Executive Summary

To model the growth of high-speed internet over the next ten years for the UK and US, our team created a model using elements of Calculus, Econometrics, Topology and Computer Simulation.

We first used an exponential approximation for a logistic curve to predict the growth of average peak bandwidth over time. We then modeled the price of internet over time using past data from the Open Technology Institute. Furthermore, we increased the robustness of our model by using statistical techniques such as blocking and taking weighted averages.

Our ultimate goal was to create a model that predicted the rate of growth of the price per unit bandwidth without assuming a hard set limit for the carrying capacity, ruling out a strictly logistical curve by assuming an exponential model up to 10 years in the future. In the future, with more data, our model would be more accurate and the carrying capacity would be better approximated.

To model the bandwidth requirements of a household, we developed a three step process, starting by estimating the effects of the COVID-19 pandemic on internet usage. After modifying 2020 internet consumption data to predict 2021 internet consumption, we moved on to a data analysis stage. We employed regression analysis to fill in missing data and took into account factors such as age, income, and occupation to estimate the hours per week spent on various internet activities. This was then converted to a proportion of time spent on the various internet activities based on total possible time for internet consumption.
Once we determined the internet needs for the activities of each member of the household, we conducted a Monte Carlo simulation to simulate different combinations of activities each family member could be doing at the same time. This created a frequency distributions that, once analyzed, uncovered the minimum bandwidth required to meet internet needs. We applied our model to three relatively disparate households and calculated the bandwidths that can satisfy both 90% and 99% of their internet needs.

To model the optimal placement of cellular towers in specific regions, we took advantage of a reverse method of Varanoi Tessellation using the given Thiessen Polygons. We first utilized the Monte Carlo Simulation from the second model to Find the total bandwidth every family in each subregion used, and used this total bandwidth to find the level of speed the cellular tower that covers the particular regional needs.

Overall, we have found that technology has not progressed to the point of the certainty of our future. However, due to the boom in technological advancements brought on by the age of information, we have just begun to realize our limits and take our knowledge of technological processes to the maximum.

Contents

Introduction

Global Assumptions

G-1 **All data is representative of their intended populations.**

- **Justification:** With the limited information and data collection in such a specific field of study, it is hard to find perfectly accurate data.

1 The Cost of Connectivity

1.1 Problem Restatement

We were asked build a model to predict the cost per unit of bandwidth in dollars or pounds per Megabits per second (Mbps) over the next 10 years for consumers in the United States and the United Kingdom.

1.2 Local Assumptions

1-1 **The Graph of the Average Peak Download Speed is continuous over time.**

- **Justification:** The peak download speed, barring any major technological developments, should increase continuously. The technological developments in the data given did not correlate to jumps in the average peak download speed.

1-2 **Price of internet is independent of the peak download speed it is able to provide**

- **Justification:** From 2012 to 2020, the average peak download speed for cities has approximately been multiplied by ten, while average monthly cost of internet has either remained the same or decreased slightly[1], showing that the changes in peak download speed have little to no influence on internet speed.

1-3 **Every city in America can be classified into either a big, small, or medium city based on their population, and each city in each classification can be blocked together for the purpose of calculating the average monthly price of bandwidth.**

- **Justification:** The size of a city has a considerable influence on the monthly cost of internet there, because higher populations mean more internet service providers. Thus, more competition and generally lower prices.

[1] https://m3challenge.siam.org/sites/default/files/uploads/breakdown/TCP_2021_data_FINAL.xlsx

1-4 **We can use the February 2021 conversion factor of 1.4 US Dollars to 1 British Pound to compare the cost per unit bandwidth values between the UK and the US**

- **Justification:** There is no reason to believe that either currency will depreciate significantly in relation to the other over the next ten years.

1.3 Variables

Symbol	Definition	Units
M_a	Average Peak Mbps in the United States	Mbps
M_b	Average Peak Mbps in the United Kingdom	Mbps
r	Growth Factor of exponential and logistic models	...
M	General Average Peak Mbps	Mbps
M_0	Average Peak Mbps at 2010	Mbps
t	Current year	years
K	Carrying Capacity of a Logistic Growth Model	Mbps
P	Predicted Average Price of Internet Service in a given year	\$/£

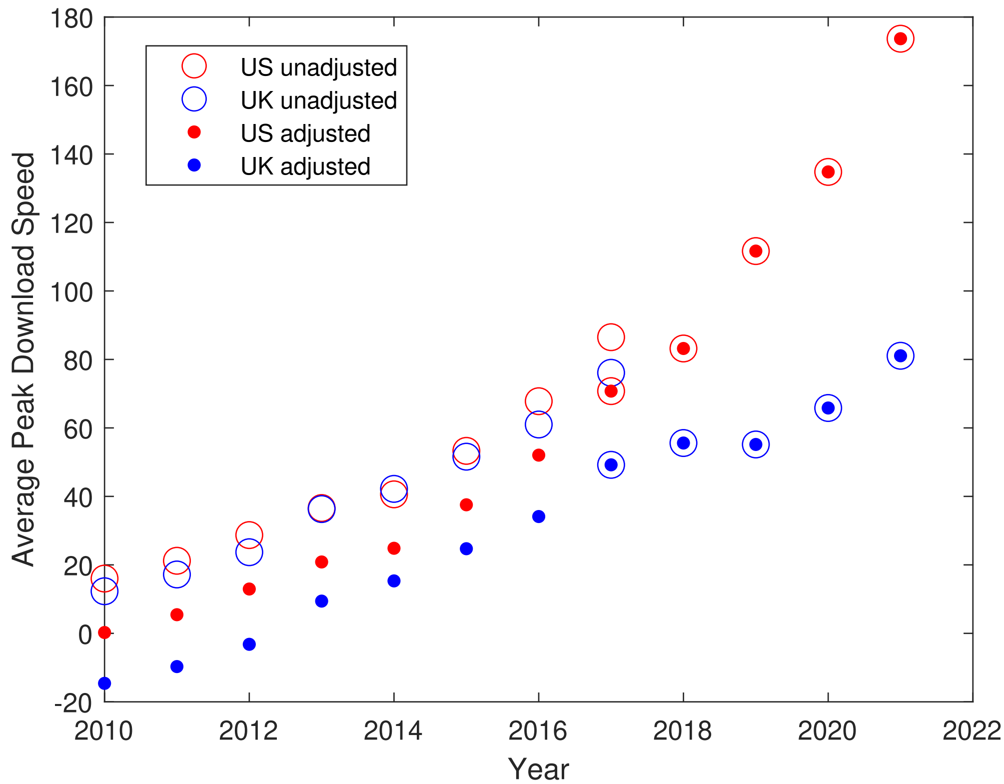
1.4 Our Model

1.4.1 Initial Thoughts

As per Assumption 1-2, price and peak download speed are largely independent from each other, so we decided that it would be best if we modeled each separately, and divided price of internet by peak download speed each year to more accurately model cost per bandwidth.

1.4.2 Getting Moore Bandwidth

We first postulated an exponential growth model, gaining inspiration from Moore's law for transistors, that the number of transistors in a dense integrated circuit doubles every two years. Using the raw data obtained from Ookla and Akamai, we modelled the average download speed for the United States and the United Kingdom from 2010 to 2021. However, there is a discrepancy caused in 2017 due to the difference in data collection methods between Akamai to Ookla, making our graph discontinuous. However, because of one of our initial assumptions that the average peak download speed with respect to time is continuous, we must account for this change. We had two options here: shift the data from Akamai down, or shift the Ookla data up. Because we are trying to predict the growth of bandwidth ten years in the future, we thought it best to shift the Akamai data down for the purposes of calculating the growth trend and preserving the Ookla data, which is more recent.



Notice that some of the data for the United Kingdom is negative; while this might seem out of place at first, we are only trying to look for the general trend in the graphs; the actual values we get will adjust for this negative value in our fitted curve.

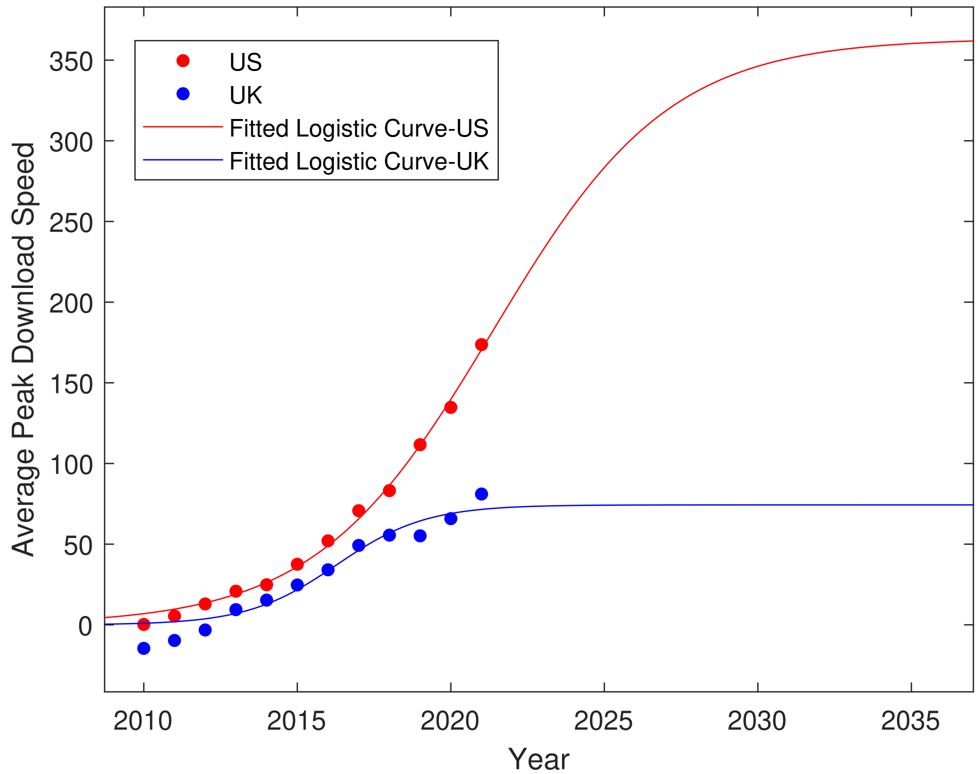
In general, the trends in the graphs do confirm an exponential growth model; the US graph in particular looks to fit the tail end of a standard exponential curve. Moore's law's exponential effect on technology, however, does not necessarily apply to bandwidth. The underlying reason for Moore's law, however, is very similar to our scenario; The physical limitations of transistors, similar to the devices for improving bandwidth, will be taken advantage of with little not no limits. However, what if there was a limit? Before our conclusion could be made, we still needed to consider a different variety of models, the first of which, and possibly the most obvious, being the logistic growth model.

1.4.3 A Brief Foray: Logistic Growth

The next best model we came up with after the exponential model was the logistic growth model. In general, the logistic growth model can be expressed in the form

$$M(t) = \frac{M_0 \cdot K}{M_0 + (K - M_0)e^{-r(t-2010)}},$$

with all variables listed in the table above. Using the MATLAB curve fitting tool on a logistic growth function, our curves indeed fit the logistic growth model almost perfectly. Below are the fitted curves for the United States and the United Kingdom.



The figure represents a plot of the following equations:

$$M_a(t) = \frac{363.4 \cdot 6.9}{6.9 + (363 - 6.9)e^{-0.35(t-2010)}},$$

$$M_b(t) = \frac{74 \cdot 0.98}{0.98 + (74 - 0.98)e^{-0.69(t-2010)}}.$$

With R-Squared values of .996 and .941, respectively, these curves seem to fit very well. However, notice that the carrying capacity for the curves are 363 and 74, respectively. However, according to the data from the average peak download speed for the US and the UK in 2020, there were several locations above these numbers, including 1096 in Lafayette, LA (US) and 104 in London (UK). A carrying capacity implies that there is a certain point where it is impossible for a population or measure (in this case, the network speed) to exceed that point.

The data we have supports the theory that *The carrying capacity is too far off in the future for us to consider a logistic regression model.* So, what is the solution to this problem? We can assume that because the carrying capacity is far away in the future, that the graph, at least until 10 years in the future, will look exponential.

1.4.4 Back for Moore

We have now ruled out a logistic growth model, on the grounds that it is impossible to predict the carrying capacity for the future. We have also postulated that the exponential model is the best possible model to predict the growth of the average peak download speed. Now, we can make an equation in the form

$$M(t) = M_0 e^{rt}.$$

Again, using MATLAB's curve fitting tool, we found the graphs of both the US and the UK, with the equations below:

$$M_a(t) = 3.2 \cdot 10^{-225} e^{0.26(t-2010)},$$

$$M_b(t) = 6.361 \cdot 10^{-165} e^{0.19(t-2010)}.$$

Plugging in $t = 2022$ to $t = 2031$ will get us the prediction of the next 10 years. We are able to make this claim because the future of our logistic growth isn't in sight; in other words, we are the the first half of the logistic growth curve, where the second derivative is positive. With no end in sight, it is clear that an exponential growth model is an adequate substitute for a logistic growth model.

1.4.5 United States Price Model

We are only provided with average monthly costs for over 2 time periods for the United States, and additionally, we are provided with data from a very disparate group of cities for each year instead of a national average. Those factors forced us to take additional measures to improve the validity of our data, instead of simply taking points and creating a Least Squares Regression Line.

We decided to average cities in each size "block" and the take a weighted average of them based on the proportion of the US population that inhabit each "block" to calculate the average cost of internet in the United States in each year[2]. City size is classified based on the following:

Size Classification	Population
Small	$\leq 50,000$
Medium	$50,000 - 500,000$
Large	≥ 500000

and the proportion of the US population that live in each block[3] is given by

Size Classification	2012 Inhabitants	2012 Weighting	2020 Inhabitants	2020 Weighting
Small	78,680,000	39.7%	79,130,000	38.2%
Medium	77,680,000	39.2%	83,040,000	40.0%
Large	41,700,000	21.1%	45,260,000	21.8%

For both years, we determined what size block each city belonged to and then calculated the simple average of the monthly internet price for each block

[2] Assumption 1-3

[3] Calculated with city population data from https://www.census.gov/data/tables/time-series/demo/popest/2010s-total-cities-and-towns.html#tables

Year	City	Classification	Average Price($)	Block Average	Weight
2020	Washington, DC	Large	70.72	62.73	21.8%
	Seattle, WA	Large	67.61		
	San Francisco, CA	Large	60.54		
	New York, NY	Large	64.94		
	Los Angeles, CA	Large	49.83		
	Lafayette, LA	Medium	80.14	77.38	40.0%
	Kansas City, MO	Medium	67.07		
	Kansas City, KS	Medium	71.15		
	Fort Collins, CO	Medium	88.99		
	Cleveland, OH	Medium	55.23		
	Chattanooga, TN	Medium	73.73		
	Atlanta, GA	Medium	105.36		
	Wilson, NC	Small	68.1	54.49	38.2%
	Ammon, ID	Small	40.88		
2012	San Francisco, CA	Large	63.07	63.48	21.1%
	New York, NY	Large	73.18		
	Los Angeles, CA	Large	54.194		
	Lafayette, LA	Medium	78.47	88.46	39.2%
	Chattanooga, TN	Medium	98.44		
	Bristol, VA	Small	100.22	100.22	39.7%

We then took a weighted average within each year, to obtain an average cost measure that can be generalized to the US as a whole.

We obtained a final average monthly cost of 88.86$ for 2012, and 65.44$ for 2020. Fitting these two points to a line, we obtain

$$P = -2.8\$ * (t - 2012) + 87.86\$$$

1.5 United Kingdom Price Model

The model for the United Kingdom is much simpler than that of the United states, as we have 23 data points to work with, which is more than sufficient to create a least squares regression line. We calculated price to be based on the equation:

$$P = -1.32£ * (t - 2014) + 33.07£$$

and used it to project the average monthly cost of internet over the next ten years. The P here is calculated in British pounds, it can be converted to US Dollars by multiplying the value by 1.4[4].

1.6 Results

Based on our model for the Average Peak Download Speed (APDS) and the Price per megabit (PPM), we came up with a table for finding the cost per unit of bandwidth in Dollars per

[4] Assumption 1-4

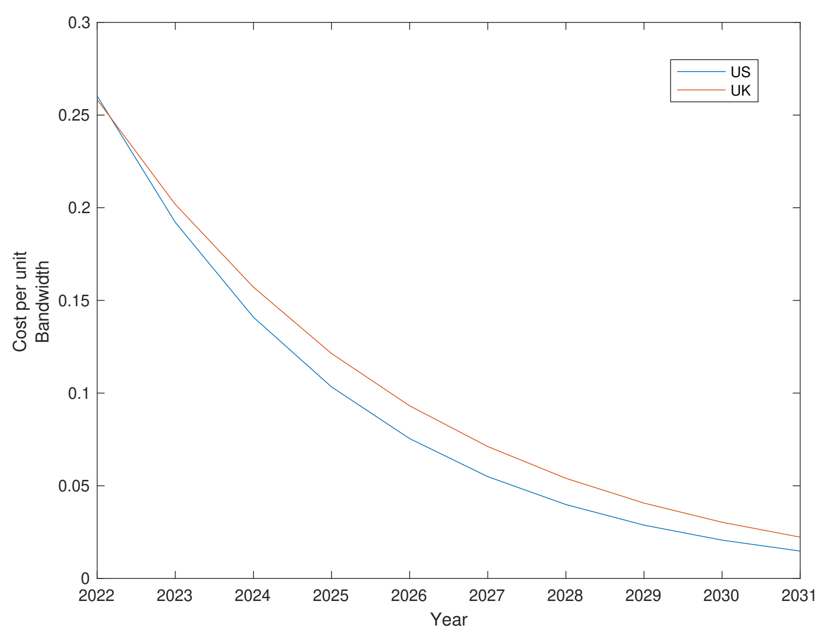
Megabits per second.

year	US APDS	US Price	US Price/APDS	UK APDS	UK Price	UK Price/APDS
2022	230	59.86	0.26	122	31.51	0.26
2023	297	57.06	0.19	147	29.67	0.20
2024	385	54.26	0.14	177	27.82	0.16
2025	498	51.46	0.10	214	25.97	0.12
2026	645	48.66	0.075	259	24.12	0.093
2027	835	45.86	0.055	313	22.27	0.071
2028	1080	43.06	0.040	378	20.43	0.054
2029	1400	40.26	0.029	457	18.58	0.041
2030	1813	37.46	0.021	552	16.73	0.030
2031	2348	34.66	0.015	668	14.88	0.022

We converted all pounds sterling to dollars for the sake of consistency for our final value. Overall, our data seems to show a general exponential downward trend in the Price/ADPS from 2022-2031, which is validated by the fact that our denominator increases exponentially while our price decreases linearly.

1.7 Discussion and Analysis

Our final results were as we expected: the cost per unit bandwidth decreased exponentially over time for both countries, as can be seen below:



The cost per unit bandwidth for the two countries seem to overlap in 2022, but every year thereafter, the cost per unit bandwidth for the US is lower than that of the UK. This can be explained away by the fact that technological development in the UK is slower than that in the US. This is confirmed by the fact that 4G was introduced to the UK 2 years after it was introduced to the US, and the fact that the average peak download speed in the UK was consistently lower than that in the US[5]. Our results, along with data in the spreadsheet, demonstrate that internet providers are experiencing an "economy of scale" of sorts in terms of providing broadband internet to their customers. We have already established that cost per unit bandwidth decreases over time, as peak bandwidth increases, furthermore, from calculating cost per bandwidth of various data plans in 2014[6]

Download Speed	Cost	$/Mbps
4-6 Mbps	34.99	5.83-8.75
15-20 Mbps	41.95	2.1-2.8
30 Mbps	54.97	1.83
50 Mbps	59.95	1.2
100-150 Mbps	69.99	0.47-0.7

We can see that this trend of cost per unit bandwidth decreasing as maximum bandwidth increases applies in smaller contexts as well. Since the "Average Total Cost" of bandwidth decreases as maximum bandwidth increases, it is likely that market incentives will drive innovations in the field for higher bandwidth for decades to come.

1.7.1 Strengths

1. The method we used to create the model for monthly internet costs over time is very robust and resistant towards cities that are, say, particularly technologically developed or charged exorbitantly high prices for their internet service. This is due to our blocking of cities by their population size, which correlates positively with technological advancement, and us taking a weighted average instead of a simple average, which further minimizes the impact that outlier data points have. Our model should hold true over time for any sample of cities in the United States, although it could be further improved if we were provided data points at more points in time.

1.7.2 Weaknesses

1. We know the graph for average peak bandwidth over time is logistic, but our model uses an exponential curve to approximate its growth over time. Although this method of approximation is efficient an accurate for predicting peak bandwidth growth for the next couple decades, as we are a long way from the bandwidth "carrying capacity" [7], we cannot expect the exponential model to hold up over long periods of time. Though when we get to that point, there will likely be much more data upon which we can build a new, more accurate model.

2. The model for predicting monthly internet price over time in the US is based on a very limited number of data points. Although we used blocking and weighted averages to

[5]https://m3challenge.siam.org/sites/default/files/uploads/breakdown/TCP_2021_data_FINAL.xlsx
[6]Ibid
[7]Refer to section 1.4.3

limit the influence of outlier cities, we have no way of combating the effects of outlier years, as we only have 2 in the sample. The validity of our model would be greatly compromised if any major events happened in either 2012 or 2020 to make the data points outliers, such as a major tech subsidy. Thankfully, we've come across no sign of such events occurring in our research, which means our model for cost of internet should be relatively accurate.

2 Bit by Bit

2.1 Problem Restatement

Different households have different levels of internet usage depending on several variables. Given a household, we were tasked with coming up with a mathematical model that predicted the bandwidth that would satisfy 90% and 99% of that household's needs. We were then asked to apply our model to three fictional but realistic households.

2.2 Local Assumptions

2-1 **As a result of COVID-19, internet activities involving streaming increased by about 16% and those involving browsing the internet increased by around 27% since the beginning of 2020**

- **Justification:** Due to the lockdown, people were required to stay indoors to stay safe from the virus. This resulted in more free time for people to spend on the internet as many outdoor activities became closed. This increase in usage was quantified using data from 2020 by the New York Times. [8]

2-2 **Age more significantly affects hours per week spent using the internet (*HPW*) compared to income and other factors**

- **Justification:** Generally, 70+ year old retired elders use most of their internet consumption on watching TV. And so even if their income is changed significantly, seniors in general will still spend close to all of their time watching television. It's even more apparent in the data provided that, relative to the amount of fluctuation in HPW from variation in income, there is much more fluctuation in HPW from variation in age. COVID-19 also has negligible effects on HPW compared to age, as COVID-19 is an external factor that one can control to a degree how much it affects their various internet usages, whereas age is an internal factor that biologically affects your various desires and natural internet consumption habits.

2-3 **After finding the hours per week spent on each online activity, that number can simply be scaled to account for changes in income**

[8] https://www.nytimes.com/interactive/2020/04/07/technology/coronavirus-internet-use.html

- **Justification:** Mainly due to lack of data, we could only assume that the income and age of an individual have no correlation. Obviously there is some slight fluctuations, but they should be relatively negligible to not dramatically influence our model's results.

2-4 **Since the onset of the pandemic, both teachers and students are spending 25 more hours hours per week using internet on their computers(not including video)**

- **Justification:** Since campuses have gone digital, most students and teachers attend school through video conferences, increasing their consumption of the internet by 5 hours per day (time spent on Zoom for school per day) for 5 days a week, resulting in a 25 hour increase in computer usage per week. And according to information from Zoom itself, the bandwidth requirement of a Zoom meeting is around 1.2 Mbps, which is close to the 1 Mbps bandwidth requirement for non-video internet and "general web surfing" that is provided to us in the given data of the problem.

2-5 **The number of hours spent awake in a week is 112**

- **Justification:** Assuming everyone gets the recommended 8 hours of sleep per night, that leaves 16 hours of awake time per day for a total of 112 hours of awake time in a week. As 8 hours of sleep per night is an extremely common standard and universally agreed upon value for a healthy life, we can safely make this assumption.

2.3 Variables

Terms and Variables	Definition	Units
I	Income of a household	\$ per year
Y	Age of an individual	Years
HPW	Hours per week	Hours per week
ISF	Income Scaling Factor (see assumption 2-2)	...

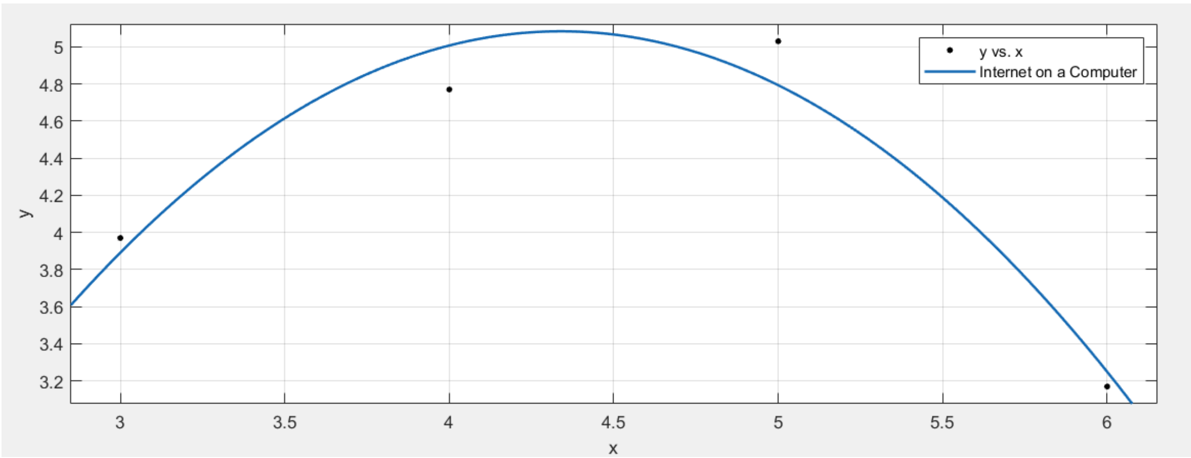
2.4 Data Collection

2.4.1 Overview

We let HPW be units of hours per week. The raw data given to us needed to be organized in a usable manner. For example, when determining the HPW watching traditional television, we wouldn't know which one of age or income should be the predominant factor. Using age, one might estimate 50 hours a week on traditional television but using income, one might estimate 20 hours a week on traditional television. We decided to value age more. However, we didn't forget about other factors, such as income and COVID-19 impact when estimating the hours spent per week on various online activities.

2.4.2 Missing Data

Starting with the data with age and HPW we needed to fill in the missing data for various internet activities in the age groups 2-11 and 12-17. To do that, we performed a quadratic regression based on the 4 other age groups that had data for each of the missing categories. A graph of the quadratic regression for one of the missing categories is shown below:



The dotted line shows the quadratic approximation. All R-squared values for all the approximations made lie between .9332 and .9922, so we assume that a quadratic regression fits the particular data set well.

2.4.3 Income

As can be seen in the data of income vs. HPW, income has a non-negligible effect on HPW. We decided to scale the values in the table of age and HPW based on household income and particular type of online activity (ISF). We were given the percent of households in each income category, so we took a weighted average across all income levels for each online activity. So for a particular activity:

$$\begin{aligned}\text{Average } HPW = .17 \cdot HPW(I < \$25,000) + .25 \cdot HPW(\$25,000 < I < \$50,000) + \\ .21 \cdot HPW(\$50,000 < I < \$75,000) + .37 \cdot HPW(I > \$75,000)\end{aligned}$$

Then, for each type of activity and income bracket,

$$ISF = \frac{HPW(I, \text{Activity})}{\text{Average } HPW(\text{Activity})}$$

And then matching the activities in the table of income and HPW with the activities in the table of age and HPW, depending on household income and type of activity, we multiply the corresponding HPW value for that particular activity in the age and HPW table by the ISF. The income scaling factor will account for income's effect on the HPW. Note that ISF is dependent on the household income and *type of activity*.

2.4.4 COVID-19

COVID-19 is still prevalent in today's society unfortunately, with many classes moved online and forcing teachers to teach from home using Zoom. We decided to account for this increased use of internet as a result of COVID-19. Streaming service consumption increased by 16% and web browsing increased by about 27% as a result of COVID-19 (Assumption 2-1). All of the categories relating to streaming service consumption, e.g. a TV-connected internet device or a video on a computer, are multiplied by 1.16 and all of the categories relating to web browsing, e.g. internet on a computer that is not watching video, are multiplied by 1.27. This scaling of the categories allows for us to account for changes in terms of internet usage due to the increase in free time for internet users during lockdown.

2.5 Model

In our model, we had 2 main inputs that affected how much **bandwidth** a single person would use up: age and income. The former is used to create the baseline HPW for each category and the latter is used to determine the level of quality for various activities such as streaming resolution. For example, if a person has an income over $75,000, then they would most likely stream videos in high definition due to not having to worry about the costs. Likewise, if a person has an income less than $25,000, then they would most accept lower quality videos in order to save money. Furthermore, income is also used to determine the scale factor for each category after age is used to create the baselines.

After finding the HPW of internet usage for a person, we used a Monte Carlo simulation in order to determine the minimum amount of bandwidth to meet their internet demands. First, we must convert HPW to a probability vector. By using Assumption 2-5, we can simply divide HPW by the total amount of hours an individual could use the internet (112 hours). Finally, start our simulation by considering the activities of each member of the household at a certain time. We randomly selected activities based on the HPW probability vector for each person, then the amount of bandwidth required for each person was summed together resulting in the following equation defining the total bandwidth used at one time:

$$\text{Required Bandwidth} = \sum_{i=1}^{n}(Bandwidth(\text{activity, age, income, } HPW)).$$

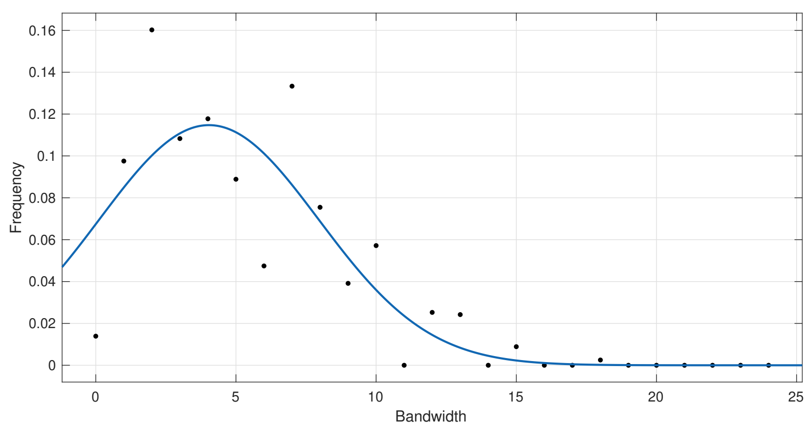
Bandwidth describes the bandwidths required for each member of the family and a function of their activity, age, income, and HPW as previously described. This summation yields the bandwidth requirement at that moment for the family of size n. We then simulate all the possible combinations of activities of the members of the household using the repeating the process for 24 hours a day, 365 days a year, and for 1000 years to ensure a large enough sample. We compile all these required bandwidths to produce a distribution. We took the floor of the required bandwidths in order for the results to be compiled more efficiently.

The resulting data can be approximated by a Gaussian Distribution, which models the true internet usage of the household. We can then find the bandwidth for which 90% and and 99% of the distribution lies below. First, we cut off the part of the distribution with a negative bandwidth as the curve is only meant to model positive bandwidth usage. By using the normal cumulative density function for tail area (10% or 1% of the area), we can then calculate the remaining area. By finally taking the inverse normal function, a cutoff for the minimum amount of required bandwidth that would cover their total internet needs 90% and 99% of the time can be created.

2.6 Results

2.6.1 Family 1

The first scenario is made up of a family consisting of a couple in their lower 30s and a 3 year old child that has a total income equal to that of a teacher (the dad is unemployed so is not making money). Furthermore, since one member of the couple is a teacher, one category of internet activity is increased by 25 hours per week. The other two members of the family will be simulated through categories that match their age and family income. Their distribution is simulated to be as followed:

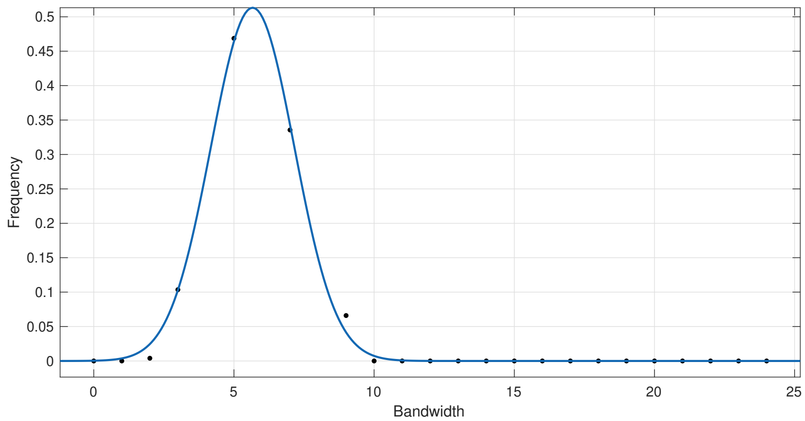


The final calculated minimum amount of required bandwidth that would cover their total internet needs 90% of the time is: **11.94 Mbps**. The minimum amount of required bandwidth that would cover their total internet needs 99% of the time is: **17.46 Mbps**.

2.6.2 Family 2

The second scenario involves a family consisting of a retired lady in her 70s and two school-aged children that will only be there for two days out of the week. The income of the household is around $22,172 a year, the amount given by a federal government pension. [9]. Furthermore, since the two school aged children are only there for two days out of the week, their probabilities for using the internet at an arbitrary time t, will be multiplied by $\frac{2}{7}$. The result of this is the following distribution:

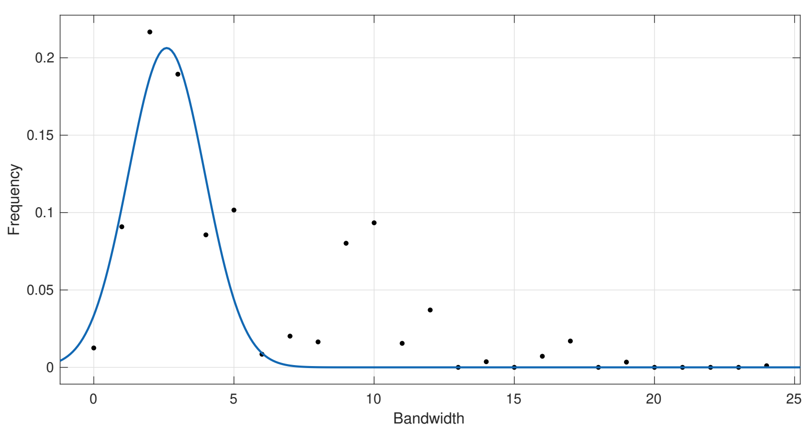
[9]https://www.personalcapital.com/blog/retirement-planning/average-retirement-income/#: :text=The%20difference%20is%20even%20more,that%20these%20are%20national%20averages



Due to excess noise, we smoothed the curve to more accurately fit a Gaussian Distribution by removing the abnormally low frequencies occurring at 4, 6, and 8 Mbps, which were most likely caused by the discreteness of the possible internet usage per activity. The results showed for 90% coverage, the required bandwidth is **8.37 Mbps**, and for 99% coverage, the required bandwidth is **10.56 Mbps**.

2.6.3 Family 3

The third and final scenario is made up of three college students who work part-time, trying their best to not end up with student loan debt by the end of college. As a result of their part-time jobs, their total income will be about $110,472, since the average income for a college part time student is $36,824. [10] The rest of their internet usage activity is standard for their age and will be based on their income. Their simulated distribution is as follows:



The internet bandwidth requirement for this "family" is **5.17 Mbps** for 90% coverage and **7.15 Mbps** for 99% coverage.

[10] https://www.ziprecruiter.com/Salaries/Part-Time-College-Student-Salary#: :text=As%20of%20Feb%2021%2C%202021,

2.7 Discussion

2.7.1 Strengths

Our model takes into account multiple factors including age, income, and occupation and combined them combined using analysis of ratios and behaviors to evaluate the activity of an individual. However, we were unable to immediately calculate the activity because the data provided was incomplete. We used a combination of linear regression models and quadratic regression models, resulting in high coefficients of determination. In addition, we were also able to account for global changes in caused by the pandemic to predict usage in 2021, which was again not provided. Scouring the internet for appropriate data to help predict trends resulted in our final data set.

We were also able to use a Monte Carlo simulation to take into account the randomness of activity of each member of the households. We thus accurately assessed over a million scenarios of activity combinations for the household and compiled results to achieve our final Gaussian Distribution model. The results for the first to families is comparable to real life statistics, as roughly 10-25 Mbps is required for a modern family (Forbes).

2.7.2 Weaknesses

Although the Monte Carlo Simulation should have resulted in roughly smooth curves, a combination of time restraint and sets of discrete measurements of bandwidths resulted in random fluctuations throughout all models. This resulted in not as accurate Gaussian Distribution curve fittings. As seen in the third family's result, the curve probably could have been better fitted with better data.

3 Mobilizing Mobile

3.1 Problem Restatement

We were asked to find the optimal plan for distributing and placing cellular nodes given different regions with certain demographics.

3.2 Local Assumptions

3-1 **The population of each subsection is uniformly distributed according to the corresponding probability density function.**

- **Justification:** The data we are given for subsections only tell us the median household income and the median age of a particular region; because of this, it is impossible to determine a distribution of population better than this in an effective manner.

3-2 **Cost per performance is constant for all three cell towers**

– **Justification:** We define performance to be a measure of both the radius and speed. Because the radius of influence and the speed are inversely related, we can determine that the cost of the three cell towers are relatively similar; that is, buying one isn't more cost efficient than buying another.

3-3 **The needs of a certain demographic conforms to its mean demand for Mbps, specifically 90% of their demand.**

– **Justification:** Because of the cost of cellular towers, the best option for a given population would be to conform to the median of the population. While this will not satisfy every person in the region 100% of the time, it is an efficient way to optimize the cost and performance of a configuration.

3.3 Variables

Symbol	Definition	Units
θ	angle made in the circle with the center and with one of the sides of a sub region as the inscribed chord.	...
x	Length of the side of the sub region we consider as an inscribed chord	miles
r	Radius of the circle, depending on the range of the cell tower	miles

3.4 Model

We first determined a way to find the distribution of the population for each sub-region; Our model for predicting the minimum amount of required bandwidth came in handy for this. We used 90% to conform to realistic expectations; being satisfied 99% of the time is unreasonable for the purposes of mass optimization.
Once we generated the required total bandwidth for each sub region, we categorized the bandwidth into three different requirements for cell towers, shown in the table below:

Band	Download Speed	Range
Low	30-250 Mbps	10-20 miles
Medium	100-900 Mbps	2-3 miles
High	1000-3000 Mbps	0.5-1 miles

We will always take the pessimistic estimate for ambiguous values, because requirements should skew to the stronger side in these situations.
The result of each sub region is in the table below.

#	1	2	3	4	5	6	7
A	424	1145	1046	579	1248	1146	...
B	1418	1345	978	721	2112	1357	1254
C	2490	2399	1717	2197	1934	1710	3035

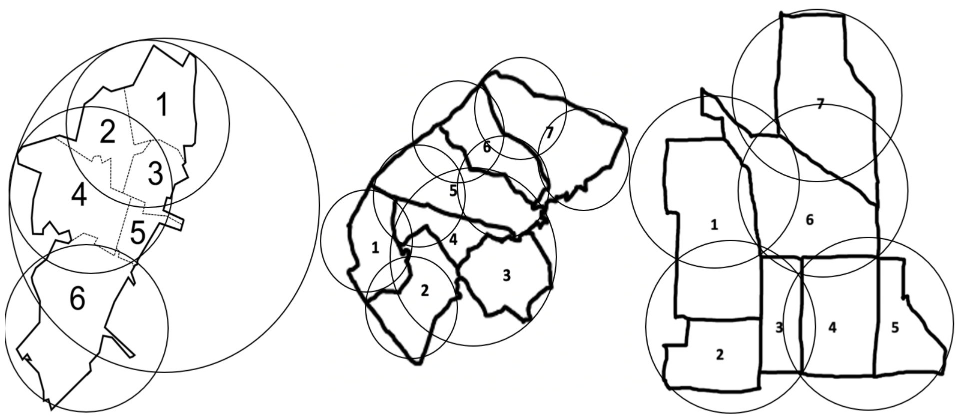
Overall, these numbers make sense; from the data we used with the respective regions, we determined that regions B and C are the most wealthy, with a median household income of more than 118000 and 102000, respectivly, while region A had a median household income of only 49670. The general trend was a medium to high speed connections, a reasonable result due to the changing technological times.
After our initial results, we applied a floodfill algorithm to combine adjacent regions with the same classification.
Then, we used arclengths of a circle with the radius depending on the classification, using the inscribed chord for reference, using the generic formula

$$\theta = \max(0, 2\arcsin(\frac{x}{2r})),$$

with the variables being defined above. We were inspired to come up with circles as our reference shapes through the reverse engineering of Voronoi Tessellation, which takes various points as centers of a circle and fills a region by continually increasing the radius. While this implementation is not the optimal solution it is very effective given the time constraints and resembles an optimization method in Voronoi Tessellation, which is the best possible optimization.

3.5 Results

As a result of our circles, these are the optimal cell towers that are needed in each region:



Region A Region B Region C

3.6 Discussion

3.6.1 Strengths

The main strength of the model is that we used a method of optimization akin to a real optimization method in Varanoi Tessellation. Our model is based on a firm foundation of topology that we defined using our standard measures of distance, which should make our model relatively accurate.

Additionally, we relied on a tested model in our model for predicting demand for bandwidth to determine our data points, which is a promising sign for the accuracy of it.

3.6.2 Weaknesses

The optimization of Varanoi points using the Thiessen Polygons only works with nodes of odd degree, something that is not necessarily true with our model. The most glaring weakness is that this is not at all an optimized method in terms of the circles, because the radii of all our circles are fixed.

We also could have made our definition of the metric in which we define the distance away from two points better; the Euclidean metric assumes no bounds for subregions and no bounds on the radii of the circles, something that we do have.

4 Appendix

```
%clear all;

NActs = 11;
Npeep = 1;
NumYears = 1;
NumDays = 365;
NumHours = 24;

pActs=zeros(1,NActs+1);  %time distribution person 1 in row 1, 2 in row 2, etc., last column is probability not using internet
bwData=zeros(1,NActs+1);  %bandwidth requirement for each activity, last one is 0

pActs(:,1:NActs)=regionA1A2';
bwData(1:NActs)=bwA1A2';
%bwData(NActs+1)=0;

% convert pActs to cdf
for i=2:NActs
    pActs(:,i)=pActs(:,i)+pActs(:,i-1);
end
pActs(:,NActs+1)= 1;

bandwidth=zeros(1,25);
currentAct=zeros(1,1);

for simYearID=1:NumYears           % number of years simulated
    if floor(simYearID/10000)*10000==simYearID
        simYearID
    end

    for simDayID=1:NumDays
%         if floor(simDayID/10)*10==simDayID
%             simDayID
%         end

        for hourID=1:NumHours            % Simulate day
            % Simulate choices in each hour
            r=rand(1,Npeep);
            findAct1=find(pActs(1,:)<r(1));
```


```
bwData(1:NActs)=bwA1A2';
%bwData(NActs+1)=0;

% convert pActs to cdf
for i=2:NActs
    pActs(:,i)=pActs(:,i)+pActs(:,i-1);
end
pActs(:,NActs+1)= 1;

bandwidth=zeros(1,25);
currentAct=zeros(1,1);

for simYearID=1:NumYears           % number of years simulated
    if floor(simYearID/10000)*10000==simYearID
        simYearID
    end

    for simDayID=1:NumDays
%         if floor(simDayID/10)*10==simDayID
%             simDayID
%         end

        for hourID=1:NumHours            % Simulate day
            % Simulate choices in each hour
            r=rand(1,Npeep);
            findAct1=find(pActs(1,:)<r(1));
```


```
            currentAct(1) = 1 + size(findAct1,2);
            bandwidth(floor(sum(bwData(currentAct(1,:))))+1)=bandwidth(floor(sum(bwData(currentAct(1,:))))+1)+1;
        end
    end




end
bandwidth=bandwidth/8760000
fig=plot(0:24,bandwidth);
```