207.355 Mbit/s	863.349 Mbit/s	2402.184 Mbit/s	1833.755 Mbit/s	1728.647 Mbit/s
Demand Per Person	.764702 Mbit/s per person	.764702 Mbit/s per person	.764702 Mbit/s per person	.963571 Mbit/s per person	.764702 Mbit/s per person	.963571 Mbit/s per person	.764702 Mbit/s per person	.963571 Mbit/s per person

Figure 6.3.3: Broadband Demand Data for Region B

	Region C	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6	Subregion 7
Population	9505	1468	1624	1012	1295	1309	1008	1789
Median Age (Yrs)	41.3	41.67	34.9	47.1	44.7	32.7	47.9	40.7
Total Broadband Demand	7268.493 Mbit/s	1122.583 Mbit/s	908.280 Mbit/s	773.878 Mbit/s	990.289 Mbit/s	732.105 Mbit/s	770.820 Mbit/s	1368.052 Mbit/s
Demand Per Person	.764702 Mbit/s per person	.764702 Mbit/s per person	.559286 Mbits/s per person	.764702 Mbit/s per person	.764702 Mbit/s per person	.559286 Mbits/s per person	.764702 Mbit/s per person	.764702 Mbit/s per person

Figure 6.3.4: Broadband Demand Data for Region C

With the information about the broadband demand of each subregion now generated via the equation from Part 2, we can now begin the process of optimizing placement of 5G microcells. To do this, we will be using a Python simulation, which will test random placements of 5G microcells in a subregion, starting with only one, and continue increasing the number and placement of the cells until coverage of the subregion is universal, and total demand for broadband internet in the subregion is satisfied. As stated in our assumptions, each 5G microcell can handle an output of up to 925 Mbit/s. Meanwhile, coverage range varies depending on the environment.

According to 5Gradar, a 5G microcell can cover a range of approximately 1,500 feet in an urban setting, but up to 2 miles in a rural setting. [5] To find which subregions are urban and rural, we will find the population density of each subregion and compare it to a test case, in this case, the fictional City of Villagetown, with a population density of 10,000 people/sq mi. Any subregion with a population density greater or equal to that of the control case will be classified as urban (and therefore will use the 5G range of 1,500 feet in the simulation), while any subregion with a population density less than that of the control case will be classified as rural (and therefore will use the 5G range of 2 miles in the simulation).

	Region A	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6
Population	6327	690	1422	1303	278	1243	1391
Area (Sq. Mi)	6.83	1.21	0.8	0.67	1.65	0.36	2.14
Population Density (People/Sq Mi)	Not Calculated	570.248	1777.5	1944.776	168.485	3452.778	650
Classification	N/A	Rural	Rural	Rural	Rural	Rural	Rural

Figure 6.3.5: Classifications and Population Density for Region A

	Region B	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6	Subregion 7
Population	15054	3873	2114	1253	1129	2493	2398	1794
Area (Sq. Mi)	33.64	3.48	4.35	4.64	2.32	7.54	3.77	7.54
Population Density (People/Sq Mi)	Not Calculated	112.931	485.977	270.043	486.638	330.637	636.074	237.931
Classification	N/A	Rural	Rural	Rural	Rural	Rural	Rural	Rural

Figure 6.3.6: Classifications and Population Density for Region B

	Region C	Subregion 1	Subregion 2	Subregion 3	Subregion 4	Subregion 5	Subregion 6	Subregion 7
Population	9505	1468	1624	1012	1295	1309	1008	1789
Area (Sq. Mi)	1.64	0.38	0.14	0.1	0.24	0.13	0.31	.34
Population	Not	3863.158	11600	10120	5395.833	10069.231	3509.677	5261.765

Density (People/Sq Mi)	Calculated							
Classification	N/A	Rural	Urban	Urban	Rural	Urban	Rural	Rural

Figure 6.3.7: Classifications and Population Density for Region C

We demonstrated the effectiveness of the model by testing Region C using a Python model. Using the known area of the region as well as its constituent subregions, we converted the original, irregular shapes of these geographic territories into rectangles with the same dimensions and area as the actual regions. These transformed, rectangular subregions were translated onto a 2D graph, where subregion is associated with its previously calculated minimum bandwidth demand, an urban/rural classification, and a “receptivity” representing the bandwidth each square mile receives based on the nodes around it. We then used the Monte Carlo method of optimization to maximize the percent of households that would have their minimum demand of bandwidth met. Functionally, that meant choosing random positions for some number of nodes, calculating the percent of households whose bandwidth needs were met, and then repeating this iteration several times to find the maximum percentage of satisfied households.

While the Simulation would yield the results we are searching for, one unfortunate downside of the program is that it requires a large amount of computational power. We are limited by the hardware we are using, and therefore cannot run the simulation to its greatest potential. However, in the limited 100,000 trials completed, we received an average of approximately 5,000 nodes needed to cover Region C, and would satisfy 83% of the broadband demand. With a superior computer, we could get more accurate results.

6.4 Validity of the Model

Strengths:

- The calculation of total broadband demand for a given subregion in accordance with the equation from Part 2 seems to be realistic and accurate. Although there is a small margin of error when comparing the added values of demanded broadband for each of the subregions as compared to the calculated value for the entire region, they are within a close margin of error, indicating the model is at the very least mostly effective.
- The Optimization Program allows for the placement of 5G microcells at the most efficient location possible, minimizing the price of installation and granting the greatest amount of coverage possible, while also satisfying the regional demand.

Weaknesses:

- The units of the model are unclear. Although the equation produces accurate results, it is unclear if those results should be measured in Mbit/s per week, as the equation for P_a produces, or if the unit needs to be brought back to Mbit/s. Despite this, the unit problem appears to be irrelevant to the validity of the actual numbers, as they are correct regardless of whether Mbit/s or Mbit/s is the unit.
- The Optimization Program assumes that the subregions are perfect squares, which is not realistic. For this model to be applied, it would have to take into account the geographical shape of the subregions.
- Used 2019 data, rather than 2020 data. This is acceptable because, according to the data on D4 of M3's spreadsheet, the first quarter of both years are very similar.
- Our computational power is weak, but with a supercomputer, we could get much greater precision.

7. Conclusion

As access to the internet becomes increasingly more crucial in educational and occupational spaces, or simply in staying informed, the importance of the cost, the quantity, and the geographical considerations for bandwidth also increase as a direct result.
By plotting the US average peak download speed against the US average download speed, created an equation that can be used to model future download speeds. By performing an exponential regression, we were able to predict the average download speed after 2021. The second model used a series of summations to find a household's requirements for bandwidth per year based on the main factor of age, which influenced the types of activities each subgroup most frequently engaged with. In applying this model to three scenarios, it was found that a retired grandmother caring for two children would require the most bandwidth because of her available average leisure time and children's growing screen time. This model was able to consider some factors contributing to bandwidth requirements but did not look at trends and fluctuations associated with changing conditions in the pandemic or accessibility to devices. Our final model utilized the equation from Part 2 to determine the broadband demand geographic territories within the primaries regions. With this information, we were able to optimize the cell node distribution about the plane using a Monte Carlo simulation in Python, with consideration for population density as well. While this method works, further improvements can be made to this simulation. For instance, the implementation of computational geometry would have allowed the model to account for the geographic figures of the subregion.

8. References

[1] “Coronavirus Disease (COVID-19).” *World Health Organization*, World Health Organization, www.who.int/emergencies/diseases/novel-coronavirus-2019.

[2] Clement, J. “Topic: Coronavirus: Impact on Online Usage in the U.S.” *Statista*, www.statista.com/topics/6241/coronavirus-impact-on-online-usage-in-the-us/.

[3] Fisher, Tim. “5G Cell Towers: Why You See Them and How They Work.” *Lifewire*, 17 Mar. 2020, www.lifewire.com/5g-cell-towers-4584192.

[4] “Small Cell 101: Crown Castle.” *Small Cell 101 | Crown Castle*, www.crowncastle.com/communities/small-cell-information.

[5] McCaskill, Steve. “5G Microinfrastructure: Microcells, Femtocells and Picocells Explained.” *5Gradar*, 5Gradar, 23 Mar. 2020, www.5gradar.com/features/5g-microinfrastructure-microcells-femtocells-and-picocells-explained#:~:text=A%20range%20of%20up%20to,the%20capacity%20of%20the%20microcell.

[6] “The Nielsen Total Audience Report: August 2020.” *Nielsen*, www.nielsen.com /us/en/insights/report/2020/the-nielsen-total-audience-report-august-2020/.

[7] “What Is the Difference between 4G and 5G?” *Just Ask Thales EN*, 13 Nov. 2020, justaskthales.com/en/difference-4g-5g/.

[8] Jdolcourt. “We Tested 5G Speeds across the Globe.” *CNET*, www.cnet.com/features/ we-ran-5g-speed-tests-on-verizon-at-t-ee-and-more-heres-what-we-found/.

[9] NerdWallet. “How to Decide What Internet Speed You Need.” *NerdWallet*, 20 Feb. 2021, www.nerdwallet.com/blog/utilities/how-to-decide-what-internet-speed-you-need/.

[10] Kharpal, Arjun. “China Starts Development of 6G, Having Just Turned on Its 5G Mobile Network.”*CNBC*, CNBC, 18 Nov. 2019, www.cnbc.com/2019/11/07/china-starts-6g-development-having-just-turned-on-its-5g-mobile-network.html.

[11] Isla McKetta. “U.S. Internet Speeds Increase 15.8% on Mobile and 19.6% on Fixed

Broadband." *Speedtest Stories & Analysis: Data-Driven Articles on Internet Speeds*, Speedtest Stories & Analysis: Data-driven articles on internet speeds, 8 July 2020, www.speedtest.net/insights/blog/announcing-us-market-report-q2-2020/#:~:text=Data%20from%20Speedtest%20Intelligence%20reveals,down%2015.2%25%20from%20Q2%202019. Accessed 1 Mar. 2021.

Code for Problem 3

```
import random

AREA_SQUARE_MILES = 1.64
SCALE = 10 # each point represents a [scale] foot by [scale] foot area
MEDIAN_HOUSEHOLD_INCOME = 102486
NODE_MAX_MBPS_OUTPUT = 925
POPULATION_PER_SUBREGION = 1468, 1624, 1012, 1295, 1309, 1008, 1789
DEMAND_PER_PERSON_PER_SUBREGION_MBPS = 0.764702, 0.559286, 0.764702,
0.764702, 0.559286, 0.764702, 0.764702
TRIALS = 100000
TOTAL_NODES = 5000

side = int(AREA_SQUARE_MILES**0.5 * 5280 / SCALE)

'''
each point in "space" has three attributes:
     space[x][y][0] representing the RECEPTION which will be incremented
depending on the proximity to a node,
    space[x][y][1] representing the DEMAND PER UNIT AREA which will depend
on which subregion the point is in, and
       space[x][y][2] representing whether whether the subregion is
classified as URBAN (True) or RURAL (False) as determined in the paper
'''

space = []
for x in range(side):
    column = []
    for y in range(side):
        receptions = []
        receptions.append(0)
        receptions.append(0)
        receptions.append(True)
        column.append(receptions)
    space.append(column)

# This part of our code assigns the "subregions" and urban/rural value to
our space array. We understand this is messy, but a more concise and
accurate assignment would require detailed information about how each of
the subregions are shaped, such as a graph with points representing a
"connect the dots" for a subregion's border. This data does exist for
real-world regions, however.
```

```
#assigning demand and rural/urban variables
for x in range(int(side*0.2)+1):
    for y in range(int(side*0.75)):
        space[x][y][1] = (POPULATION_PER_SUBREGION[0] / AREA_SQUARE_MILES)
* (DEMAND_PER_PERSON_PER_SUBREGION_MBPS[0]) * (1/5280.0)**2 * (SCALE)**2
        space[x][y][2] = False
for x in range(int(side*0.25)):
    for y in range(int(side*0.75+1),side):
        space[x][y][1] = (POPULATION_PER_SUBREGION[1] / AREA_SQUARE_MILES)
* (DEMAND_PER_PERSON_PER_SUBREGION_MBPS[1]) * (1/5280.0)**2 * (SCALE)**2
for x in range(int(side*0.25+1), int(side*0.475)):
    for y in range(int(side*0.60+1), side):
        space[x][y][1] = (POPULATION_PER_SUBREGION[2] / AREA_SQUARE_MILES)
* (DEMAND_PER_PERSON_PER_SUBREGION_MBPS[2]) * (1/5280.0)**2 * (SCALE)**2
for x in range(int(side*0.475+1), int(side*0.7)):
    for y in range(int(side*0.6+1), side):
        space[x][y][1] = (POPULATION_PER_SUBREGION[3] / AREA_SQUARE_MILES)
* (DEMAND_PER_PERSON_PER_SUBREGION_MBPS[3]) * (1/5280.0)**2 * (SCALE)**2
        space[x][y][2] = False
for x in range(int(side*0.7+1), side):
    for y in range(int(side*0.4+1), side):
        space[x][y][1] = (POPULATION_PER_SUBREGION[4] / AREA_SQUARE_MILES)
* (DEMAND_PER_PERSON_PER_SUBREGION_MBPS[4]) * (1/5280.0)**2 * (SCALE)**2
for x in range(int(side*0.2+1), int(side*0.6+1)):
    for y in range(int(side*0.6+1)):
        space[x][y][1] = (POPULATION_PER_SUBREGION[5] / AREA_SQUARE_MILES)
* (DEMAND_PER_PERSON_PER_SUBREGION_MBPS[5]) * (1/5280.0)**2 * (SCALE)**2
        space[x][y][2] = False
for x in range(int(side*0.60+1),side):
    for y in range(int(side*0.6)):
        space[x][y][1] = (POPULATION_PER_SUBREGION[6] / AREA_SQUARE_MILES)
* (DEMAND_PER_PERSON_PER_SUBREGION_MBPS[6]) * (1/5280.0)**2 * (SCALE)**2
        space[x][y][2] = False


def calculate_reception(x_house, y_house, x_node, y_node, isUrban):
    global NODE_MAX_MBPS_OUTPUT, SCALE

    if isUrban:
        NODE_MAX_RANGE_FEET = 1500
    else:
        NODE_MAX_RANGE_FEET = 6547 # 1.24 miles

    NODE_MAX_RANGE = NODE_MAX_RANGE_FEET / SCALE
```

```
        distance  =  ((x_house-x_node)**2  +  (y_house-y_node)**2)**0.5  #
Calculate distance between a point and the node that was just placed
     if distance < NODE_MAX_RANGE: # If the house is within range of the
tower
         return 925 - (NODE_MAX_MBPS_OUTPUT / NODE_MAX_RANGE) * distance #
Return a reception as a function of distance, assuming that as you move
further away from a node your reception decreases linearly such that once
you're at the max range your reception is 0
    else:
        return 0


max_percent_houses_with_demand_met = 0.0 # value we want to maximize

best_nodes = []

for trial in range(TRIALS):

    nodes = []

    for node in range(TOTAL_NODES):
        # "Throw a dart" and that's where a node is
        x_node = random.randint(0,side)
        y_node = random.randint(0,side)
        trial_node = [x_node, y_node]

        nodes.append(trial_node)


            #  Iterate  over  all  points  in  space  and  calculate  how  much
reception each point gets from the node that was just found
        for y_house in range(side):
            for x_house in range(side):
                   space[x][y][0] += calculate_reception(x_house, y_house,
x_node, y_node, space[x][y][2])


    this_trials_points_with_demand_met = 0
    for y_house in range(side):
        for x_house in range(side):
            if space[x_house][y_house][0] > space[x_house][y_house][1]:
                this_trials_points_with_demand_met += 1
```

```
                        this_trials_percent_houses_with_demand_met      =
(this_trials_points_with_demand_met / side**2) * 100
                   if     this_trials_percent_houses_with_demand_met     >
max_percent_houses_with_demand_met:
                                  max_percent_houses_with_demand_met     =
this_trials_percent_houses_with_demand_met
        best_nodes = nodes


print(max_percent_houses_with_demand_met)
print(best_nodes)

print("completed")
```