PREVIEW PAPER: EXCELLENT

It was noted that the quality of the paper is relatively consistent throughout. There was no one aspect of the paper that stood out, though.

With respect to the first question, the team provided some justification for their choices. It is relatively straightforward to follow their process and determine how to construct their model.

For question two, it would be more difficult to replicate the results. Also, the complexity of their model does not easily lend itself to be generalized to another situation.

For question three, it is difficult to determine the assumptions that lead to the results. On the positive side, the results are well presented. Additionally, this was one of the few teams that tried to address the optimality of the placement using a structured approach. They were not able to make good use of their proposed approach, but they recognized the issue and made a good attempt to search for a solution to the problem.

Defeating the Digital Divide: Internet Costs, Needs, and Optimal Planning

Executive Summary

High-speed internet is increasingly being seen as less of a commodity and more of a necessity. However, its prohibitive cost has created dramatic inequalities in coverage between demographic groups. The COVID-19 pandemic has only exacerbated the situation by forcing schools, workplaces, and companies to transition to entirely virtual/remote operations. Access to broadband internet should be need-based: large families and families below the poverty line are all examples of families that could benefit the most from high-speed internet access.

However, the problem of internet access is not a trivial one: there exist numerous factors and variables involved. Because technology is such a rapidly expanding field, information about the future of internet technology is difficult to ascertain. However, said data is extremely critical in guiding decisions surrounding internet access.

We first predicted the download speed in the internet over the next decade. By creating an exponential regression, we were able to create a general prediction model in which it takes in the number of years and outputs the predicted price ($ or £) per Mbps. The predictive model showed that the price of internet plan can decrease as low as $0.04 in the US and £0.02 in the UK in 2030.

With the dawn of widespread internet usage, the applications have multiplied; nowadays, the internet is not just merely for entertainment, but for work and education as well. In our second mathematical model, we modeled the amount of time spent on video conference, web surfing, and entertainment/news usage based on the profession of each person. We created Gaussian stochastic behavior to model the inherent "randomness" involved in everyday decision-making. However, this stochastic behavior also took into account productivity and stress trends over the day in order to create a more accurate model. Through our application of the aforementioned models to three sample households of very different demographics, we determined the average Mbps required to sustain a 5 day work week and a 2 day weekend. We then accounted for coverage inefficiency and came up with the 90% and 99% coverage requirements for each scenario.

Lastly, the culmination of our project was creating a model to design cellular networks for a region. Given the region's population density and age distribution, we expressed each cell within a region as a vector, then used the directional derivative to identify the needed bandwidth. We used an exponential regression with distance from cellular tower to calculate a cell's provided bandwidth, then used linear algebra and gradient descent to find the optimal value.

Contents

1 The Cost of Connectivity **3**
1.1 Defining the Problem . . . 3
1.2 Assumptions . . . 3
1.3 Variables Used . . . 3
1.4 Model . . . 4
1.4.1 Average Internet Speed Model . . . 4
1.4.2 Average Price Model . . . 5
1.4.3 Average Price per Mbps Model . . . 6
1.5 Testing the Model . . . 7
1.5.1 Testing the Exponential Relationship Assumption . . . 7
1.5.2 Sensitivity Testing . . . 8
1.6 Model Limitations . . . 9

2 Bit by Bit **10**
2.1 Defining the Problem . . . 10
2.2 Assumptions . . . 10
2.3 Variables Used . . . 11
2.4 Developing the Model . . . 11
2.4.1 Work Function . . . 12
2.4.2 Education Function . . . 12
2.4.3 News and Entertainment . . . 12
2.5 Testing the Model . . . 13

3 Mobilizing Mobile **16**
3.1 Defining the Problem . . . 16
3.2 Assumptions . . . 17
3.3 Variables Used . . . 17
3.4 Model . . . 17
3.5 Testing the Model . . . 18

4 Conclusion **20**

Bibliography **21**

A Code **21**

1 The Cost of Connectivity

1.1 Defining the Problem

Our task is to predict the cost per unit of bandwidth in dollars or pounds per Mbps over the next 10 years (2021–2030) for consumers in the United States and the United Kingdom.

1.2 Assumptions

Assumption 1: The peak speed is the maximum speed a consumer gets below the speeds they pay for.

- **Justification:** While in reality, the peak speed someone gets is likely above the speed that they actually pay for, we need this assumption to better harmonize datasets D1 and D2.

Assumption 2: Internet speeds will increase exponentially over time and the cost will decrease exponentially over time.

- **Justification:** Data compiled from top broadband speeds over the the last decade by the NCTA (The Internet & Television Association) indicates an exponential rate of growth [1]. Specifically, Nielsen's law states: "A high-end user's connection speed grows by 50% a year." Similary, Eldholm's law predicts a doubling in bandwidth every 18 months. The study outlining Nielsen's law found that the law fits data from 1983 to 2019 [2], while a paper on Eldholm's law found that it has held true since 1970 [3]. Thus, it is likely that this established trend will continue for the next 10 years. Furthermore, it is reasonable to assume that a future exponential increase in internet speed and bandwidth would drive an exponential decrease in price. This is supported by empirical evidence, as a study over internet costs from 1995 to 2003 found that an exponential growth in internet correlated with an exponential decline in cost [4].

Assumption 3: The effect of the COVID-19 pandemic on internet costs over the next 10 years is negligible.

- **Justification:** Anthony Fauci, Director of the National Institute of Allergy and Infectious Diseases, predicts a return to "normality" by the end of 2021 [5]. It is reasonable to assume that internet usage – which rose during the pandemic – should return to normal levels by that time. Thus, we would expect a negligible impact on internet costs in the period 2022–2030.

Assumption 4: Those who have access to broadband internet, either have fast connection or super-fast connection.

- **Justification:** Although ultra-fast connection broadband exist, these broadbands were grouped into super-fast connection since data for ultra-fast connection were not given.

1.3 Variables Used

Name	Definition	Units
P	Monthly price of internet plan	$
S	Average peak download speed	Mbps
P(t)	Predicted price of internet plan at time t	$
S(t)	Predicted peak download speed at time t	Mbps

Table 1: Variables used in our model for Level 1

1.4 Model

The average price per Mbps can be predicted by creating a model for average prices and a model for average internet speeds in Mbps. For any given year, the ratio between the predicted average price and the predicted average speed would be the predicted average price per Mbps.

1.4.1 Average Internet Speed Model

D1 consists of data from both Ookla and Akamai, but with Ookla data covering only 2017–2021 and Akamai covering 2010–2017. Because Ookla and Akamai use different methodologies for measuring their average peak download speed (Assumption 1), we came up with a method to combine both for one model. For our model, we set 2010 to be our starting point as year 0. We separated the Akamai and Ookla values for average wired peak download speeds into the following tables:

Year	Average Peak Speeds (Mbps US)	Average Peak Speeds(Mbps UK)
0	16.0	12.3
1	21.2	17.2
2	28.7	23.7
3	36.6	36.3
4	40.6	42.2
5	53.5	51.6
6	67.8	61.0
7	86.5	76.1

Table 2: Akamai-Only Peak Internet Speeds

Year	Average Peak Speeds (Mbps US)	Average Peak Speeds (Mbps US)
7	70.	49.2
8	83.2	55.6
9	111.7	55.2
10	134.8	65.8
11	173.7	81.8

Table 3: Ookla-Only Peak Internet Speeds

After isolating the Akamai and Ookla speeds, we performed exponential regression on the data to get the following equations:

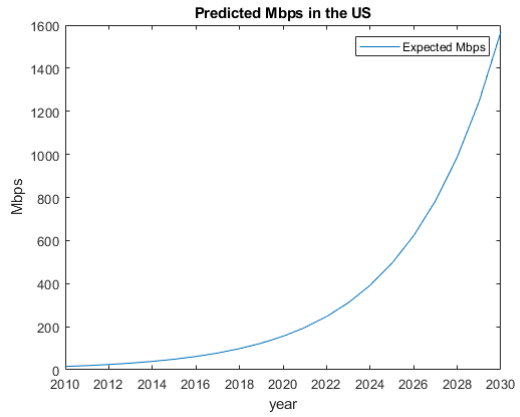
$$S(t)_{AUS} = 16.884e^{0.233x}$$
$$S(t)_{OUS} = 14.0241e^{0.228x}$$
$$S(t)_{AUK} = 13.891e^{0.257x}$$
$$S(t)_{OUK} = 21.139e^{0.117x}$$

We then combined the equations from the Akamai and Ookla exponential regression by averaging both the base and exponent coefficients.

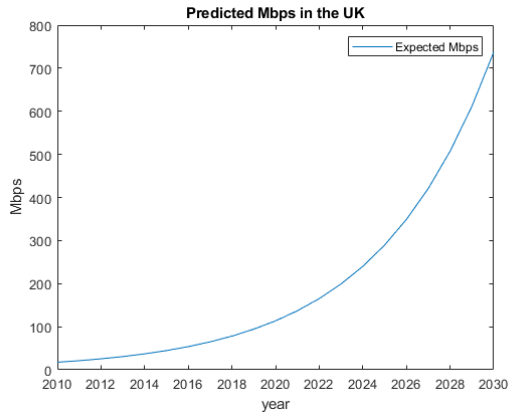
$$\hat{y}_1 = Ae^{Bx}$$
$$\hat{y}_2 = Ce^{Dx}$$
$$\hat{y} = \frac{(A+C)}{2}e^{\frac{(B+D)}{2}x}$$

We applied this basic method to obtain the following models for the US and UK:

$$S(t)_{US} = 15.454e^{0.231x}$$
$$S(t)_{UK} = 17.515e^{0.187x}$$



(a) predicted Mbps in the US



(b) predicted Mbps in the UK

Figure 1: Predicted Average Peak Internet Speeds (Mbps)

1.4.2 Average Price Model

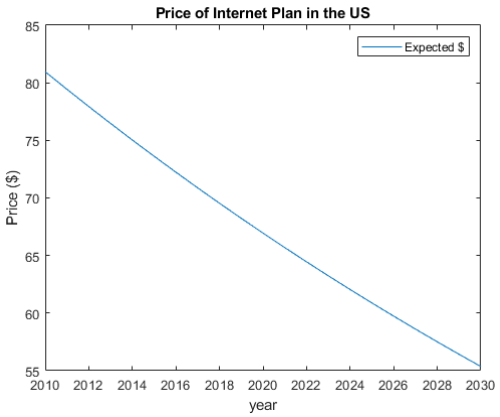
Under the assumption that the price of internet plan decreases exponentially (Assumption 2), using the prices at two different time stamps, the model for predicting the price can be determined using exponential regression. However, since the US and the UK were given different types of data, the price model for the US and the UK were generated separately.

For the US price model, D2 was used to determine the change in the price between two years, 2012 and 2020. Denoting the average price as $\bar{p}$, the average change in the price of internet plan was determined. Using exponential regression, $P(t)_{US}$ for the change in price over 8 years, a model to predict the price at certain year t was found Figure 2a:

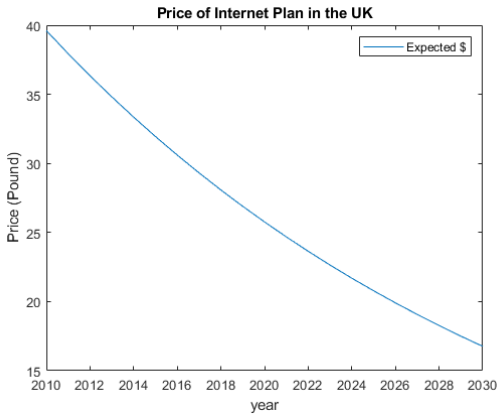
$$P(t)_{US} = 80.954e^{-0.019t}$$

For the UK price model, D3 was used to determine the change in the price among three years, 2017, 2018, and 2019. Along with the assumption that the users in UK either use fast connection, or superfast connection (Assumption 4), the weighted average between the price of fast connection and superfast connection was taken to determine the change between 3 years [6]. Similar to the US model, the exponential regression was taken to determine the predicted internet plan price:

$$P(t)_{UK} = 39.605e^{-0.0435t}$$



(a) predicted price of internet plan in the US



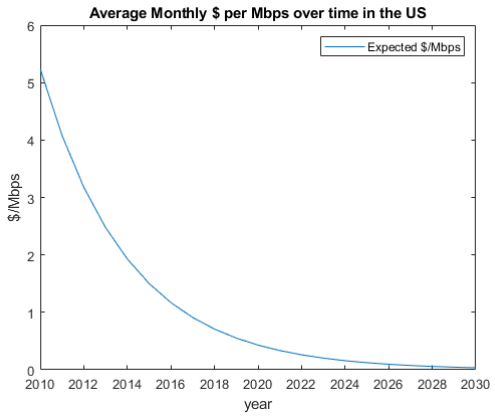
(b) predicted price of internet plan in the UK

Figure 2: Predicted Price of Internet Plan of US and UK

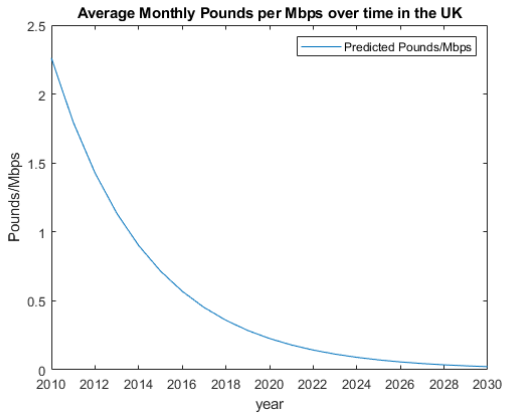
1.4.3 Average Price per Mbps Model

With both the Average Price Model, $P(t)$ and the Average Internet Speed Model $S(t)$, the final model to compare the price per Mbps was developed by dividing the two models as shown in Figure 4.

$$\frac{P(t)_{US}}{S(t)_{US}} = \frac{15.454e^{0.231t}}{80.954e^{-0.019t}}, \frac{P(t)_{UK}}{S(t)_{UK}} = \frac{17.515e^{0.187t}}{39.605e^{-0.0435t}}$$



(a) predicted $/Mbps in the US



(b) predicted $/Mbps in the US

Figure 3: Predicted Price of Internet Plan of US and UK

With our model complete, the following table has the predicted cost per unit of bandwidth per dollars or pounds for the next 10 years in the United States and the United Kingdom.

Year	$/Mbps in the US	£/Mbps in the UK
11	0.337	0.180
12	0.262	0.143
13	0.204	0.114
14	0.159	0.090
15	0.124	0.072
16	0.097	0.057
17	0.075	0.045
18	0.059	0.036
19	0.046	0.029
20	0.036	0.023

Table 4: Predicted Price per Mbps

1.5 Testing the Model

1.5.1 Testing the Exponential Relationship Assumption

To support the validity of our model, we performed linear regression t-tests on the regression models of the average wired internet peak speeds for the US and UK samples from both Amakai and Ookla. We performed the following transformation on all of the exponential regression equations so that the linear regression t-test was applicable:

$$S(t) = \hat{y} = \hat{\alpha}e^{\hat{\beta}t} \longrightarrow \ln(\hat{y}) = \ln(\hat{\alpha}) + \hat{\beta}t$$

We tested the null hypothesis that the true slope $\beta = 0$ against the alternative hypothesis that $\beta \neq 0$. For all samples, we found that:

- The shapes of the scatterplots suggested a linear relationship.
- The lack of fanning in the residuals plots implied that the error around the line was independent of x (homoscedasticity).
- The shape of the histograms of the residuals implied that the error around the line was approximately distributed normally for every value of x.

Thus, the assumptions for the linear regression t-test were met for all samples. We then used the following formula to obtain the t-test statistics:

$$T = \frac{\hat{\beta}-0}{SE(\hat{\beta})},\ SE(\hat{\beta}) = \frac{\sqrt{\frac{\sum(\ln(y)-\ln(\hat{y}))^2}{n-2}}}{\sqrt{\sum(t-\bar{t})^2}}$$

Where SE stands for the standard error function, n the population of the sample, y the observed average peak speed, $\hat{y}$ the predicted average peak speed, t the observed time, and $\bar{t}$ the average of all observed times. For each sample, the resulting test statistic was calculated with a degree of freedom of $n-2$ to obtain the p-value. The resulting p-values were less than 0.01 for all tests. These p-values were significantly low enough to indicate a linear relationship between $\ln(\hat{y})$ and t, which implies an exponential relationship between $\hat{y}$ and t. Thus, the results of the linear regression t-tests fail to reject the usage of an exponential regression to model average internet peak speeds over time. This strengthens our model's assumption of exponential internet speed increase, which in turn strengths the dependent assumption of exponential price decay.

1.5.2 Sensitivity Testing

To assess the potential impact of errors in data or modeling, sensitivity testing was conducted. The testing was conducted exclusively on the US average price model. This model was one of the weakest as it was extrapolated using only two data points. A variance factor was determined by the cumulative distribution function of a random variable X on the standard normal distribution scaled to the range $[0, 2]$:

$$CDF(X) = 2\int_{-\infty}^{X} \frac{1}{\sqrt{2\pi}} e^{-\frac{X^2}{2}}$$

The variance factor was multiplied against the observed average price in 2012 ($77.93). This data point was chosen over the average price in 2020 as it was determined by fewer data points, and was thus more prone to error. Once the 2012 average price was adjusted by the variance factor, it was used to redetermine the US average price model, which in turn, was used to redetermine the US average price per Mbps model. The new 10-year cost predictions were then compared against our original cost predictions by taking the residual sum of squares:

$$RSS = \sum_{i=1}^{10}(y_i - \hat{y}_i)$$

where, for the ith year after 2020, y_i is the new cost prediction after sensitivity testing, and $\hat{y}_i$ is the original cost prediction. This testing was repeated 100 times, each with a new

random variable X. The results with outliers removed is depicted in the scatterplot Figure 4.

The sensitivity testing results indicate that decreases in observed average price generally have a greater impact on model result deviation than increases in observed average price. Like the models themselves, the scatterplot follows an exponential trend.

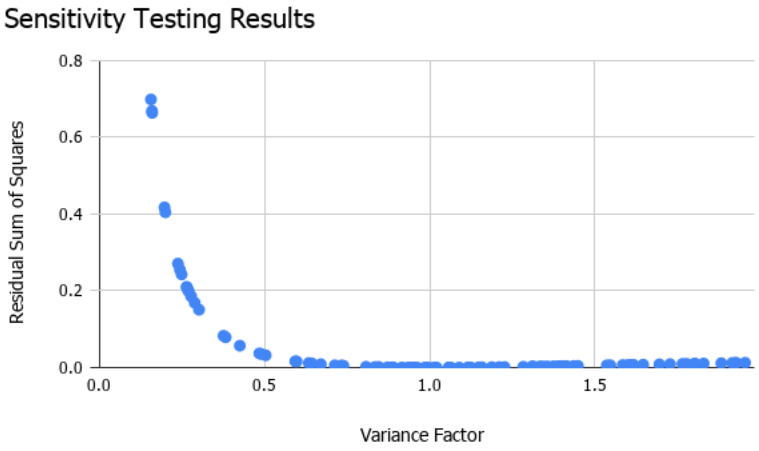


Figure 4: Scatterplot of residual sums of squares against variance factor in sensitivity testing.

1.6 Model Limitations

One of the limitations in our model is that it does not take into account mobile (wireless) download speed. This is because of the very limited data set given on mobile data, and the inconsistent trend with inexplicable factors. Upon evaluation of the peak download speed in the UK between 2015 and 2016, there was a significant drop from 90.9 Mbps to 66.5 Mbps. Also, there was a gap in the data in the year 2017, which was followed up by another significant drop to 25.8 Mbps. Without a justification to these unexpected drops in the data download speed, it was determined that modeling based on this data was unreliable and unfeasible.

Year	Peak Mobile Mbps
2014	34.6
2015	90.9
2016	66.5
2017	–
2018	25.8

Table 5: Average Peak Mobile Download

Other key limitations of this model include the integration of the Ookla and Akamai models into a single average internet speed model. This was particularly troublesome for the UK peak average download speeds, as the exponential regression for the Ookla sample

predicts significantly different prices compared to the Akamai exponential regression by year 2030 ($S(20)_{OUK} \approx 275,\ S(20)_{AUK} \approx 3948$). This suggested that the differing methodologies used by Ookla and Akamai likely could not be accurately reconciled using our naive coefficient averaging.

2 Bit by Bit

2.1 Defining the Problem

Our task is to identify the amount of bandwidth used in a household throughout the day, then identify the 90th percentile and 99th percentile in bandwidth use.

2.2 Assumptions

Assumption 1: People who live together use the internet at the same time.

- **Justification:** The average American household has 10 internet-connected devices, according to a Statista survey [7]. With multiple devices being used at the same time, family members will be able to share the internet concurrently, rather than taking turns.

Assumption 2: During work and school, individuals will do a combination of web surfing and video conferences.

- **Justification:** Workers will always be either working asynchronously or synchronously. While working asynchronously, they use the internet to find resources and produce documents. While working synchronously, they use video conferencing to speak with others.

Assumption 3: Productivity is at a peak in the middle of the morning and the middle of the afternoon. The closer a worker is to these peak work hours, the more productive they will be.

- **Justification:** Individuals are less productive at the start of a work session and at the end because there are more distractions. Typically, an individual "zones in" to their work in the middle of their work time.

Assumption 4: Individuals need more bandwidth for work when they are more productive.

- **Justification:** A completely unproductive person will not complete any bandwidth-demanding tasks, such as creating documents or researching online. As a person becomes more productive, they will do more of these tasks.

Assumption 5: News and entertainment are the main uses of bandwidth when an individual is not working, learning, eating, or watching TV.

- **Justification:** It is common knowledge that people read news, consume entertainment, and browse social media when they are distracted from work or not working.

Assumption 6: Workers have a consistent quota of work that they have to complete in a day.

- **Justification:** Rather than letting workers run out the clock, most business organizational structures are oriented so that workers must complete a combination of tasks throughout the day.

Assumption 7: Entertainment desire follows a sinusoidal pattern.

- **Justification:** Given that each person has a desire for some kind of source of entertainment, the accumulation of that desire will lead to the person taking a break for entertainments. However, after some period of time, the desire for entertainment would decrease. With this trend, the desire will continue to increase and decrease, creating a sinusoidal pattern.

2.3 Variables Used

Name	Definition	Units
t	time	hours
R_w	Percentage of worktime spent on web surfing	%
R_v	Percentage of worktime spent on video conferencing	%
b_w	Bandwidth for web surfing	Mbps
b_v	Bandwidth for video conferencing	Mbps
$\varphi(t)$	Normal distribution	Unitless
$\Phi(t)$	Cumulative Normal distribution	Unitless
$\phi(t)$	Normal distribution	Unitless
α	Skew of each peak	Unitless
D	Daily bandwidth consumption	Mb
K	Annual bandwidth consumption per household	Mb
E	Efficiency	Unitless
$P(t)$	Productivity at time t	%
$E(t)$	Education at time t	Mb
$W(t)$	Work at time t	Mb
$N(t)$	Newpaper and Entertainment at time t	Mb
$S(t)$	Newspaper and Entertainment on no work nor school at time t	Mb
$B(t)$	Total bandwidth usage at time t	Mb

Table 6: Variables used in our model for Level 2

2.4 Developing the Model

We calculate the total amount of bandwidth needed for work, education, and news/entertainment in a household at every minute of the day.

2.4.1 Work Function

For every point of time during the day, the work function outputs the amount of bandwidth used by an individual for work purposes.

We first calculate productivity, a unitless scalar that ranges from 0 (completely unproductive) to 1 (maximally productive). We obtain the productivity from a bimodal Gaussian time plot, where the two centers are the morning peak productivity hour (10 AM) and the afternoon peak productivity hour (2 PM). The function P(t) is defined as follows:

$$P(t) = 2S[\varphi(t-10) * \Phi(\alpha_1 t) + \varphi(t-14) * \Phi(\alpha_2 t)]$$

where lowercase phi is the normal distribution $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}$ and uppercase phi is the cumulative distribution function $\Phi(t) = \int_{-\infty}^{t} \varphi(t)dt$. S is the scale factor (1.3) and alpha 1 and alpha 2 represent the skew of each peak. The resulting bimodal distribution is plotted as follows: We multiply the productivity scalar by the amount of bandwidth used (Mbps) at peak productivity. This was calculated through a weighted sum of video conferencing bandwidth and web surfing bandwidth:

$$W(t) = (r_w b_w + r_v b_v) * P(t)$$

Where rw and rv are the percent of time spent web surfing and video conferencing in a day, respectively, and bw and bv are the bandwidth used by surfing and video conferencing, respectively.

2.4.2 Education Function

Much like with work, bandwidth used for educational purposes varies proportionally with productivity. Productivity is higher near the peak hours of 10 a.m. (morning) and 2 p.m. (afternoon). Furthermore, the ratio of video conferencing to web surfing is different for students, resulting in the equation:

$$E(t) = (r_{2w} b_{2w} + r_{2v} b_{2v}) * P(t)$$

The unique aspect of education is the leftover homework students have to complete after school.

2.4.3 News and Entertainment

We start by calculating the amount of news and entertainment bandwidth used during the work/school day. The amount of bandwidth supplied to news and entertainment is inversely proportional to the individual's productivity (Assumption 5). Therefore, N(t)'s trajectory curve can be defined as:

$$N(t) = (r_{3w} b_{3w} + r_{3v} b_{3v}) * (1 - P(t))$$

However, the amount of bandwidth spent on news and entertainment decreases after many hours of lost productivity, in order for a student or worker to make up for productivity (Assumption 6). This is because workers must fulfill a quota, and the closer they get to the

deadline the more pressured they feel to complete that quota (Assumption 7). For our preliminary testing, we set a quota of 70% productivity, and this is tested in the sensitivity test later on. As a result, they tend to spend more time working and less time on entertainment. This quota-pressure or deadline factor influences the model as follows:

$$d(t) = \frac{\int_0^t V(t)dt}{t+1} - \frac{quota}{W}$$

Where $\frac{quota}{W}$ is the expected productivity per hour. The rationale is that the less productive a worker has been up to the point t when compared to a baseline of the expected productivity, the less likely they are to watch entertainment or news. The error on $N(t)$ is then bounded below by the worst-case volatility function, $V(t) = \frac{d(t)}{|N'(t)}$. It is worth noting, however, that this is not the case for students as a student rarely has any cumulative deliverable due at the end of the day. This lack of quota influences the stochastic behavior of the model. After the work day ends, news/entertainment is the principal use of the internet. There are two interruptions to news/entertainment use: eating dinner, which we place at 8 p.m. EST, and watching television, which peaks at 9 p.m. EST (primetime). Students additionally need to do online homework or a part-time job,which usually happens between 5 and 7 PM. It is worth noting that after the activity is completed, entertainment rises up sharply to pre-activity levels. Finally, for someone who neither works nor goes to school, News and Entertainment are the only factors. This is usually a mix of News and Entertainment and non-Internet based activities. As a result, we can anticipate their usage to follow a simple oscillatory sinoform S(t), defined as the addition of two sine waves. The first sine wave represents their desire for news and the second their desire for entertainment. This function is then:

$$S(t) = (r_w b_w + r_v b_v) * (0.5 + 0.25 * sin(\frac{2}{pi} * t) + 0.25 * cos(\frac{8}{pi} * t))$$

On weekends, entertainment and news is provided by S(t) for all classifications. Total

We calculate each person's bandwidth at each time point as a sum of work, education, news, and entertainment. At a time point t for a family of n people, we calculate the total bandwidth used by summing the work, education, and news/entertainment bandwidth uses from all family members.

$Bandwidth = \sum_{i=1}^{n} W(A_i, t) + E(A_i, t) + N(A_i, t)$

2.5 Testing the Model

Member	Classification	R_v, R_w	Average Mbps(D)
Scenario 1			
Teacher	Work	0.7, 0.2	14.8679
Job Searcher	Work	0.3, 0.5	10.3828
Toddler	Neither	0.4, 0.0	8.3078
Household	–	–	33.5584
Scenario 2			
Retiree	Neither	0.2, 0.2	5.8155
Schoolchild 1	School	0.6, 0.1	12.2009
Schoolchild 2	School	0.6, 0.1	12.2009
Household	–	–	30.2173
Scenario 3			
Student 1	School	0.5, 0.5	13.1222
Student 2	School	0.5, 0.5	13.1222
Student 3	School	0.5, 0.5	13.1222
Household	–	–	39.3666

Table 7: Model testing on the situations provided

D*E then is the rate needed to get E as the proportion of traffic, where E is the efficiency. Furthermore, the speed in Mbps can be 20-50% slower based on distance, d, from the router. Assuming a large distance, the requirements must be increased by 100% to compensate, When 0.9 and 0.99 are tested as values of E, the following results are obtained.

Scenario #	90% Mbps Requirement	99% Mbps Requirement
1	60.4051	66.4456
2	54.3911	59.8303
3	70.8599	77.9459

Table 8: Final results for each of three scenarios.

For sensitivity testing, we created a gradient heat map that represents the required Mbps taking account of R_v and R_w. he heat map was formed from the equation, $f(x, y) = 2.5x + y$ because the Mb consumption of R_v is 1 Mb, while the Mb consumption of R_w is 0.25Mb. As represented in the heat map, the maximum Mbps is required when both video conferencing R_v and web surfing R_w is at its maximum productivity, represented with yellow.

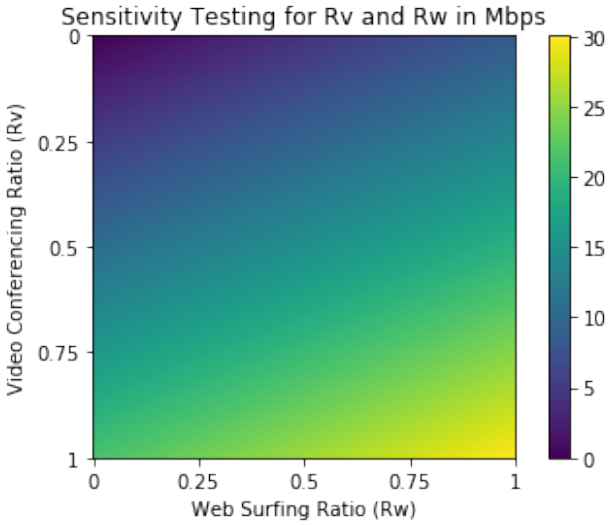
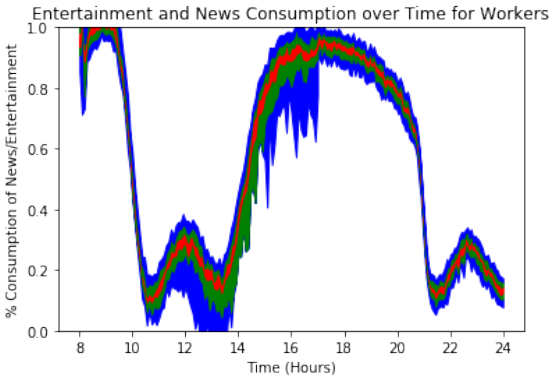


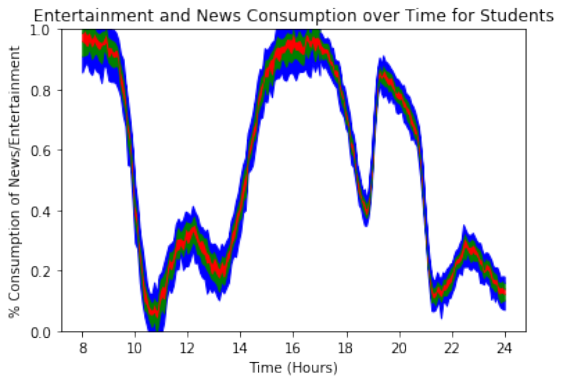
Figure 5: Sensitivity testing heatmap for the effect of video bandwidth and web surfing bandwidth consumption ratios, denoted Rv and Rw respectively, on. Increased Rv had a much stronger effect on higher required Mbps than Rw, consisted with respective bandwidth requirements for the two activities.

To sensitivity check the daily consumption of news and entertainment for workers, stochastic calculus was used to view the probability of different rates of consumption happening throughout the day, red being most probable, green being probable, and blue being unlikely. From the general trend, consumption of entertainment and news is at the peak around 8am and 5pm, and at the trough around 12pm and 10pm. This shows a pattern that most people tend to consume entertainment in the morning, and in the evening where they finish their daily quota. It is also noticeable that generally, the green and the blue areas are below the red, the most probable outcome. With this, it could be seen that it is most likely that the probability of news and entertainment consumption is skewed to the right. Therefore, it is more probable that the workers will spend less than the expected consumption. This is because of the rushing workers who needs to meet the quota – in other words they do not have the time for news and entertainment consumption.

The sensitivity check for daily consumption of news and entertainment for students were similar to workers, in a way that there were spikes in the amount of consumption, in 8am and 5pm. Just like the workers, students also have a trough in the afternoon, due to school obligations and homework. However, since the students don't have a "quota" of work to fill out before the end of the day, the red, green, and blue area evened out normally, showing that it is equally probable that students will spend more or less time on news and entertainment consumption.



(a) Sensitivity testing for news and entertainment consumption and determining the probability of each events happening for workers.



(b) Sensitivity testing for news and entertainment consumption and determining the probability of each events happening for students.

Figure 6: Sensitivity testing for news and entertainment consumption

We also performed best, worst, and average case sensitivity analysis of the aforementioned stochastic models and their effect on required speed in Mbps. Though average case was used in all our findings, we found a significant increase in required Mbps for worst case.

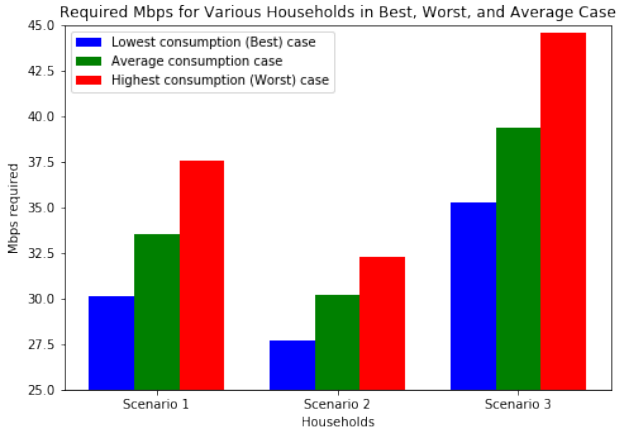


Figure 7: Best, worst, and average case sensitivity testing for stochastic behavior in model.

3 Mobilizing Mobile

3.1 Defining the Problem

Our model's task is to identify the cellular network that optimally distributes bandwidth according to need.

3.2 Assumptions

Assumption 1: As distance from cell site increases, signal speed decreases exponentially.

- **Justification:** Dr. Prashant Krishnamurthy, Chair of the Department of Informatics and Networked Systems at the University of Pittsburgh, claims that loss in signal strength is a logarithmic function of distance [8]. Though Dr. Krishanmurthy does not specify which types of signals he is working with, it is reasonable to assume that the same principle applies to cell signal speeds.

Assumption 2: The optimal cell tower is the one in which the provided bandwidth in each cell is close to the needed bandwidth in each cell.

- **Justification:** The goal of cell companies on the market is to provide for the need of their consumer, nothing more and nothing less.

3.3 Variables Used

Name	Definition	Units
$\vec{cell}$	Vector with cell characteristics	of People
$a\vec{v}g$	Vector with average characteristics	of People
∇B	Vector of partial derivatives for changes in characteristics$t$	of People
d	Distance from cell to closest cell tower	mi.
s	Typical download speed	Mbps

Table 9: Variables used in our model for Level 3

3.4 Model

We start by defining the function $B(ag1, ag2, ag3, ag4, ag5)$, where $ag1$, $ag2$, $ag3$, $ag4$, and $ag5$ are the number of people in the 2-11 age group, 12-17 age group, 18-34 age group, 35-49 age group, 50-64 age group, and 65+ age group.

We then define the gradient vector ∇B. For every component in B, the corresponding component in ∇B is the estimated change in B for a small change in the component, all else held constant. This is also known as a partial derivative.

For each cell in our region, we extract the appropriate amount of mobile bandwidth needed in the region. We start by creating a vector $\vec{u}$, such that for each component $\vec{u}_i$:

$$\vec{u}_i = \vec{cell}_i - a\vec{v}g_i$$

We then normalize $\vec{u}$ by dividing every component by the vector's magnitude.

$$\hat{u} = \frac{\vec{u}}{u}$$

We then take the dot product of $\hat{u}$ and ∇f to get the directional derivative:

$$D_{\vec{u}}B = \nabla B \cdot \hat{u}$$

Finally, we add the directional derivative to the baseline value bandwidth (B_{avg}) to get needed bandwidth (B_{cell})

$$B_{cell} = B_{avg} + D_{\vec{u}}B$$

We begin identifying cell tower locations by randomly choosing 10 cells in which we can place a cell tower. In doing so, we can calculate the closest cell tower to every cell in the map by calculating all of the distances and choosing the minimum one. Then, we define the estimated bandwidth provided to a cell (P_{cell}) using a curve fitted from an exponential regression:

$$s = 2275e^{-0.23d}$$

Where s is the typical download speed in Mbps and d is the distance in miles from the closest tower. This equation was derived from a regression analyses performed on the data from the 2019 Venture Beat Article [9]. Finally, we create a vector V_n with the needed bandwidth for all cells and a vector V_p with the provided bandwidth for all cell. To calculate the optimality score for this map, we calculate the distance between these two vectors, and reciprocate the output so that higher optimality scores reflect better networks.

$$\text{Optimality} = \frac{1}{|V_p - V_n + 0.001|}$$

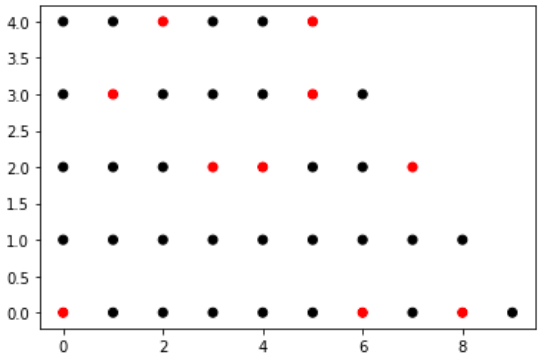
Then, we randomly select one cell tower to move by one unit in a random direction, and recalculate optimality. If the new optimality is lower, then we will revert to the previous map (prior to moving the cell tower). Otherwise, we will keep the change. Regardless of whether the change is made or not, we will then repeat this random selection-and-comparison process with new cell towers at least 10 times. This will ensure we find the local optimum.

After finding and storing the local optimum, we clear the cell map and randomly select 15 cells in which to put cell towers. We then repeat our local optimum identification, and repeat the process of clearing-and-restarting the map 10 times.

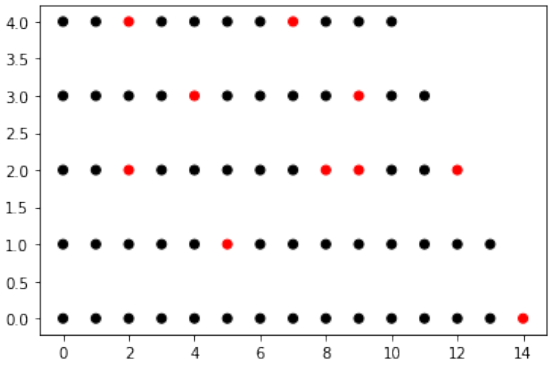
We end with 10 maps and their corresponding optimality scores. We choose the map with the highest optimality score as the "optimal" map.

3.5 Testing the Model

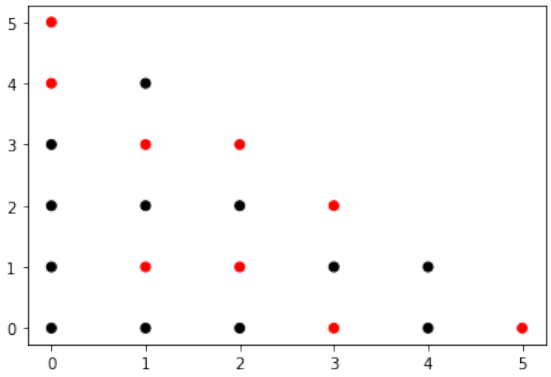
We created 3 regions and divided each region into 40 cells, 65 cells, and 21 cells, each cell being 5 miles in width and 5 miles in length. Using the cell population and cell age distribution, we identified its "needed bandwidth." Using its distance from the closest cell tower, we identified the provided bandwidth. The following figures show our 3 regions in which each red dot denotes a tower in a cell.



(a) Region 1



(b) Region 2



(c) Region 3

Figure 8: Cell phone tower placement by region

In order to sensitivity test our model, we adjusted our exponential model by changing the input points for the regression function. After doing so, we obtained a new equation, and a different configuration of cell towers for region 1. However, it is important to note that even with the same equation, two drastically different cell maps could have similar optimality scores. The new equation was:

$$s = 2344.91e^{-0.25d}$$

Applying this formula to our model, we get the following configuration:

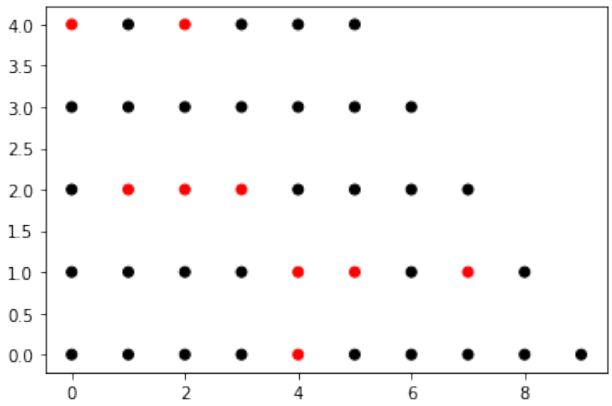


Figure 9: Region 1 Sensitivity Testing

The testing shows that a change in the speed does have a noticeable impact on the configuration of the cell towers.

4 Conclusion

Internet access has allowed the world to function through a global pandemic. In these three problems, our team analyzed download speeds, bandwidth allocation, and cellular network configuration.

In our first model, we used used the the average peak wired internet speed and average monthly costs in the both the United States and United Kingdom to create two models: an internet speed model and a price model. With these two models we are able to predict the cost per bandwidth for the next 10 years. Model weaknesses identified include the lack of wireless internet data and the combination of data from different methodologies.

In our second model, we identified the 90th and 99th percentile in household bandwidth usage. By making use of a clever insight – that productivity is highest close to peak hours – we were able to stochastically model the amount of bandwidth used for work, education, and news/entertainment.Our stochastic model also took into account phenomena such as stress from the pressure to meet a deadline.

In our third model, we were able to determine the optimal internet tower placement for various regions using a vector differential model and gradient descent. This model is extremely flexible/generalized and, as a direct result, can easily be generalized to different regions and demographics. We anticipate this model aiding lawmakers and companies when determining how to maximize broadband internet access, especially to those who need it the most.

With the utilization of our models, we were able to determine the price trend, the minimum bandwidth usage, and the optimal location for internet tower placement. By using these three facts, it is possible to determine how to provide high-speed internet to places where internet access is scarce. With a promising price decline in the internet plan, within a decade, an affordable internet could be achieved.

Bibliography

[1] NCTA. *Superfast Internet*. URL: https://www.ncta.com/positions/the-future-of-superfast-internet.

[2] Jakob Nielsen. *Nielsen's Law of Internet Bandwidth*. URL: https://www.nngroup.com/articles/law-of-bandwidth/.

[3] Steven Cherry. "Eldholm's Law of Bandwidth". In: *IEEE Spectrum* 41 (7), pp. 58–60. DOI: 10.1109/MSPEC.2004.1309810. URL: https://ieeexplore.ieee.org/document/1309810/figures#figures.

[4] Kenneth C. Laudon and Jane P. Laudon. *Management Information Systems: Managing the Digital Firm*. Chap. 4.

[5] Alvin Powell. "Fauci says herd immunity possible by fall, 'normality' by end of 2021". In: *The Harvard Gazette* (). URL: https://news.harvard.edu/gazette/story/2020/12/anthony-fauci-offers-a-timeline-for-ending-covid-19-pandemic/.

[6] *UK Broadband Statistics: Increase Broadband Speed*. Feb. 2021. URL: https://www.increasebroadbandspeed.co.uk/broadband-statistics.

[7] *Average number of connected devices residents have access to in U.S. households in 2020, by device*. URL: https://www.statista.com/statistics/1107206/average-number-of-connected-devices-us-house/.

[8] Prashant Krishnamurthy. *Large Scale Fading and Network Deployment*. URL: http://www.sis.pitt.edu/prashk/inf1072/Fall16/lec5.pdf.

[9] *The definitive guide to 5G low, mid, and high band speeds*. URL: https://venturebeat.com/2019/12/10/the-definitive-guide-to-5g-low-mid-and-high-band-speeds/.

A Code

```
#!/usr/bin/env python
# coding: utf-8

# In[6]:


import numpy as np
import matplotlib.pyplot as plt
import matplotlib
get_ipython().run_line_magic('matplotlib', 'inline')

import math
from random import gauss
import scipy
from scipy.stats import skewnorm
#from skewnorm import stats

import random

import pandas as pd

#Establish a region

xcoords = []
ycoords = []
for y in range(6):
    for x in range(6-y):
        xcoords.append(x)
        ycoords.append(y)

NUM_CELLS = 21-1


#Randomly generate a map of cell towers

list_of_inds = []
labeled_x = []
labeled_y = []
for count in range(10):
    ind = random.randint(0,NUM_CELLS)
    while (ind in list_of_inds):
        ind = random.randint(0,NUM_CELLS)
    labeled_x.append(xcoords[ind])
    labeled_y.append(ycoords[ind])
    list_of_inds.append(ind)
label = [0 for x in range(len(xcoords))]
for l in range(len(label)):
    if l in list_of_inds:
        label[l] = 1
color = ['black','red']
plt.scatter(xcoords,ycoords,c=label,cmap=matplotlib.colors.ListedColormap(color))
```



```
#Randomly generate a map
def createMap():
```

```
    list_of_inds = []
    labeled_x = []
    labeled_y = []
    for count in range(10):
        ind = random.randint(0,NUM_CELLS)
        while (ind in list_of_inds):
            ind = random.randint(0,NUM_CELLS)
        list_of_inds.append(ind)

    return list_of_inds



#Calculate the Mbps provided to each cell

print(len(xcoords))

def provided(xcoords,ycoords,list_of_inds):

    cp = [] #vector of provided bandwidth
    for count in range(len(xcoords)):
        x = xcoords[count]
        y = ycoords[count]
        allDists = []
        for index in list_of_inds:
            try:
                dist = math.sqrt((x-xcoords[index])**2+(y-ycoords[index])**2)
                allDists.append(dist)
            except Exception as e:
                print(e)
                print(index, xcoords, ycoords)
                print(list_of_inds)
                raise e
        lowestDist = min(allDists)
        speed = 2275*math.e**(-0.23*lowestDist)
        cp.append(speed)

    return cp



#Calculate the optimality score

df = pd.read_csv("Regions4.csv")

prov = provided(xcoords,ycoords,list_of_inds) #Vector of provided bandwidths
print(len(prov))
```

```
def optimality(xcoords, ycoords,list_of_inds):

    need = df["Needed Bandwidth"] #Vector of needed bandwidths
    prov = provided(xcoords,ycoords,list_of_inds) #Vector of provided bandwidths
    #Calculate vector difference
    difference = 0.0
    for count in range(len(prov)-1):
        p = float(prov[count]) #Component from vector of provided bandwidth
        try:
            n = need[count] #Component from vector of needed bandwidth
        except Exception as e:
            raise e
        difference += math.sqrt((p-n)**2)

    return(1/(difference+0.001))

#Find local maximum for optimality

def total_max(xcoords,ycoords):

    bestOpt = 0 #Highest optimality score
    bestList = [] #Best list of cell towers

    for a in range(20): #10 map selections
        list_of_inds = createMap() #Create a map randomly
        opt = optimality(xcoords,ycoords,list_of_inds)

        for b in range(20):
            pointMoved = random.randint(1,len(list_of_inds))


            if (pointMoved<10 and list_of_inds[pointMoved]<NUM_CELLS): #Making sure we don't go off the edge of the boa
                list_of_inds[pointMoved] += 1 #Moving the point in a random direction
            if optimality(xcoords,ycoords,list_of_inds) < opt:
                list_of_inds[pointMoved] -=1
            else: #Replace optimum value
                opt = optimality(xcoords,ycoords,list_of_inds)

        if opt > bestOpt:
            bestOpt = opt #Best optimum
            bestList = list_of_inds #Best list

    return list_of_inds

#Print the CONFIGURATION with the maximum optimality
print(total_max(xcoords,ycoords))

#Coloring map
color = ['black','red']
tm = total_max(xcoords,ycoords)
labels = []
for count in range(0,NUM_CELLS+1):
    if count in tm:
        labels.append(1)
    else:
        labels.append(0)
plt.scatter(xcoords,ycoords,c=labels,cmap=matplotlib.colors.ListedColormap(color))
```