Profit-Generating Capacity for a Freight Railroad

Melody Khadem Sameni (corresponding author)
PhD student
Transportation Research Group
School of Civil Engineering and the Environment
University of Southampton
Southampton, UK
SO17 1BJ
Tel: (0044)23 – 8059-3013 Fax: (0044)23-8059-3152
Email: m.sameni@soton.ac.uk

Mark Dingler
Graduate Research Assistant
Railroad Engineering Program
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign
B-118 Newmark Civil Engineering Laboratory
205 N. Mathews Ave., Urbana IL, 61801 USA
Tel: (001)217- 244-6063
Email: dingler2@uiuc.edu

John M. Preston
Professor
Director-Transportation Research Group
School of Civil Engineering and the Environment
University of Southampton
Southampton, UK
SO17 1BJ
Tel: (0044)23-8059-4660 Fax: (0044)23-8059-3152
Email: jpreston@soton.ac.uk

Christopher P.L. Barkan
Professor
Director – Railroad Engineering Program
Department of Civil and Environmental Engineering
University of Illinois at Urbana-Champaign
1203 Newmark Civil Engineering Laboratory
205 N. Mathews Ave., Urbana IL, 61801 USA
Tel: (001) 217-244-6338 Fax: (001) 217-333-1924
Email: cbarkan@uiuc.edu

Word count: 7400 (4650 words + 4 Tables + 6 Figures)
Submission date: 1 August 2010
Abstract

In the face of enormous and continuously growing demand but limited infrastructure, railroads worldwide are struggling with the capacity challenge. As expansion projects are time consuming and capital intensive, existing infrastructure should be used as efficiently as possible. Different categories of metrics exist for measuring how well the infrastructure is utilized including throughput (trains, tons, cars, train-km, ton-km), level of service (terminal dwell, velocity, delays) and asset utilisation (velocity, occupation time of infrastructure). However, current metrics cannot provide enough clues for the railroad authorities to choose the best combination of heterogeneous traffic. Different combinations of traffic incur different costs, revenue, delays and capacity utilisation. This article studies the new concept of profit-generating capacity and introduces a novel approach to measure capacity utilisation by means of profit. It would bring capacity analysis in line with the operational goal of railroads and enable considering different values of trains, seeing the big picture and more efficient decision making for getting the maximum value. A case study for heterogeneous traffic of bulk and intermodal trains is simulated by Rail Traffic Controller (RTC) for various traffic levels (8 to 48 trains a day) and incremental levels of heterogeneity (0% percent of bulk trains to 100%). The early results of studying this new concept are presented and discussed.

Key words: railroad capacity, profit, heterogeneity, rail traffic controller, simulation modeling, freight train

231 words
1. Worldwide Railroad Capacity Challenge

Expansion of railroad infrastructure has not kept up with the growth in passenger and freight demand in recent years leaving many railroads around the world with the “capacity challenge”. In the European Union since 1995, many railroads have had considerable growth in passenger-km and freight ton-km. The fastest growth in passenger-km since 1995 belongs to Great Britain (42%), Ireland (41%), Belgium (40%), France (34%) and Spain (28%) and the fastest growth in freight ton-km belongs to Great Britain (61%), Switzerland (57%), Austria (43%), Hungary (33%) and Germany (30%). For more details and analysis of European rail growth trends please refer to [1]. These trends have caused saturated railroad networks in many European railroads and continuous growth trends would exacerbate the situation in coming years.

The same trends can be seen in North American railroads. The US Department for Transport estimates that total railroad freight tonnage will increase 88% by 2035 requiring an investment of $148 billion for infrastructure expansions. Without any improvements, 25 percent of primary corridors would operate at or near capacity and 30 percent above capacity as shown in Figure 1. [2]

![Figure 1- Volume of primary corridors in 2035 without infrastructure expansions](image)

In the face of the railroad capacity challenge, proper utilisation of infrastructure is extremely important. Railroad researchers have been studying different methods of capacity analysis and improvements which would be summarized in the literature review section.

2. Literature review

Abril et al. [3] review different methods for railroad capacity analysis. Capacity analysis methods are mainly dependent on whether the railroad is run through an exact timetable or
not. Passenger focused railroads, as in Europe, are run through timetables whereas freight focused railroads, as in North America, mainly operate without timetable and on-demand. Therefore, we divide the methods into the two above mentioned approaches.

a. Timetable based capacity analysis methods

For the railroads that are run through timetable, most capacity analysis methods focus on analyzing and improving timetables. Extensive research has been carried out on different aspects of timetables and different surveys have been published to summarize them. Cordeau et al. [4] review train scheduling literature as a sub category of their comprehensive survey. Tornquist [5] provides an overview of research in the field of railroad scheduling and dispatching. Hansen et al. [6] present different state-of-the-art techniques on timetable design principles, infrastructure modeling, running time estimation, timetable stability analysis, optimization models for railroad timetabling, simulation, rescheduling and performance evaluation. Lusby et al.[7] provide a recent survey on track allocation models and methods.

Capacity analysis of timetables ranges from simple theoretical formulae to sophisticated simulation software. Two major theoretical methods for analyzing capacity utilisation of a timetable are capacity utilisation index (CUI) and UIC 406 methods. The underlying concept of both of them is identifying how much the infrastructure is occupied.

Capacity utilisation index (CUI), as shown in Figure 2, is defined as the time taken to operate a ‘squeezed’ timetable to the time taken to operate the actual timetable[8].

![Figure 2- Capacity utilisation index](image)

The most famous theoretical formula for capacity analysis is UIC 406 capacity method developed by UIC [9]. After generation of the timetable, the railroad network is divided into sections at:

- Junctions
- Change of train order
- Change in number of trains
- Single ↔ double track changes

For the next phase, timetable is compressed. All train paths are “pushed together to the minimum headway” without any changes in the running times, running time supplements, dwell time at stations and block occupation time. Finally, the total percentage of occupation time of infrastructure is calculated [9, 10].
Major software used in European railroads to generate, analyze and compare timetables, identifying bottlenecks and analyzing capacity utilisation are:

- RailSys by Rmcon, Germany
- DONS and its SIMONE module for capacity by Railned, Netherlands
- PETER by Delft University, Netherlands
- VIRITATO and its CAPRES module for capacity by EPFL and SMA and partner, Switzerland
- DEMIURGE by SNCF, France
- RAILCAP by Stratec, Belgium
- CMS by DeltaRail, UK

b. Non-timetable capacity analysis methods

Krueger [12] built a parametric model for capacity planning at Canadian National Railway (CN). It takes into account plant parameters such as block length, meet-pass point spacing and uniformity, signal spacing and percentage of double tracks as well as traffic and operating parameters such as traffic peaking, priority of trains, speed ratio and running times. McClellan [13] reviews different aspects of railroad capacity. Harrod [14] studied railroad capacity management in the context of North America. Lai [15] developed a decision support framework to optimize investing in different capacity expansion schemes for North American railroads. It has three modules: “Alternative Generator” that lists all expansion options along with their costs and impact on capacity; “Investment Selection” that identifies parts of the network that need to be upgraded and the proper capacity improvement option; and the “Impact Analysis” module that investigates the costs of delay versus investment. Dingler et al. [16] studied the effect of heterogeneity on railroad capacity utilisation for a typical North American freight railroad that can be used to find the best combination of traffic that causes least delays. Gorman [17] used econometric methods to analyze congestion delays in BNSF, a major American freight railroad.

The main infrastructure management software used by all Class I railroads in North America is Rail Traffic Controller (RTC) developed by Berkeley Simulation Software [16].
3. Current metrics of railroad capacity for a freight railroad

It is not precise and scientific to define the capacity in railroads as the maximum number of trains because it is dependent upon the level of service, types of trains, length of trains, etc. Railroad capacity can be expressed in different metrics; each metric has specific applicability and weaknesses, and analyzing trends using a single metric fails to capture the complexity of rail performance[18]. Within each metric there are several different units that can be used to understand and quantify that metric (Table 1). Each metric focuses on one aspect of railroad operations, and unfortunately they are not directly convertible. Instead each metric reveals something different and is important to a different group in the railroad industry.

In this section an overview of existing metrics of railroad capacity is presented. For more details please refer to [19].

Table 1 - Current metrics of railroad capacity

<table>
<thead>
<tr>
<th>Metric</th>
<th>Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Throughput</td>
<td>Trains, tons, cars, train-km, ton-km</td>
</tr>
<tr>
<td>Level of service</td>
<td>Terminal dwell, velocity, delay</td>
</tr>
<tr>
<td>Asset utilisation</td>
<td>Velocity, Percentage of occupation time of infrastructure</td>
</tr>
</tbody>
</table>

a. Throughput

Throughput is a measure of how much material can be transported over a route in a specific period of time. Throughput can be measured in trains, tons or cars. The number of trains per day is often used when discussing capacity on a route and has the advantage of being intuitive, easily measured and understandable by the general public. Tonnage is often used when considering the ability to move cargo over a route and is categorized either as gross or revenue tons. Cars are most useful when considering terminals. Cars is often used as the standard unit of terminal capacity; much like trains per day is used for line capacity. In general, cars is not useful as a unit of line capacity because two trains with different numbers and sizes of cars use similar amounts of capacity. Train-km and ton-km are more detailed measures of throughput.

The key weakness of using throughput as a capacity metric is its variability depending on the specific traffic and operations. Different train types use different amounts of capacity, carry different amounts of cargo and have different numbers of cars. Throughput is a strong metric of capacity as it is a direct measure of the movement of goods across a line or through a terminal, but has limitations since it is highly influenced by the train and car types.

b. Level of Service

Level of service is a measure of the reliability and timeliness of transportation. Long travel times or unreliable deliveries are unacceptable to shippers. The units of level of service are
terminal dwell, punctuality, reliability and train delay. Level of service is an indirect measure of capacity since it only measures the effect of insufficient or excess capacity and not capacity itself.

Most traffic has scheduled arrival windows for each customer. Variability in the travel time makes arriving in the time window difficult. As defined by the American Association of Railroads [20], terminal dwell is the average time a car resides at the specified terminal location. The measurement begins with a customer release, received interchange, or train arrival event and ends with a customer placement (actual or constructive), delivered or offered in interchange, or train departure event. While velocity measures the line-haul movement between terminals. Velocity is calculated by dividing train-miles by total hours operated, excluding yard and local trains, passenger trains, maintenance of way trains, and terminal time. Since velocity and terminal dwell are dependent on many factors and not comparable between railroads, it is the change in these values that is important. In general, falling average speeds or increased terminal dwell imply problems, and rising average speeds or reduced terminal dwell imply improvements [18].

The other unit of capacity that measures level of service is delay. Delay is additional travel time due to scheduled and unscheduled events. Delay is the summation of the length and number of late arrivals, therefore delay directly measures the magnitude of the variability in travel times. Delay is important since it is the primary output from simulation modeling of railroad operations. Change in delay is often used by the railroads to determine its benefit of a project. A literature review on costs of delays to freight railroad, causes, measures, impacts on different types of trains, importance and value of reliability has been presented by Johnson and Fowkes [21]. Major causes of freight congestion delays has been extensively studied and statistically analyzed by Gorman [17].

To shippers, level of service is the most important metric of capacity. Different shippers have different level of service requirements. Intermodal customers require highly time sensitive shipments while coal deliveries are less time sensitive. Level of service alone does not allow for any estimation of the capacity of network and offers no information on the movement of goods.

c. Asset Utilisation

Asset utilisation measures the usage of the railroad’s assets. A railroad’s assets include the infrastructure, rolling stock, motive power and personnel. Railcars and locomotives are costly to purchase and maintain and railroads want to make efficient use of them. A railroad can move sufficient cargo at a high level of service but still be underutilizing its assets. Asset utilization, like level of service, is an indirect measure of capacity.

The main unit that measures asset utilization is velocity. With a constant amount of traffic, an increase of velocity shortens cycle times on cargo allowing more efficient use of the assets. At one major railroad an increase of one mile per hour in velocity could mean that 250 locomotives, 5,000 freight cars, and 180 train and engine employees would be freed up to move additional traffic [22].

Occupation time of infrastructure can be a measure of asset utilisation but is mainly used for passenger focused railroads that run through timetable as previously discussed. Velocity is useful for measuring asset utilization, it is a weak measure when it comes to capacity. Velocity is highly dependent on the number and type of trains being operated. The other disadvantage of using velocity is that it does not provide any information on the quality of
service. Two trains can have the same average speed but if the service demands of one train type are greater than the other, this is unacceptable from the business standpoint.

4. Profit-generating capacity: a new metric

In the end, to railroads, capacity is not about throughput, level of service or asset utilization, it is about money. Currently the metrics have to be considered together as each provides different information on the profitability of the infrastructure asset. Consequently by considering the profit generation on a route all the other metrics can be considered together while providing a railroad data on the most important aspect of capacity, financial return.

Railroad infrastructure is a limited and expensive resource that is allocated to trains; it should be utilized the best way possible to get the maximum possible value. However, current methods of capacity analysis and metrics do not provide a proper tool for advising freight railroad how to get the maximum value from this resource.

Value is an expression of “the relationship between function and resources where function is measured by the performance requirements of the customer (such as quality of service) and resources are measured in materials, labour, price, time, etc. required to accomplish that function”[23].

It can be represented as

Value = Functions / resources  [23]

Freight railroads are usually faced with different options for using the railroad infrastructure: different train types incur different costs and revenues. On the other hand, combinations of different train types reduce the levels of service. In North America a wide range of commodities are transported by rail (Figure 4). Choosing the best combination of train types that offer the maximum value and profit for the railroad is an important decision. For example bulk trains are on average longer and slower but the costs of delays for intermodal trains are more than twice as much as bulk trains [16]. Effects of different levels of heterogeneity and combinations of bulk and intermodal trains on delays has recently been studied by Dingler et al.[16]. However, total revenue, costs and profit should also be considered to assist the railroad authorities to choose the best combination of traffic.
This study introduces the concept of profit-generating capacity as a new metric for measuring capacity in the category of throughput. In this regard capacity of infrastructure is defined as “the ability of infrastructure to generate profit”. Contrary to other metrics of throughput, this metric would use a currency unit to measure capacity utilisation. This would provide a better metric for capacity analysis than current ones by:

- Using currency unit for capacity which is in line with the operational goal of railroads
- Considering different types of trains and their values
- Seeing the big picture of using the infrastructure
- Capturing the complex nature of railroad capacity more
- Enabling more efficient decision making for getting the maximum value

5. The case study: Calculating profit-generating capacity

To analyze capacity utilisation by means of the profit generated, the following steps were taken. All the calculations were normalized for 100 miles.
1) Simulating incremental traffic (8 to 48 trains) and heterogeneity levels (0% of bulk trains to 100%)
2) Extracting fuel consumption, total delays and calculating running time
3) Calculating total costs
4) Calculating total revenue
5) Calculating total profit

a. Characteristics of the case study

Bulk and intermodal traffic account for roughly 60% of the American railroad’s revenue, 75% of the tonnage and 80% of the carload [20]. However, they utilize the capacity of infrastructure in different ways and incur different costs and revenue. In this study tested the concept of profit-generating capacity to figure out the best combination of bulk and intermodal trains that generate the maximum profit for the railroad.

Simulations were run by Rail Traffic Controller (RTC) developed by Berkeley Simulation Software at different levels of traffic from 8 to 48 trains as well as 0%, 12.5%, 25%, 50%, 75%, 85.5% and 100% of each train type. Train composition characteristics are described in Table 2. They were equally distributed over a 24 hour period.

Table 2 – Train composition characteristics in the simulation

| Intermodal | Bulk
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>16 3-pack spine cars</td>
<td>115 loaded hopper cars</td>
</tr>
<tr>
<td>9 5-pack well cars</td>
<td></td>
</tr>
<tr>
<td>5,659 ft</td>
<td>6,325 ft</td>
</tr>
<tr>
<td>5,900 tons</td>
<td>16,445 tons</td>
</tr>
<tr>
<td>3.64 HPTT</td>
<td>0.78 HPTT</td>
</tr>
<tr>
<td>5 4,300 HP Locomotives</td>
<td>3 4,300 HP Locomotives</td>
</tr>
<tr>
<td>Maximum Speed: 70 mph</td>
<td>Maximum Speed: 50 mph</td>
</tr>
</tbody>
</table>

Track was chosen to a single-track mainline subdivision with the following attributes:

- 262 miles long
- 10 miles between siding centers
- 8,700 ft signaled sidings with 24 powered turnouts
- 2.75 mile signal spacing
- 2-block, 3-aspect signaling
- 0% grade and curvature

In order to figure out how well the capacity is utilized, total profit at each level of traffic and different heterogeneities is estimated. Total costs and total revenue is calculated to infer the profit for the stakeholder, the railroad company.
b. Calculating total costs

General workflow of calculating total costs is shown in Figure 6

Figure 6- General workflow of calculating total costs

Total costs of running freight traffic on a line can be simplified as:

\[ TC = a \times VM + b \times VH + c \times V + d \times RM \]

TC: total costs
VM: vehicle-mile
VH: vehicle-hour
V: vehicle
RM: route-mile
a,b,c,d : parameters

In this study we considered major costs including fuel, locomotive, car and crew costs. The values used in the study are calculated as in Table 3. For more details please refer to [19].
c. Calculating revenue

i. Elasticity of revenue to train frequency

Total revenue should be calculated for each traffic and heterogeneity level. Due to the different nature and sources of traffic, bulk and intermodal trains are considered to be independent from each other, hence revenue would be independent of heterogeneity level. But it should be investigated whether or not the revenue is elastic with regard to train frequency. The authors could not find any reference about elasticity of revenue to train frequency in the literature. Therefore it is deduced by price elasticity from the following formula:

Frequency elasticity = Price Elasticity × Value of frequency × (Price/Frequency)

\[
\text{Value of frequency} = \frac{\partial U}{\partial F} \cdot \frac{\partial P}{\partial F}
\]

\[
\frac{\partial Q}{\partial F} \cdot \frac{F}{Q} = \frac{\partial Q}{\partial P} \cdot \frac{P}{Q} \cdot \frac{\partial P}{\partial F} \cdot \frac{P}{F}
\]

F: number of trains (intermodal/bulk) per day
Q: revenue (dollar per container/car)
P: price (dollar per container/car)
U: utility

Rail price elasticity for intermodal trains (nondurable manufactures) is assumed to be -0.6 \cite{25}. Studies such as Zhong \cite{26} have been unable to find statistically significant frequency parameter; based on his work value of frequency for intermodal trains \( \frac{\partial P}{\partial F} \) is calculated as 0.157 pounds per container per departure per day. Exchange rate of pound to dollar is considered to be 2 for 2007 (the year of that study). Average revenue per container and average distance for intermodal trains (Extracted from Class 1 Railroad statistics [20]), yields average freight rate (P) for a container for 100 miles as $112 per container.

At traffic level of 28 intermodal trains per day, elasticity of revenue with respect to train frequency is:

\[
\frac{\partial Q}{\partial F} \frac{F}{Q} = -0.06 \times 0.157 \times 1/2 \times 112/28 = -0.018 \approx 0
\]

In this regard, intermodal revenue is considered to be relatively inelastic to train frequency. It was not possible to calculate elasticity of bulk trains with regard to train frequency as no reference in literature was found regarding value of frequency for bulk trains. However, bulk (coal) trains are less time-sensitive compared to intermodal trains, therefore their elasticity to train frequency must be less than intermodal trains. Hence, it can be inferred that revenue from bulk trains is inelastic to train frequency as well.

ii. Considering Empty Return Ratio

The average revenue per loaded train for bulk trains is more than intermodal trains. However it should be noted that bulk trains have a higher Empty Return Ratio\(^1\). This fact should be considered when calculating total revenue for different combination of trains to reflect their real value. This is because the capacity of infrastructure would be wasted when trains are hauled empty. By considering the Empty Return Ratio and average length of haul, revenue is adjusted for each type of train for 100 miles as in Table 4.

\(^1\) Empty Return Ratio is defined as total miles divided by loaded miles.2. Cambridge Systematics, *National rail freight infrastructure capacity and investment study*. 2007, Association of American Railroads: Cambridge, USA.
Table 4- Adjusted average revenue per type of train

<table>
<thead>
<tr>
<th></th>
<th>Bulk Trains</th>
<th>Intermodal Trains</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average revenue per train</td>
<td>$203,182</td>
<td>$128,246</td>
<td>[20]</td>
</tr>
<tr>
<td>Average length of haul (mile)</td>
<td>707</td>
<td>828</td>
<td>[27]</td>
</tr>
<tr>
<td>Empty Return Ratio</td>
<td>2.03</td>
<td>1.13</td>
<td>[2]</td>
</tr>
<tr>
<td>Adjusted revenue per train (100 miles)</td>
<td>$14,156</td>
<td>$13,706</td>
<td></td>
</tr>
</tbody>
</table>

6. Results and conclusions

Based on the simulation results from RTC software, total costs, revenue and profit were calculated for 77 different scenarios of traffic level (total number of trains) and heterogeneity (percentage of bulk/intermodal trains). Figure 7 shows how total profit varies against different traffic and heterogeneity levels.

![Figure 7- Total profit against different heterogeneity and traffic levels](image-url)
The following results can be concluded from this study:

- When revenue is adjusted by Empty Return Ratio and average length of haul, revenue per bulk train and per intermodal train are very close to each other (bulk train being slightly more).
- Within the same level of traffic, total profit (total revenue minus total costs) increases as the percentage of intermodal trains increases and the percentage of bulk train decreases. This would lead to a better utilisation of infrastructure capacity.
- There is significant increase in total delay and total costs between 25% and 75% of heterogeneity.
- Static costs of delay (eg. $1392 per hour for intermodal trains and $586 for bulk trains [16, 19]) are negligible against revenues of extra trains. Therefore total profit would increase as the number of train increases. Dynamic costs of delays that vary according to the level of traffic would better reflect the consequences of adding extra trains.
- At the same level of traffic, best utilisation of capacity in terms of profits generated is heterogeneity of less than 25% or more than 75%.
- Within the same level of traffic, total profit generated can be a good indication of how well the infrastructure is utilized hence it can be used for finding proper heterogeneity level. However, further research is needed for finding the optimum level of traffic (total trains per day) as currently in the literature costs of delays are mainly considered static which falls far below the added revenue of an extra train. For tackling this issue, maximum allowable delay, inventory costs, yard time and added cycle time should be considered to calculate dynamic costs of delay that varies according to the traffic level.

The authors are currently working on a sensitivity analysis that costs of delays are based on “the value of time per hour per ton of shipment”[28] which results in higher costs of delays.

Further research is suggested for:

- Considering yard times in total costs and profits and how different levels of traffic affect yard times and inventory costs.
- Research on dynamic costs of delays according to the level of traffic.
- Calculating value of frequency for bulk and intermodal trains in a freight railroad.
- Considering maximum allowable delay and dedicating infinity costs to the delays more than the set limit.
7. References

15. Lai, Y.C., *Increasing railway efficiency and capacity through improved operations, control and planning*, in Civil Engineering. 2008, University of Illinois at Urbana-Champaign: Urbana-Champaign, USA.
23. SAVE, *Value standard and body of knowledge*. 2007, SAVE International: USA.