Proceedings of The National Conference
On Undergraduate Research (NCUR) 2013
University of Wisconsin La Crosse, WI
April 11 – 13, 2013

Traveling on the O.R Express: An Excursion in Commuter Rail Design

Bryan McComb
Department of Mathematics and Computer Science
Muhlenberg College
2400 Chew Street
Allentown, PA 18104 USA

Faculty Advisor: Dr. Linda McGuire

Abstract

Public debate about providing commuter rail service between eastern Pennsylvania and New Jersey has hinged on factors such as cost, time, predicted usage, and profit. Using mathematical optimization techniques from Operations Research (O.R), together with modern technology like Google Earth, we designed possible rail connections between Easton, PA and existing commuter lines in New Jersey that minimize cost and maximize travel efficiency. Despite claims that the cost would be prohibitive and that train service would not be superior to less-expensive bus service, we will demonstrate that, under certain conditions, it is possible to design rail service which can satisfy both consumer and producer.

Keywords: Railroads, Optimal, Mathematics

1. Introduction:

Since the early 19th century and the first introduction of the railroad, a consistent goal has been to optimize traffic flow and to balance volume while ensuring that systems are cost efficient and profitable. While Europe has a longstanding tradition of on-going railroad usage, this pattern is less well established in the United States for many reasons, for example, local and regional governments struggle with construction decisions.

One issue with railroad optimization is that each system has unique specifications and therefore, each line requires its own specialized analysis. This is not to say, however, that lines lack shared characteristics. Fortunately, there is a long academic history of considering these issues. In a series of papers written by Felix Schmid, et. al., several factors to recognize when one wants to strategically design and manage railway systems are considered[10, 13, 14]. Before a design can be planned, a decision must be made regarding the use of freight and passenger tracks, since each track contains different requirements. Schmid continues to write that proper maintenance and management assessments, as well as human resource management, should be properly addressed within the analysis. Additionally, components such as safety, marketing, and environmental issues should aim to satisfy the conditions set by each individual company. To synthesize these ideas, Schmid, et.al., promote two key qualities:

1. *Dependability*: Combination of acceptable levels of regularity, reliability, punctuality, and safety.

2. *“Good” Train*: Must optimize speed, acceleration, braking, and capacity.

Another primary factor that prevents a unification of optimization techniques in rail design is the variation in costs when addressing urban rail, capital costs, per-kilometer rail costs, and per-station costs[5]. Flyvbjerb states that the total capital costs for per route-kilometer in European projects are between $50-100 million while US projects range between $50-150 million. Reasons for the large variation include: differences between the ratio of underground to above-ground construction, environmental and safety constraints, and labor costs. Pickerell approached this issue from a strictly statistical standpoint. He utilized least-square estimates, point estimates and outlier analysis to compare costs between heavy and light rail project projections[11].

One way of modeling rail systems is to optimize the periodic train schedule[1, 12]. Barber and Salido, et. al., used a combination of topological constraints and software tools like LINGO and ILOG Concert Technology (CPLEX) to optimize their system. This technique, however, requires an already established line and that these tools are readily available for use. Bussieck and Zimmermann address the same issue of scheduling, but approached it in a discrete mathematical sense[4]. When they designed their schedule, they addressed line planning and also determined passenger demand using mathematical models. Although it was not a key feature of their paper, Bussiek and Zimmermann applied some graph theoretical approaches when addressing the flow of a system. Similarly, they have approached the problem using combinatorial methods[3].

This analysis of existing literature led to an interest to modeling rail connections between New Jersey (NJ) and Pennsylvania (PA).When designing and optimizing rail systems, we incorporated many of these methods into our analysis. Furthermore, we synthesized existing ideas with easily accessible modern technologies to design a feasible and cost-efficient line.

2. Hypothesized Design and Reasoning

An article in *International Construction* illustrated how the proposed rail network by the California High-Speed Rail Authority has increased in proposed cost from $36.4 billion in 2010 dollars to $65.4-74.5 billion in 2011 dollars[16]. Our goal was to address this problem by working to implement a commuter rail system that connects eastern Pennsylvania (PA) to New York City (NYC) which would be optimized for time, cost, and ridership.

New Jersey Transit (NJT) is a well-known statewide public-transportation system with stations in eastern NJ, which operates commuter rail service between NJ and NY. One of NJT's most notable routes is between its NJ connections and Penn Station, NYC with an average weekday boarding of 77,058 people[15]. Within NJT, there are two connections of interest with endpoints closest to Pennsylvania: the Gladstone Branch, which terminates at Gladstone, NJ and the Raritan Valley Line, which terminates at High Bridge, NJ. There has been significant discussion in the local media about the desire for installation of an express train into NYC, as well as installation of tracks that leads westward into PA. PA has had a similar discussion about possible railroad extensions that would directly connect PA into NYC at these two points, as the only mode of transportation into NYC is currently either bus or private transportation.

This project sought to design and optimize commuter railroad connections between NJ and PA by creating viable models using mathematical technique from statistics, operations research and graph theory, and available technology such as Google Earth. For this model, our connection in Pennsylvania will be Easton, PA, which, according to the U.S. Census Bureau, has a population of 26,995, and has a growing white-collar community which is seeking a NYC rail connection[9, 18]. For NJ, we will make use of the Gladstone and High Bridge locations due to their proximity to Easton. We will design not only a local connection, but an express line that is a "good" train and will optimize dependability.

To build this model, we implemented the drawing tool in Google Earth to obtain an accurate layout of the landscape and to gather estimates of the distances between rail stations. Lines were broken down by individual stops, and a total distance of the path created by the drawing tool indicated the amount of miles between each stop. An example of the proposed model can be seen in Figure 1, where the yellow and pink lines indicate the existing NJT line, and the orange and purple lines are the created lines that connect Gladstone and High Bridge to Easton. Each created line utilized the topological properties of Google Earth to minimize the environmental damage caused by the imposed system.

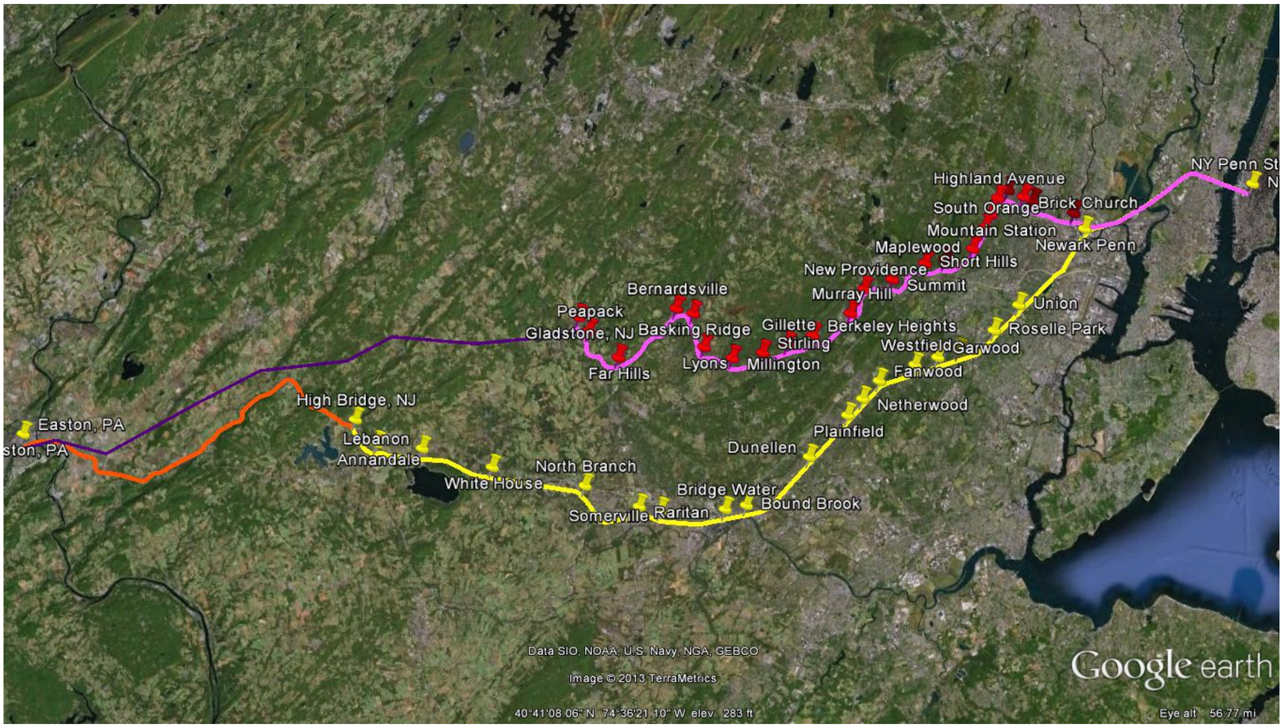

Figure 1- Theoretical transit model which connects New Jersey to Pennsylvania.

3. Key References and Simplifying Assumptions

Most public information available regarding NJT is accessible on their website[19]. For information not found there, upon request NJT would often provide additional data by email. A similar process was used to gather information from Pennsylvania by contacting the Pennsylvania Department of Transportation (PennDOT).

One document that provided crucial information throughout this project was the *Lehigh Valley Surface Transportation Plan 2011-2030* (LVSTP). Within this report, the Lehigh Valley Planning Commission discusses a similar model to the one we designed. They indicate, however, that this line is not feasible because the daily 2030 ridership is predicted to NYC from Easton is 800 riders, plus an affixed 395 from Phillipsburg, Bloomsbury, and Hampton, NJ, which are additional stations between Easton and High Bridge. Further, they predicted the estimated rail time to be 2.1 hours from Easton to High Bridge. Therefore, when designing this new model, such information was used as the lower bounds for the system to satisfy the requirements laid out by the LVSTP.

Simplifying assumptions made at the onset include the following premises. The train speed is constant throughout a trip, except at the endpoints. Also, this design focuses only on commuter rail, and freight information will not be included. When analyzing commuter traffic, we focused on strictly weekday traffic, since the weekend traffic into NYC is highly variable. One also must assume that all of the public information gathered is accurate and reliable.

4. Mathematical Analysis

One of the benefits of using existing rail in our model is the initial stability that it should provide. To test this, we ran two-way ANOVA analyses comparing the average ridership to each station and year. We found, overall, that the ridership throughout NJT has low variability, indicating a stable system. We also ran a one-way ANOVA analysis comparing the variability of average ridership from 2001-2011 on both the High Bridge and Gladstone lines. On both lines, we found that the variability sufficiently minimal so that ridership is stable.

Initial data gathered on miles, minutes, rate, ridership, and monthly cost can be seen in Table 1 and Table 2. The first task was to properly distribute the monthly cost model to extend westward to Easton. Secondly, we then needed to properly model the rates in relation to the express train.

On the Easton to Gladstone Line, we considered the monthly cost from Newark Broad Street and created a linear equation to represent the monthly costs. An ANOVA analysis calculated an F = 62.14, which correlates to a p <.0001, indicated a significant difference between the means. The predicted model, in accordance to LVSTP had an r^2= .7299 and was calculated to be:

Gladstone Monthly Cost = 114.936 + 6.02250*(Total Miles)

We took the values from the linear model and subtracted it from the expected amount, from LVSTP, in Easton, Gladstone, and Peapack, to get a value of 210.37 dollars extra to disperse among the other stops. Then, after running several experimental procedures to optimize fare distribution, we took 5.5% of the value and added it to the expected values from Highland Ave to Far Hills, netting in a total dispersed amount of 196.69 dollars. Additionally, we took the remaining 13.68 dollars, and distributed 33.33% to each stop. When running an ANOVA test on the suggested model, we calculated an F = 181.36 and p <.0001, which indicates an appropriate alteration to the original regression. The predicted monthly cost, after all the changes, had a r^2value of .897 and was modeled as:

Gladstone Adjusted Monthly Cost = 95.128 + 6.77691*(Total Miles)

On the Easton to High Bridge Line, we first ran an ANOVA test and calculated an F = 105.14, and a p < .0001. Furthermore, we calculated r^2= .827 for the LVSTP monthly cost and modeled the equation as:

High Bridge Monthly Cost = 48.735 + 8.505*(Total Miles)

Then, we tried to create two equations to mirror the estimated cost values. We split the equations between NY Penn Station to Raritan and Raritan to Easton since the larger distance values caused the equation to expand beyond acceptable bounds. The ANOVA test calculated an F statistic of 2930.43 and 129.72, respectively, and with a p < .0001, as desired. No distribution of values was necessary, and the equations had a r^2value of .9978 and .869 and modeled as the following:

High Bridge Lower Monthly Cost = 139 + 8.219*(Total Miles)
High Bridge Upper Monthly Cost = 414 + 1.380*(Total Miles)

Table 1- Data table for the Gladstone line.

	Between adjacent stations (first being Easton to Philipsburg)					Grouping Prices Together	add 5.5% from 8 to 24
	Miles	Minutes	Rate (Average mph)	Ridership	Monthly Cost	y = 139 + 8.2192x	Distributed Price
Easton	0	0	0	800	378	673.25	484.56
Gladstone	30.5			173	414	422.56	418.56
Peapack	0.93	3	18.6	49	414	422.56	418.56
Far Hills	2.38	4	35.7	147	408	395.36	406.93
Bernardsville	4.13	6	41.3	186	400	361.41	372.98
Basking Ridge	1.17	3	23.4	99	400	361.41	372.98
Lyons	1.92	3	38.4	435	361	336.01	347.58
Millington	1.64	3	32.8	161	361	336.01	347.58
Stirling	1.6	3	32	97	361	336.01	347.58
Gillette	1.36	3	27.2	153	324	298.21	309.78
Berkeley Heights	1.38	3	27.6	504	308	286.86	298.43
Murray Hill	2.37	4	35.55	549	284	267.38	278.95
New Providence	1.57	3	31.4	563	273	254.48	266.05
Summit	1.68	4	25.2	3565	273	254.48	266.05
Short Hills	2.25	4	33.75	1392	233	222.18	233.75
Millburn	1.16	3	23.2	1687	233	222.18	233.75
Maplewood	1.61	3.5	27.6	3037	208	199.41	210.98
S. Orange	1.3	3.5	22.29	3495	193	188.73	200.3
Mountain Station	0.76	2	22.8	303	193	188.73	200.3
Highland Ave	0.93	2	27.9	231	193	188.73	200.3
Orange	0.71	2	21.3	1150	169	169	169
Brick Church	0.89	3	17.8	1523	169	169	169
E. Orange	0.52	2	15.6	298	169	169	169
Newark Broad Street	2.24	4.5	29.87	2316	139	139	139
NY Penn Station	10.3	22.5	27.47	77058	0	0	0
						r-squared = .99656	r-squared = .897

Table 2- Data table for the High Bridge line.

	Between adjacent stations (first being Easton to Philipsburg)						
	Miles	Minutes	Rate (Average mph)	Ridership	Monthly Cost	y = 139 + 8.2192x	y = 414 + 1.38018x
Easton	0	0	0	800	480*	707.44	457.848
Philipsburg	2.3	3	46	395	472*	688.54	454.674
Bloomsbury/Bethlehem	7.6	9	50.67	395	456*	626.07	444.184
Hampton	9	9	60	395	436*	552.1	431.76
High Bridge	4.9	7	42	73	425	511.82	425
Annandale	1.81	4	27.15	78	425	511.82	425
Lebanon	2.37	4	35.55	28	425	511.82	425
White House	3.79	6	37.9	120	414	446.32	414
North Branch	4.9	7	42	80	414	446.32	414
Raritan	3.47	9	23.1333	622	408	377.52	
Somerville	1.17	3	23.4	651	408	377.52	
Bridgewater	3.34	4.5	44.53	338	386	340.45	
Bound Brook	1.11	3	22.2	620	361	331.33	
Dunellen	4.24	5	50.88	948	324	296.48	
Plainfield	2.95	4	44.25	897	308	272.23	
Netherwood	1.14	5	13.68	534	284	262.86	
Fanwood	1.28	3	25.6	966	273	252.34	
Westfield	2.1	4	31.5	2300	248	235.08	
Garwood	1.2	3	24	101	248	235.08	
Cranford	1.15	3	23	1189	233	215.77	
Roselle Park	2.16	5.5	23.564	864	208	198.01	
Union	1.84	4	27.6	1265	193	182.89	
Newark Penn	5.34	12.5	20.671	26581	139	139	
NY Penn Station	9.96	20.5	29.15	77058	0	0	
						r-squared = .9978	r-squared= .869

According to the LVSTP data, it would take the new train 2.1 hours to travel from Easton to NYC, while it would take the bus 1.7 hours to travel the same distance. Thus, if t is our time in hours, and m is the distance, in miles, we know that the below inequality is satisfied:

$$\sum_0^m t < 1.7.$$

While the data correlates with LVSTP predicted time, we decided that the train system would be more efficient if express trains were installed to run during times of high volume traffic. Therefore, two separate express train systems were created and modeled after NJT: one going from Easton to the Gladstone connection, and the other leading from Easton to the High Bridge line.

Express trains contain fewer stops per trip, and stops are selected for such a route based upon high average passenger ridership. Thus, the High Bridge express line will stop at Easton, High Bridge, Raritan, Westfield, Newark Penn Station, and NY Penn Station, while the Gladstone express will stop at Easton, Gladstone, Summit, Newark Broad Street, and NY Penn Station. According to NJT, the average speed of a train ranges between 60-79 mph[20]. Using this information, we calculated the estimated time of travel of each express train. Since each train takes about a mile to start and stop, we considered the start and stop distances separately from the travel miles, for higher accuracy. For the start and stop miles, we totaled the number miles and divided by half of the total speed traveled for that estimated system.

For example, the total distance of the High Bridge express train is 79.12 miles, and there are four stops (not including the endpoints). We removed two miles from each stop to include in the separate calculation, making it a total of eight miles, while adding on another two miles for the start time from Easton and the stopping time to NYC. Therefore, if the estimated speed of the train is 60 mph, we would calculate 10 hours/30 mph to determine the total time it would take to start and stop the train. In addition to how long it would take to accelerate to 60 mph, it is calculated that it would take 1.485 hours to travel 60 mph from Easton to NYC. If the speed is increased to 79 mph, it would take a total of 1.1281 hours per trip.

Similarly, since the Gladstone express train is 75.3 miles long, and there are three stops, not including the end points, a total of eight miles will be separated from the system for separate calculations. Therefore, if the train is moving at 60 mph, it will take 1.3884 hours to travel, while if the train is moving at 79 mph, it will take 1.0544 hours to finish the trip. Since our predicted times are below the estimates of the LVSTP data, we decided to utilize the express train data when preparing our system.

For transportation and assessment problems, we utilized the simplex method, which is a famous technique in operations research and used specifically for linear programming. We first implemented the simplex method to minimize maintenance costs. Due to limited information regarding maintenance, we decided this would not be an appropriate model. The simplex method, however, was effective for modeling our capacity model. We will define our variables as the following:

x_1 = The number of additional riders from Easton to High Bridge,
x_2 = The number of addition riders from Easton to Gladstone.

Therefore, x_1 and x_2 are positive integers since ridership must only be in whole numbers and can only be positive. From this, we can derive our profit model as:

Maximize $P = 22.893x_1 + 24.228x_2 - 62{,}082.33$
Subject to: $1{,}200 \leq x_1 \leq 14{,}515$ $800 \leq x_2 \leq 9{,}410$ $0 \leq x_1$ $0 \leq x_2$

Both 1,200 and 800 originate from the LVSTP plan where they indicated that ridership could not surpass those two values. Thus, for our system to be successful, the model must at least exceed those values. On the other hand, 14,515 and 9,410 were established from calculating the maximum number of people that could be transported during one work day. This calculation was done both by hand and excel, with a P = \$119,565,069. This profit seems appropriate as it aligns with the amounts NJT usually nets.

We also used advanced variations on the simplex method, such as the network simplex method. This technique is used to solve for the minimal cost flow across a weighted network. We considered the flow through a directed edge $i \rightarrow j$; we use both upper and lower bounds on the basic variables and "+" and "−" signs in the relational equations capture the directed nature of the edges. The algorithm has a "start vertex" and adds edges to the tree that reduce the cost function while adhering to the traversal order of the directed edges until we obtained our optimized model. In Figure 3, the "start vertex" would be Easton, and by expanding a tree outwards from that point, the most optimal solution would be Easton→ Clearwater/Lebanon→High Bridge. By analyzing Figure 4 and utilizing the same "start vertex" at Easton, the most optimal solution would be the suggested express route between Easton and High Bridge.

The network simplex method provided a strong foundation upon which to apply graph theoretical approaches to our models. We applied short path algorithms like Kruskal's algorithm (edge first) and Prim's algorithm (vertex first) on Figure 3 and Figure 4**,** which were designed using Geometer's SketchPad. In Figure 3, the most optimal path from Easton to High Bridge would be Easton→ Clearwater/Lebanon→High Bridge, and the optimum path from Easton to Gladstone is Easton→Clearwater/Lebanon→Gladstone. Figure 4 extends this model by including every point between Easton and NYC. Including both cost and profit into the model, the optimal solution would be to design High Bridge as the express train, and Gladstone will remains strictly as a local train. The advantages of the express train will serve as an incentive for commuters.

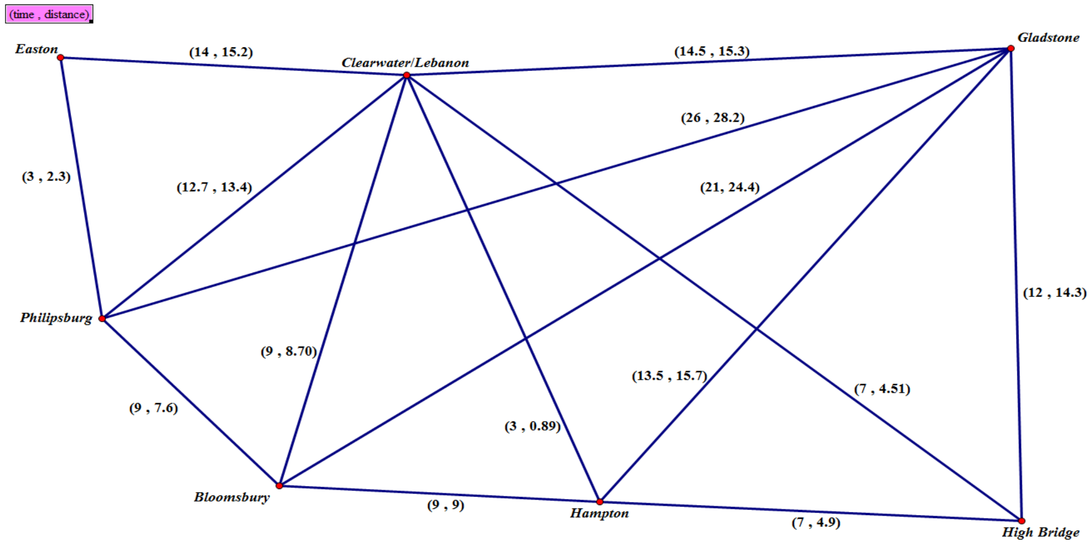

Figure 3-Graph of total connections between Easton and High Bridge/Gladstone with the assumed model.

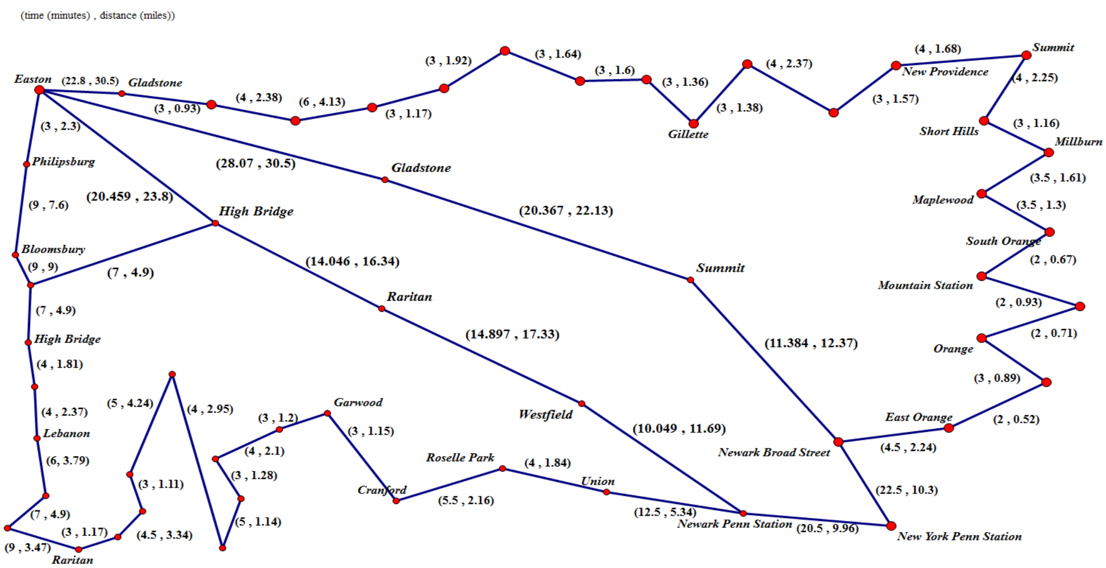

Figure 4-Total connections between Easton and New York Penn Station

5. Discussion

There were various roadblocks which limited the effectiveness of applying operations research techniques. As stated before, despite all the information provided, there were still substantial gaps which hindered more thorough analysis. Due to the assumptions that needed to be followed, the model is idealized, and thus may not properly address some of the key factors in railway design. Additionally, to utilize the simplex method, it was assumed that the model was linear, which may well not be the case.

There are, however, many positive outcomes that should be noted. Through this research, we can conclude that this rail system is feasible and that a workable system can be designed. Since there is public interest, public work programs could be created which would spur job growth. Similarly, this would also have a positive economic impact on the community. Currently, white collar workers in the region are commuting by highway. As this population grows in Eastern PA, a commuter rail service will be crucial to handle the high capacity of people. Additionally, the system we designed aligns with the environmental guidelines in the LVSTP.

6. Future Directions

If data and tools were more accessible, we would design more sophisticated systems where linearity would not need to be assumed. We could also expand this model into central PA, and not just Easton, PA. One aspect that was overlooked is how taxes would be affected by the introduction of a new rail and what effect this would on the general population. Additionally, we could change the focus of the model to maximize freight line usage, which would require us to navigate private ownership issues.

7. Acknowledgement

I thank Dr. Linda McGuire for guiding me throughout this project which enabled me to complete these analyses.

8. Bibliography

1. F. Barber, M.A. Salido, L. Ingolotti, M. Abril, A. Lova, P. Tormos, *An Interactive Train Scheduling Tool for Solving and Plotting Running Maps*, Current Topics in Artificial Intelligence, Vol. 10, 646-655, 2004.
2. Barrow, Keith, *Look—no wires!*, Railway Age, Vol. 210, Issue 9, 58-61, September 2009.
3. M.R. Bussiecky, U.T. Zimmermann, *Combinatorial Optimization Methods for Optimal Lines in Real-World Railway Systems*, Technical Report TR-95-03, 1996.
4. M.R. Bussiecky, U.T. Zimmermann, *Discrete optimization in public rail transport*, Mathematical Programming 79(3), 415-444, 1997.
5. Flyvbjerg, N. Bruzelius and B. van Wee, *Comparison of capital costs per routekilometre in urban rail*, European Journal of Transport and Infrastructure Research, 8:1, 17-30, 2008.
6. F.P. Garcia Marquez, F. Schmid and J. Conde Collado, *A reliability centered approach to remote condition monitoring. A railway points case study*, Reliability Engineering and System Safety, 33-40, Elsevier Science, Amsterdam, 2003.
7. P. Gordon and P.E. Kolesar, *A Note on Rail Transit Cost-Benefit Analysis: Do Nonuser Benefits Make a Difference?*, Public Works Management and Policy, 16, 100-110, April 2011.
8. M. Grunwald, *Back on Tracks*, Time, Vol. 180, 2012.
9. Lehigh Valley Planning Commission, *Lehigh Valley Surface Transportation Plan 2011-2030*, October 2010.
10. M. McHutchon, W. Stasziewski, F. Schmid , *Signal processing for remote condition monitoring of railway points*, Strain, Blackwell Science, May 2005.
11. D. Pickrell, *Estimates of Rail Transit Construction Costs*, Transportation Research Board, 54-60, Washington, 1985.
12. Salido, Miguel A., M. Abril, F. Barber, L. Ingolotti, P. Tormos, and A. Lova.*Applying Topological Constraint Optimization Techniques to Periodic Train Scheduling*. Technical Report. Universidad de Alicante, Spain, 2005.
13. F.Schmid, *Train Control Research in Europe*, Proceedings of the Institution of Railway Signaling Engineers for 2001-2002,52-66, IRSE, London, 2002.
14. F. Schmid and N. König, *Reliability demands the simplification of train control techniques*, Railway Gazette International, Vol. 151, No. 9, 561-564, September 1995.
15. ------, *NJ Transit Facts At A Glance*, Fiscal Year 2011, Revised February 2012.
16. ------, *Rail Costs Rocket*, International Construction, Vol. 50, Issue 10, 7-7, December 2010.
17. ------, *Rail Operations and Infrastructure Planning Services: Flemington Corridor Transit Study*, DRAFT Task 10: Phase 2 Report, 16-28.
18. http://quickfacts.census.gov/qfd/states/42/4221648.html.
19. http://www.njtransit.com.
20. http://www.railroad.net/forums/viewtopic.php?f=137&t=84327.