FLEXIBLE APPOINTMENT BASED SYSTEM WITH ADAPTIVE RESPONSE TO TRAFFIC AND PROCESSING DELAYS

Amrinder Arora, DSc
NTELX – Research and Development,
1945 Old Gallows Rd, Suite 700, McLean VA 22182, USA
+1 703 356 5050 x116, aarora@ntelx.com

ABSTRACT
We present an adaptive freight traffic management system that uses flexible appointments and adjusts the schedule based on traffic and processing delays. The system provides a form of first-in-first-out fairness while trying to maximize system throughput. The inputs to the system can include, for example, the ocean vessel and train arrival data, the processing times at various terminals, the real-time traffic information, the location information from participating trucks and statistical records of past transit times. The system applies the algorithms presented herein on the inputs, and continues to adjust the appointment slots and communicates in real-time with participating trucks to update the appointments. An implementation of this system is currently in use at the Jordan-Syria free trade zone near the town of Nasib. Upon its deployment, the zone administrators have expressed greater comfort with orderliness of zone operations and the local councils have confirmed reduction in congestion. Results also confirm an increase in the volume of trucks handled by the free trade zone.

INTRODUCTION

In this paper, we focus on efficient management of freight traffic. Freight traffic, of which trucks and freight trains are primary examples, is a special class of traffic that shares many similarities with general traffic but also differs in two significant ways. Firstly, the freight traffic shares the transportation infrastructure with other classes of traffic, and the amount of overall traffic has a sizeable impact on the freight traffic. Secondly, freight traffic chains display some characteristics that are not always shared with general traffic. For example, a typical freight traffic chain may involve a truck leaving a trucking yard, arriving at a container terminal to drop a container, arriving at a different terminal to pick up a container, and dropping that container to a distribution center before returning to the trucking yard. Each of the operations in the freight traffic chain is expected to take a certain amount of time and the goal of the freight traffic management system is to maximize the system objectives that relate to the specific freight traffic chains.

There are many approaches to managing traffic, and specifically, freight traffic. For example (1) describes network models and how they are applied in transportation, (2) describes a capacity management approach, (3) describes a variable tariff based approach, and (4) describes an appointment based approach. (5) presents an interesting study of container transfers at multimodal terminals that uses some of the similar concepts. As a way of classification, most of the freight traffic management approaches can be classified into three broad categories: (i) capacity based approaches, (ii) appointment based approaches, or (iii) incentive (or variable fee) based approaches. A system may also use a combination of these approaches.

All of these approaches, and by extension, the systems that implement these approaches, have certain advantages and disadvantages. For example, a capacity based system manages traffic by limiting entry to the system, and is extremely resilient to changes within the traffic network. However, it does not provide any mechanism to schedule a specific time interval access to any resource. For example, suppose a capacity management system is being used to manage a freight traffic chain that involves access to specialized crane equipment, and that the equipment is usually unavailable from 12:00 to 13:00. In this case, the system would need to mark the "capacity" of the relevant portion of the traffic chain to zero to indicate the unavailability of the equipment during the hours. While this workaround is not necessarily incorrect, still, the flowing aspect of capacity management makes specific location based scheduling difficult, or at least, artificial.

Similarly, an appointment based scheduling system (sometimes simply referred to as appointment based system) is generally used in cases where specific operations need to be scheduled at specific resources. It can be very effective if the capacity of a single resource is a dominant factor in the system. An appointment based system can decrease the congestion and waiting time very quickly, but if the schedule is too conservative, it can lower the system throughput significantly. An appointment based system can also suffer from the classic problem of no-shows, in which case the system throughput can further diminish.

Finally, an incentive based (variable tariff) system uses market forces of demand and supply to regulate and manage traffic. For example, it can impose a higher tariff during the normally congested hours, and impose a lower tariff during the hours that have surplus capacity. A significant advantage of such systems is that they do not try to impose change; instead, they create incentives for other actors to change their behavior. One disadvantage of these systems is that they can require frequent "calibrations", for example, the values of the different tariffs may need to be continuously adjusted to cause desired behavior. Thus, a concern with incentive based systems is that if the system is too dynamic, then the users of the system may not find the system easy to comprehend. If, on the other hand, the system is too static, it may not meet its objectives. An interesting study of the impact that incentives have on toll road usage by trucks is presented in (6). It may also be observed that High Occupancy Vehicle programs, even though they are outside the scope of freight traffic management, use an incentive based approach by dedicating resources (typically lanes) for vehicles meeting the occupancy requirements. A usual dilemma that the traffic administrators face regarding HOV programs is the determination of the number of passengers that are required, for the vehicle to be considered high-occupancy. For example, it may happen that HOV-2 program does not lead to any significant incentive as the number of cars meeting that criteria exceeds the capacity of the dedicated lane(s), and that HOV-3 program creates strong incentive but very few cars are able to utilize that incentive. Similarly, the times that the HOV program should be enforced may also need to be regularly updated for maintaining effectiveness, and that needs to be balanced with the need of commuters to have a predictable schedule.

A schematic of these three categories is presented in Figure 1.

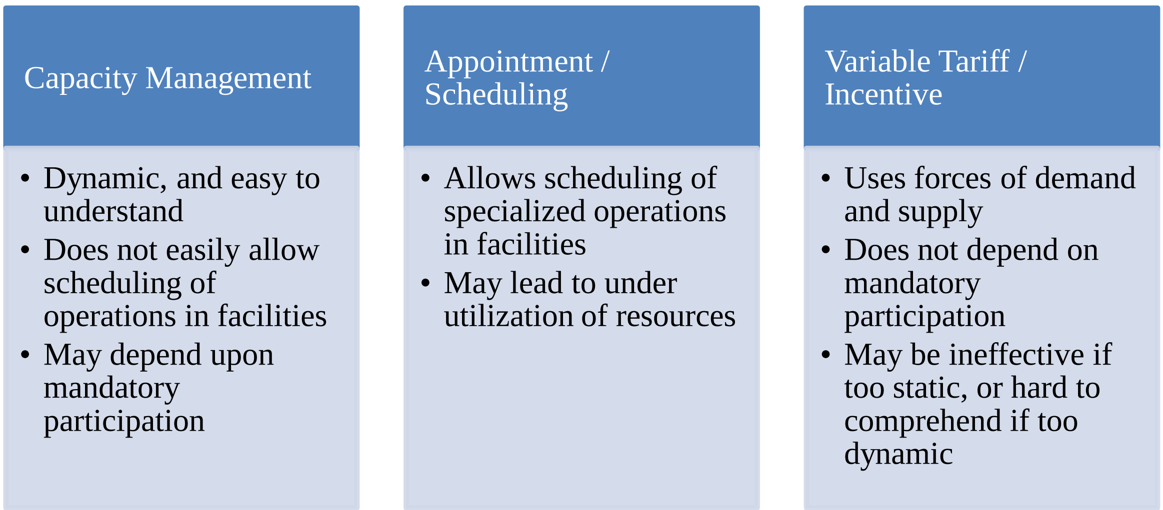

Figure 1: Three main categories of traffic management approaches

A flexible, adaptive appointment based system can maximize the system throughput while minimizing the congestion at the same time. There have been previous attempts at adaptive appointment based systems in different verticals, and have been vertical specific (for example, a system for doctors, technicians, laboratories, etc.). This is, at least in part, due to the significant role that the vertical specific insights play in devising the appropriate adaptation strategy. We observe the same in our practice area of freight transportation as well, and as we discuss further, the strategy to shuffle timeslots in case of delays and changes relies significantly on the business specific insights.

MOTIVATION

The system was motivated directly by a traffic congestion problem at the Syria-Jordan free trade zone near the town of Nasib. Before the deployment of the system, the trade zone officials and customs officials were concerned by a sense of disorder; the trip times were long and indeterminate due to truck/load and documentation problems. There was poor coordination between trucks and businesses within the trade zone and the residents reported frequent altercations with the truck drivers. From a system volume perspective, the free trade zone was unable to keep up with demand, and there were no performance measurements that could be used to objectively measure the situation on a given day, week or month.

BUSINESS OBJECTIVES

The system had the business goal to maximize the system throughput while minimizing traffic congestion and wait times. Additionally, the system has following objectives: (i) users should not be able to make appointments and then sell their slots to others, (ii) while the schedule may not always provide the first-in first-out service, it should respect the intent of a first-in first-out system.

SYSTEM MODEL

Next, we describe the system model. We begin with the physical system model that motivated this study, and then present the logical (generic) system model that is used in problem statement and the solution design.

PHYSICAL SYSTEM MODEL

The physical system consists of trucks which need to perform one or more transportation related activities at specific locations inside the free trade zone. The free trade zone has multiple locations, each with a certain capacity (number of trucks that can be in the loading/unloading area of the location at the same time). The time that the truck needs to spend inside the location is variable but has low variance and hence can be treated as a constant. The access roads to the trade zone can easily get congested if too many trucks are on the road, and the trucks on the road move at largely the same speed (which can vary by time, but at a given time, does not vary significantly between trucks). The trucks arrive at the trade zone from a finite number of origin points, and their preceding legs or dependencies (such as ocean vessel or a container terminal) are known to the system which can aid in calculating their expected arrival times. Thus, the physical system model has the following distinct inputs:

- Vessel data
- Terminal processing times and delays
- Real-time traffic delays and vehicle speeds on access roads

- Location from participating trucks
- Road closures and detours

LOGICAL SYSTEM MODEL

In logical system model, we consider a set of m resources with respective capacities c(1..m). We consider a set of n jobs, each of which needs access to one of the m resources and consider an arrival model a(1..n), where each arrival model contains information about its expected time and other statistical parameters. We model a schedule S as a mapping of resources, jobs and timeslots, where the value S(j,r)=t, if job j is scheduled to use resource r at timeslot t, and is 0 if job j has no current appointment at resource r. We observe that the logical system model has many similarities to the standard job scheduling problem in theoretical computer science, but the objective function is quite different due to the focus on the throughput and the correlation of the arrival of jobs. In the traditional job scheduling problem, the arrival process of jobs is used in the scheduling algorithm, but the arrival process is not itself impacted by the presence of other jobs.

We further define the expected **value** of schedule S over a time range T, E(S,T) to be the expected number of jobs that will be successfully completed over the time range T.

PROBLEM STATEMENT

We formulate the problem purely as an online problem. In the online problem, we are given a baseline schedule S, and a "change" event and we are required to calculate a new schedule S'. The change can represent: (i) the completion of a job, (ii) initiation of a new job, or (iii) an external event that represents the change in an arrival process of a job. The objective is to maximize the expected value of schedule S' while respecting the business constraints specified in Section 0.

It may appear that we also need an offline version of the problem which calculates an initial schedule S, but we can observe that the online version of the problem suffices by using an empty schedule and modeling initial set of jobs as change events of type job arrival.

PROPOSED SOLUTION

Before we describe the algorithm that calculates the schedules, we describe the physical business process and some of the vertical specific heuristics that aid the algorithm. Each trucking company submits a request (job) by specifying a requested arrival time, and also specifies a range of timeslots. Further, it specifies the route and the preceding logistics constraint (such as a ship arrival). The incentive to specify this optional information is that if this extra information is given and the job is delayed (for example, by a traffic incident), the system tries to **accommodate** the job (truck) at a timeslot permissible by system capacities. If the applicant fails to specify any of the optional information and gets delayed, it loses its appointment, the system makes no attempt to accommodate the job, the applicant needs to reschedule the job from beginning which may cause significantly longer wait. While not a technical aspect of the solution, this business aspect can be a key in creating an incentive for providing optional information that helps the system in providing high throughput.

ALGORITHMS

Here we describe the two main algorithms. First algorithm MultiCurveGA is a genetic algorithm that takes historical data and finds correlations between event types and arrival models. The second algorithm uses the correlation models previously computed by MultiCurveGA algorithm, and using the current set of reservations and a newly received change event, computes new set of reservations.

Algorithm MultiCurveGA

The output of this algorithm is a mathematical function for each event type that encapsulates the effect of that specific event type on arrivals. As an example output, the event type of vessel departure of a 5000 TEU ocean liner may yield an arrival probability distribution function of a normal distribution that peaks at a time 48 hours before the vessel departure, and has a variance of 24 hours. It may be quickly observed that the exact arrival process depends not only on the policies and processes in effect at the port, but also on the trade parties involved, and that the model is only a useful approximation.

The event types need not be limited to transportation events, other events such as a sporting event or simple day of the week, week of the year "events" may lead to insights such as the following:

- On the tenth Wednesday of the year, at 8 AM, the average truck driving time from point A to marine terminal B is 48 minutes.
- In case there is a traffic incident of severity X on the highway, the average truck driving time from point A to terminal B at 8 AM is 40% more on days there is no such incident.

The MultiCurveGA algorithm can be set to terminate if the aggregate of all curves fits the historical data within a margin of X%, where X is a system constant. A value of X that can be used in many circumstances is 10, although care must be taken to avoid underfitting and overfitting. We observe that the MultiCurveGA is used to create the models, and is not used operationally by the system for each transaction. Thus, it is acceptable for the MultiCurveGA algorithm to take significant amount of time (for example, a few hours) to run. Once the models have been created, they can be refreshed on a weekly or monthly basis depending upon the age of the data in the system.

Algorithm AdjustAppointments

The "Adjust Appointments" algorithm is a branch and bound algorithm that adjusts the system schedule based on a certain change event that is received. The change event can simply be a location event from a participating truck, a change in vessel schedule, or a truck arrival event. The AdjustAppointments algorithm consists of the following steps.

1) Receive a change event
2) Update all arrival processes that may have an impact from the change event using the algorithm MultiCurveGA presented above.
3) Check the expected **value** of schedule S over a time range T, E(S,T) that is the expected number of jobs that will be successfully completed over the time range T.
4) Compare E(S,T) to the maximum number of jobs theoretically possible (the upper bound) over the time range T. Typically, this can be computed using the system through put.
5) If the expected value is lower than the maximum by more than a threshold value, then do the following steps
   a) Using a branch and bound approach, check for each possible reshuffle of appointments. A business rule that is used in this regard is that an appointment for a particular truck will not be shuffled more than twice.
   b) After each reshuffle, check the expected value against the maximum possible value. Use the normal rules of branching and bounding to eliminate shuffle (branching) nodes. If the shuffle leads to increase in the expected value, add the shuffle to a list of "desirable shuffles".
6) For each reshuffle in the desirable shuffles list, communicate with the truck and other required parties in real time. If the shuffle cannot be committed, abandon the shuffle, and proceed to the next possible shuffle in the shuffle list. If a shuffle is committed, complete the algorithm (do not attempt other shuffles), and trigger the AdjustAppointments algorithm again, using an empty event as a change event.

We observe that AdjustAppointments algorithm works as a tail-recursive algorithm. Using this mechanism, it can make many shuffle events, that is, adjustment events for each change event received.

PROCESS

Next, we describe the actual physical process that is used by various parties involved in the system.

1. The process begins when trucking company dispatchers request appointment with specific FTZ business.

2. The system checks capacity at gate and business, and shows the list of available appointments.
3. The dispatcher selects one of the appointments, and reconfirms request.
4. The system finalizes the appointment.
5. At the time of the appointment, the trucking company dispatchers guide the truck drivers to arrive at the preselected location. The system assists with historical driving times and alternative route suggestions in case of traffic incidents.
6. Trucks that arrive at their preselected location for appointments enter FTZ without delay.
7. System continues to monitor and report truck, system and people performance using the algorithms presented above.

CONCLUSIONS

In this paper, we have presented a flexible appointment scheduling system that manages access of trucks to a free trade zone set in a congested area. The system manages the traffic while adapting to delays due to traffic delays, natural incidents and processing times at terminals. The proposed system provides a form of first-in-first-out fairness and prevents malicious actors from making appointments for profit. While it can shuffle the sequence of appointment slots, it does so in a manner that respects the priority of initial sequence. Using the ship arrival data, the processing times at various terminals, the real-time traffic information, the location information from participating trucks and statistical records of past transit time, the system applies the presented algorithms on the inputs, and continues to adjust the appointment slots and communicates in real-time with participating trucks to update the appointments. Preliminary results from an implementation of this system at the Jordan-Syria free trade zone confirm the positive attributes of the system by way of an increase in the volume of trucks handled by the free trade zone and a simultaneous decrease in the number of incidents between truck drivers and local resident councils.

ACKNOWLEDGEMENTS

The author would like to thank the NTELX board for providing the resources to understand the system, the George Washington University faculty members for valuable discussions around traffic management and the Intelligent Transportation Society of America president Scott Belcher for his inputs and comments on the freight traffic management approaches.

REFERENCES

(1) Magnanti, T and Wong, R, "Network Design and Transportation Planning: Models and Algorithms", *Transportation Science,* INFORMS, February 1984.

(2) Arora, A and Patel, D, "Applying Stochastic Capacity Management to Manage Truck Traffic around a Sea Port", *Proceedings of Hamburg international Conference on Logistics.* Hamburg, 2008.

(3) Giuliano, G and O'Brien, T, "Extended gate operations at the ports of Los Angeles and Long Beach: a preliminary assessment", *Maritime Policy and Management,* Volume 35, Issue 2, Informaworld, 2007.

(4) Huynh, N and Walton, C, "Robust Scheduling of Truck Arrivals at Marine Container Terminals", *Journal of Transportation Engineering,* Volume 134, Issue 8, ASCE, 2008.

(5) Kozan, E, "Optimising container transfers at multimodal terminals", *Mathematical and Computer Modeling,* Volume 31, Issue 10, ScienceDirect, 2000.

(6) Zhou, L and Burris, M and Baker, T and Geiselbrecht T, "Impact of Incentives on Toll Road Use by Trucks", *Transportation Research Board,* Volume 2115, 2009.