AUB-LM-13-077

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

Wave Release Strategies to Improve Service in Order Fulfillment Systems

17 May 2013

Kevin R. Gue, Associate Professor

Erdem Çeven, PhD Candidate

Auburn University

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

Abstract

Distribution centers in the Department of Defense (DoD) have become critical components in getting supplies to the warfighter. In the case of repair parts and other essential items, they become a critical component of the operational availability of major weapons systems.In general, the more quickly the logistics system responds to requests, the higher the availability because total downtime is reduced through reduced mean logistics delay time. The desire to fulfill warfighters' orders immediately, however, must be tempered by the need to provide service at a low cost, which means taking advantage of economies of scale in warehouse picking operations. To strike this balance, distribution centers in the DoD release orders in large batches called waves.Despite their ubiquity in military and commercial warehouse operations, there are no analytical models to determine the optimal number and timing of these waves, especially to maximize performance against deadline-oriented metrics such as Next Scheduled Departure, which is used at the Defense Logistics Agency.We address this deficiency by developing methodologies to determine the optimal number and timing of order releases in a distribution center.

Keywords: Operational Availability, Order Fulfillment, Optimal Wave Release Policies, Simulation

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

The authors thank managers and analysts at DLA Distribution, Susquehanna, PA, for providing data and for many helpful discussions about wave planning systems. We also thank the Acquisition Research Program for providing funding.

THIS PAGE INTENTIONALLY LEFT BLANK

About the Authors

Kevin Gue is the Tim Cook associate professor of industrial engineering at Auburn University. He graduated from the U.S. Naval Academy in 1985 with a bachelor’s degree in mathematics. He received his PhD from the School of Industrial & Systems Engineering at Georgia Tech in 1995. From 1995 to 2004, he was on the faculty of the Graduate School of Business & Public Policy at the Naval Postgraduate School. Dr. Gue’s research interests include logistics modeling and optimization, with applications in distribution, warehousing, and material handling. He is a past president of the College-Industry Council on Material Handling Education.

Erdem Çeven is a PhD candidate in the Department of Industrial & Systems Engineering at Auburn University. His primary research goals are directed toward controlling logistics systems, especially in order fulfillment systems.

THIS PAGE INTENTIONALLY LEFT BLANK

AUB-LM-13-077

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

Wave Release Strategies to Improve Service in Order Fulfillment Systems

17 May 2013

Kevin R. Gue, Associate Professor

Erdem Çeven, PhD Candidate

Auburn University

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Wave Release Strategies to Improve Service in Order Fulfillment Systems

Introduction

The overarching goal of this research is to provide methodologies that analysts at the Defense Logistics Agency (DLA) can use to establish operating parameters that maximize the "operational availability" of supported systems. Operational availability (A_o) of a system is defined as the fraction of time or probability that a system's capabilities will be available for operational use.

Operational availability of an end item is a function of mean time between failures (MTBF), mean time to repair (MTTR), and the mean logistics delay time (MLDT). MTBF is a measure of "up time," whereas MTTR and MLDT are measures of "down time." Operational availability (Department of the Navy, 2003) is defined by

$$A_o = \frac{MTTF}{MTTF+MTTR+MLDT}. \quad (1)$$

Of interest in our work is the observation that as MLDT goes down, via a more responsive order fulfillment system, A_o is higher.

Intuitively, it seems that a logistics system is most responsive when a distribution center minimizes the average flow time, which is the time between arrival of the order and the time it is ready to ship. As Doerr and Gue (in press) observe, this intuition is only partially right. If reducing flow time for an order results in the order getting on a truck it otherwise would have missed, then response time improves; otherwise, it does not. For example, processing an order in one hour instead of three provides no operational benefit if the order is not scheduled to leave the warehouse for another 10 hours.

One explanation for this insight is to realize that warehouse operations are effectively "continuous," in that completed orders arrive at a shipping dock, more or less, in a continuous stream.By contrast, transportation is a cyclical process, due to the need to achieve economies of scale. To coordinate the internal, continuous operations of its distribution centers (DCs) with the cyclical transportation schedules of its transportation providers, the DLA uses a metric called Next Scheduled Departure (NSD), which measures the fraction of orders arriving during a specified 24-hour period (before the cutoff time) that are processed before a specifictruck departure (Doerr & Gue, in press; see Figure 1).

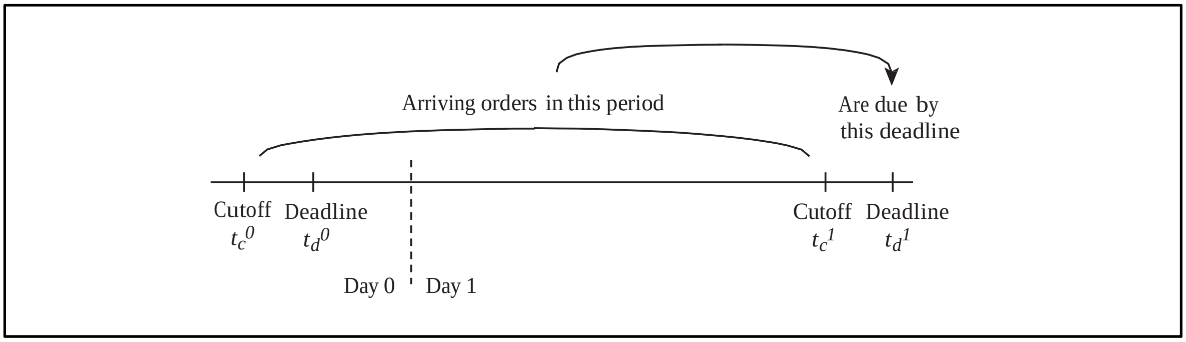

Figure 1. The Next Scheduled Deadline Metric

By definition, an increase in the metric means that more orders make their last departing trucks and that some customers receive their orders before they otherwise would have. Consequently, improving NSD of a distribution reduces MLDT (by definition) and, therefore, increases operational readiness.

Because picking a large batch is much more efficient than picking a single order, making orders immediately available to pickers effectively reduces their productivity, which increases costs. The benefits of large batches, however, must be weighed against the queueing time necessary to form the batch. To strike this balance, DCs at the DLA release orders in large batches called *waves*.

Wave Operations at DLA Distribution

DLA Distribution Susquehanna, Pennsylvania (DDSP), is an extremely complex distribution operation, handling more than one million stock keeping units (skus) stored among dozens of warehouses. DDSP receives orders 24 hours a day, seven days a week, serving a region large enough to require multiple days of transport to distant customers.

Customers are grouped into ordering clusters, and each cluster receives service at a frequency determined by distance from DDSP. In other words, some customer clusters receive service daily, some every two days, some twice per week, and so on. Therefore, each arriving order can be assigned a “next scheduled departure” corresponding to the departure time of a truck to the order’s destination.DDSP refers to customers receiving scheduled deliveries as dedicated truck (DTK) customers. DTK operations constitute one of the major service offerings at DDSP.

Order Flow Analysis

In collaboration with the managers at DDSP, we first focused on the order flow timing of DTK customers. Our interviews with analysts and our data analysis enabled us to understand the current practice and to generate input for our analytical models. Distribution centers of the DLA typically have outbound processes that include picking, packing, order consolidation, and shipping. Example flow timing data are given in Table 1.

Table 1. Sample Order Flow Timing Data

AODORD	NSDDT	NSDTI	XMITDT	XMITTI	RLS_DT	RLS_TI	PICK_DT	PICK_TI
W8002R005	2010061	235900	2010060	121206	2010060	122633	2010060	135404
W8002R007	2010061	235900	2010060	121207	2010060	122633	2010060	135052
PACK_DT	PACK_TI	OFFER_DT	OFFER_TI	MISSION_DT	MISSION_TI			
2010060	153610	2010060	161740	2010060	201700			
2010060	153737	2010060	161740	2010060	201700			

The first entry in Table 1 refers to the order ID. Date and time fields are Julian day and military time (e.g., 2010257 refers to September 17, 2010; 230140 refers to time 2301 and 40 seconds).The next two fields refer to the scheduled departure time, followed by the arrival date and time of the order. The field RLS refers to the date and time that the order is released for picking. Pick completion date and time are given in PICK_DT and PICK_TI. Packing completion date and time are given in fields PACK_DT and PACK_TI. Because orders wait for consolidation, there is a consolidation date and time stamp (given with OFFER_DT and OFFER_TI). The last two data fields correspond to the actual shipment date and time. Figure 2 is an illustration of order flow.

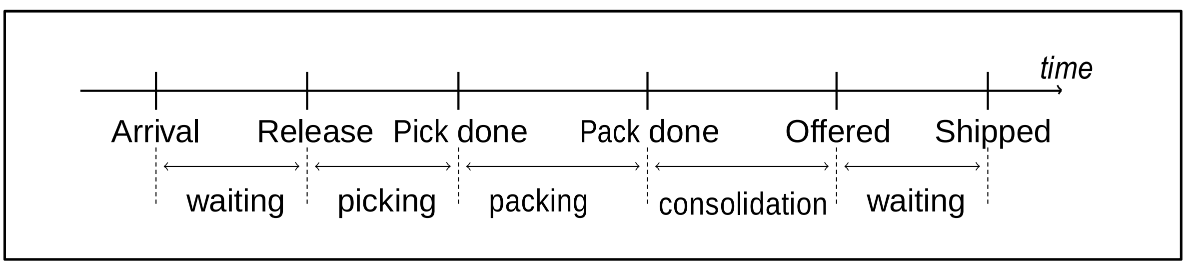

Figure 2. Timeline of an Order Through Arrival-to-Ship Process

We were provided with three months of order flow data from January–March 2010 for dedicated truck (DTK) operations—a total of 402,406 orders. Of those orders, 351,866 arrived in 2010 (87.44% of total) and 351,530 (87.36% of total) were shipped during the three-month interval and were the subject of our analysis.

The leftmost pair of columns in Figure 3 demonstrate the total number of arrivals and releases on Mondays (followed by other days). DDSP observes a heavy workload on Mondays; on the other hand, a relatively small number of orders are released for picking. In an attempt to maintain service promises, the number of releases is increased for the other days in the week; however, internal operations can only catch the arrivals on Friday. Further, the workload on Saturdays is not heavy, but there are many releases on this day, indicating an attempt to improve the low service performance from previous days.

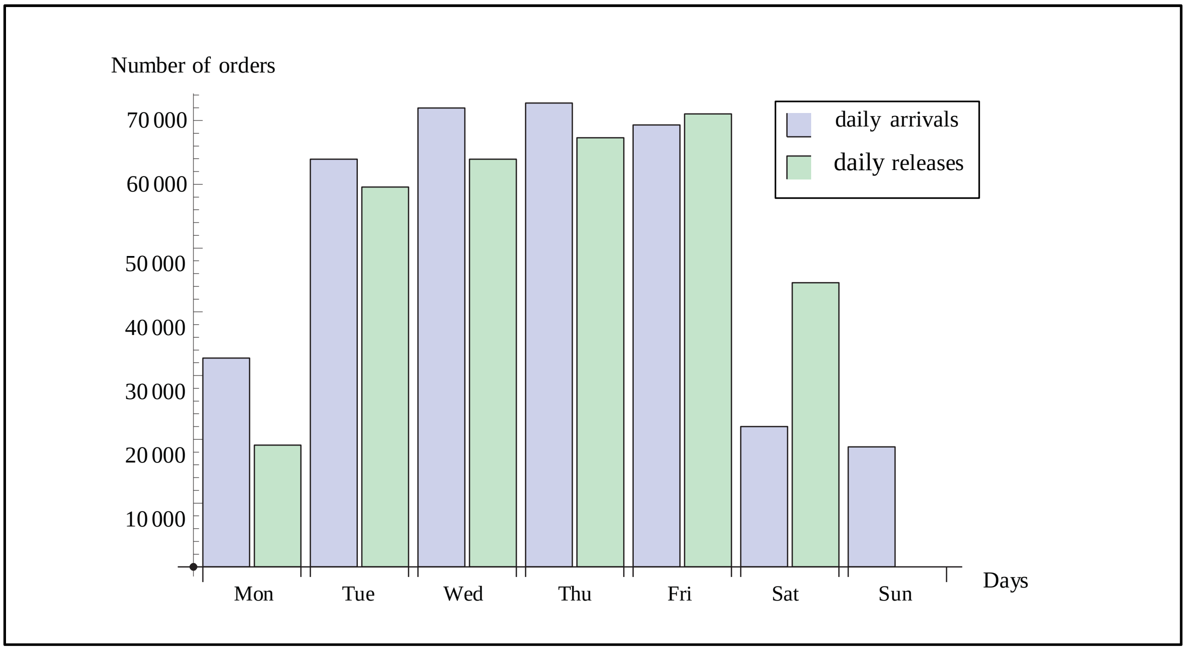

Figure 3. Arrivals and Releases for Each Day of the Week

As a consequence of service requirements and order release imbalances, we observed that NSD was highly variable throughout the length of study, within a range of [22.7%, 100%]. DDSP managers reported that the average NSD was around 72% for the three-month period (75.0% in January, 57.9% in February, and 70.4% in March 2011; however, on some days NSD dropped below 60% (Figure 4).

In addition to daily workload fluctuations, the order stream within a day is also non-stationary. For example, many customers place their orders around 0800 and 1800 (Figure 5a). Figure 5b shows the scheduled wave release times at DDSP at the time of our study—0000, 0400, 0930, and 1600. In addition to these scheduled releases, orders were occasionally released manually at around 0700 and 0900 to balance the workload.

Although arrivals within a day are almost continuous, the majority of releases are at discrete times: around 0001, 0400, 1000, and 1600. These release times correspond to planned releases and constitute the majority of orders. At other times, orders are released manually by wave planners in order to maintain high picker utilization and to process urgent orders.

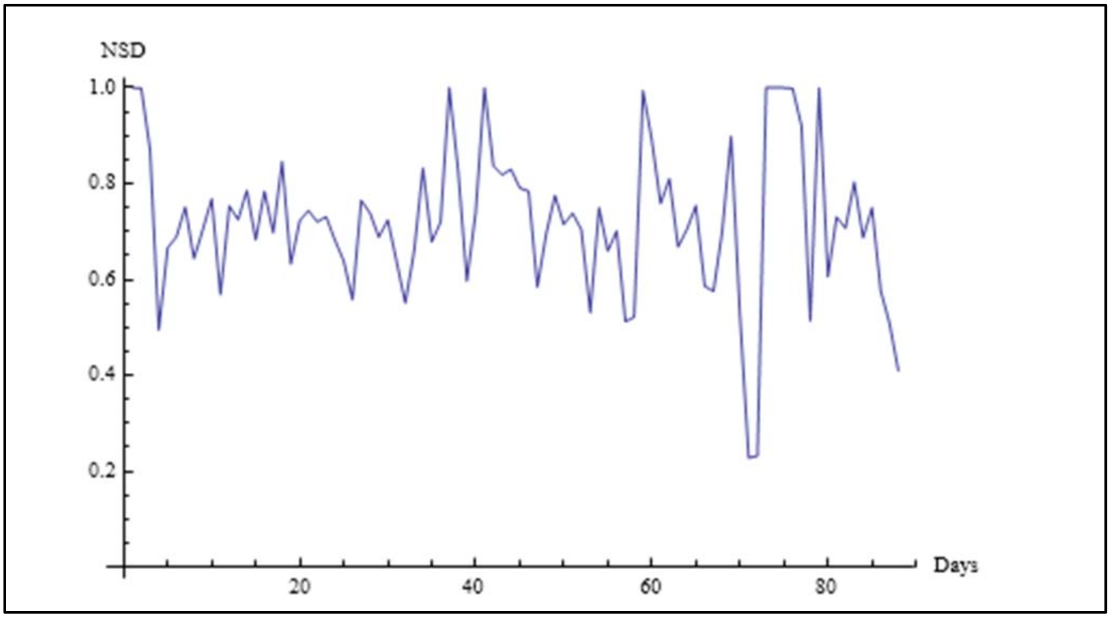

Figure 4. Recorded Daily NSD at DDSP

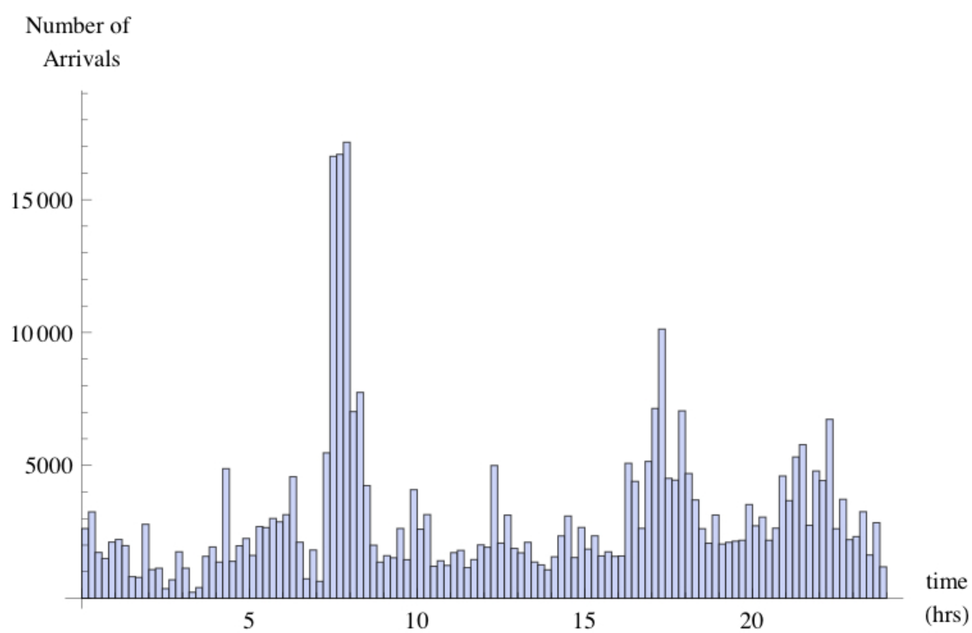

Figure 5. Number of Orders Arrived Within a Day

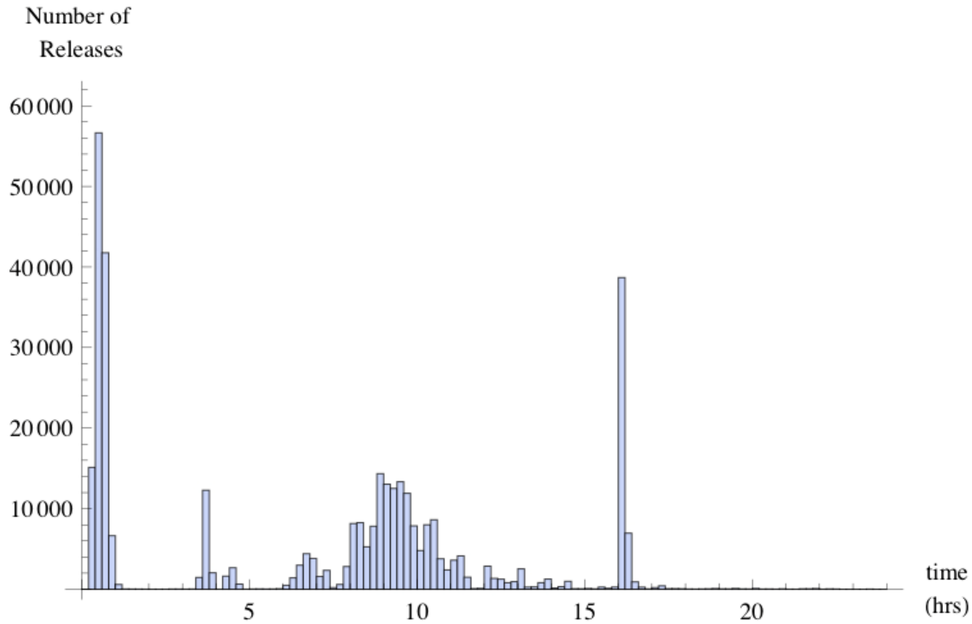

Figure 6. Number of Orders Released Within a Day

Problem Definition

Managing the release of work to the system in order to improve NSD and maintain internal operations is a difficult task, to say the least, and especially so in the presence of waves.In a typical distribution operation, including at DDSP, there are two–six waves per day, depending on the workload and number of destinations that must be served.

DDSP has a fixed capacity every day, which depends on the size of the workforce and is expected to be sufficient to process required orders. Arriving orders accumulate in a Warehouse Management System (WMS) virtual queue until the next wave is released, at which time the quantity of orders in that wave decreases at rate μ until the wave is complete.Waves in this model are not allowed to overlap; that is, the current wave must be complete before a new wave can begin. While the pickers are working a wave, orders in the next wave accumulate, and the cycle continues. Figure 6 illustrates the inventory graph in each of three waves.

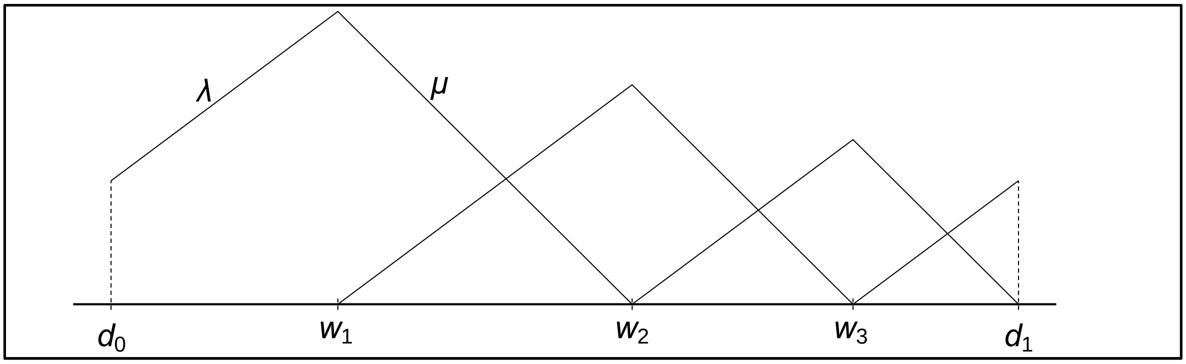

Figure 7. A Three-Wave System

Note. Variable $d_i$indicates the deadlines on Day i; and w_i is the release time for wave j.

Order picking in most areas at DDSP (and in the majority of commercial operations) is accomplished with *picker-to-parts* operations, in which pickers travel to the items. A picker-to-parts system typically has an associated minimum cycle time, based on the time required to walk a picking route (the minimum time to complete packing and shipping operations can be included in this fixed time component).This implies that there is a minimum time between wave releases, and also that pickers are less productive in smaller waves.(This observation is certainly true in practice.) Too few waves means orders arriving near the deadline have no chance at all to make the deadline, but too many waves means reduced capacity in operations. Lower capacity necessarily leads to longer waiting times and consequently lower NSD.

Already we have a very complex control problem: determine the number and the timing of waves to maximize NSD. Together with the workload and the size of workforce fluctuations,this problem becomes more complicated.It is our experience that the current number and timing of order releases is based on intuition and management experience. Could NSD be improved if the release times were changed? What level of benefit is possible?

We know of no scientific studies that have investigated systematic wave planning in the presence of deadline-oriented operations (Çeven and Gue [2013a] provide a literature review). We believe our work is the first comprehensive study of wave planning in a distribution environment, and certainly the first that addresses deadline-oriented operations.

Optimal Wave Release Policies

We first discuss some major results from Çeven and Gue (2013a), in which the arrival process is assumed to be stationary with rate λ orders per unit time. The reader is encouraged to refer to that paper for technical details and full model development.

To maintain stability, the server's capacity is assumed to be $\mu > \lambda$, where λ is the average arrival rate.

Çeven and Gue (2013a) proposed a fluid approximation model in which individual orders are indistinguishable. The fluid model is appropriate for a number of reasons:

- DDSP operations are too complex to analyze with exact models. An exact model would be intractable even with the strongest assumptions.
- We are interested in systems with stationary and non-stationary arrivals.
- The latter poses an especial challenge for exact queuing analysis.
- Our measure of interest (NSD) is not a typical metric of interest for pure queueing models. However, we can determine NSD with a fluid approximation.

As the authors (Çeven & Gue, 2013a) stated, when the ratio of average processing time to length of the final wave is less than about 1/3 (as in the case of DDSP), the fluid model approximates the system NSD within less than a 1% error.

We began with the simplest version of the problem, in which there is a single class of orders and the goal of finding a single optimal wave release time that maximizes NSD. By definition,

$$\text{NSD} = \frac{\text{\# orders worked today that arrived today}}{\text{\# orders that arrived today}}. \tag{2}$$

A workload of λ orders arrive, and the system (the server) processes orders with a capacity of μ. The server is busy λ/μ of the time for processing orders, which is utilization ρ. When there is a single wave, completion of the wave is at time $w_1 + \rho$. (Throughout, we scale a 24-hour period to a unit-length of one.) Çeven and Gue (2013a) show that the server should complete the wave exactly at the deadline, and, therefore,$w_1 = 1 - \rho$. Thus the optimal NSD for a single wave system is $w_1^* = \text{NSD}^* = 1 - \rho$.

The above result for a single wave can be generalized to multiple wave systems: The final wave ends exactly at the deadline in an optimal solution. Any idleness between wave completion and a release reduces the capacity. Therefore, as expected, the server should be idle only between the deadline and the first wave of the next day (a formal proof is provided in Çeven and Gue [2013a]).

Çeven and Gue (2013a) built a numerical procedure that finds a closed form for optimal wave release times for a system with stationary arrivals:

$$w_j = \begin{cases} \frac{j-1}{N} & \text{, for } \lambda = \mu \\ 1 - \frac{\rho^j - \rho^{N+1}}{1-\rho^N} & \text{, for } \lambda < \mu \end{cases} \quad (3)$$

Figure 8 shows the optimal release times when system utilization $\rho = 0.75$. Each horizontal line shows a different system with a different number of waves (from one to eight). As the number of waves increases, the first wave release time does not change, and because the last wave release time approaches the deadline, NSD is improved.

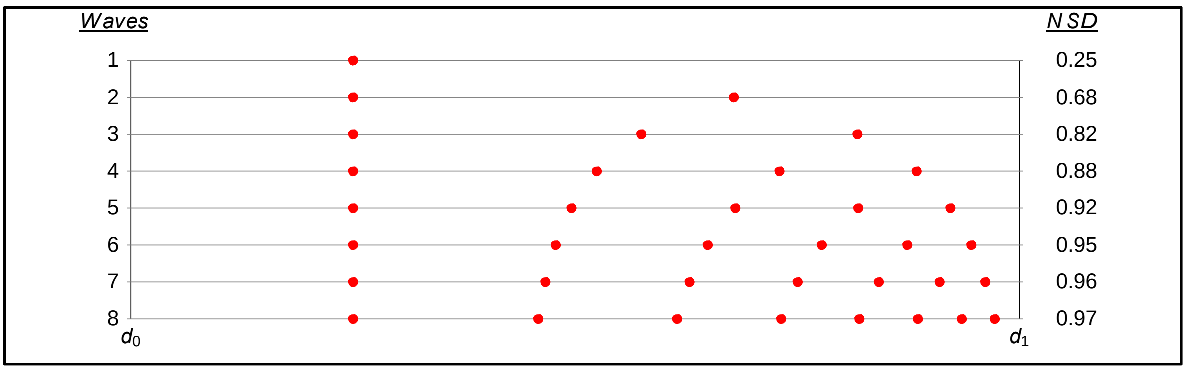

Figure 8. Optimal Wave Release Times for $\rho = 0.75$
(Çeven & Gue, 2013a)

Note. Although based on a simple operational model, Figure 8 offers useful insight for operations at a DLA distribution center: More waves produce a higher NSD, and waves should decrease in size as a deadline approaches.

The initial model for multiple wave systems does not address a requirement for minimum cycle time. For example, when there are eight waves (the system shown on the bottom in Figure 8), the server cannot complete the last wave before the deadline if this fixed time plus the processing time is longer than the length of the wave. As a consequence, Çeven and Gue (2013a) modified the initial model to incorporate a minimum cycle time component.Because each picking tour of length T means non-productive time, a system with N waves will consume NT time, and, therefore, the number of waves in a multiple-wave system has an upper bound of

$$N^* = \frac{1 - \lambda p}{T}. \quad (4)$$

The initial models also assume stationary arrivals. In practice, orders arrive in a non-stationary stream (Figure 5a shows this complication for DDSP). To address non-stationary arrivals, Çeven and Gue (2013a) again assumed a constant workforce size represented by capacity μ and replace λ with $\lambda(t)$ in the models (t denotes discrete time intervals).

As we discuss in the section Order Flow Analysis, daily workload is also uncertain, which, combined with the unexpected changes in workforce levels (due to absence, for example), leads to uncertainty in utilization. Çeven and Gue (2013a) discussed a procedure to adjust the release times in the presence of uncertainty.These modified models could be used at DCs such as DDSP, if changes in workforce levels and arriving workload can be estimated. This level of data collection was beyond the scope of our project.

Simulation Studies

We built simulation models for multiple reasons. Our first objective was to test the validity of the analytical models in Çeven and Gue (2013a). We modelled a simple version of DDSP operations and estimate the simulated NSD. We also compared optimal policies with an intuitive policy. After verifying the analytical models in a simulated environment, we considered DDSP's non-stationary arrival stream to validate the mathematical models.Because DDSP operations are quite complex and we lacked required data on processing times, we used Monte Carlo simulation to estimate the performance of DDSP operations in addition to the discrete event simulation model.

Model Verification

We modelled the order fulfillment system as a three-stage queueing system corresponding to the picking, packing, and shipping processes.We assumed 20 servers per stage and identical exponential processing time distributions in each stage. This choice was arbitrary, of course, but in the absence of real data, we had no justification for another choice.Arriving orders are stored in a virtual queue and released in the next wave.Once an order is released for picking, available workers start picking orders.Completed orders are sent directly to packing and then to shipping.Because daily workload at DDSP varies, we tested different levels of utilization,ρ = 0.5, 0.75, and 0.95. We adjusted the (exponential) processing rate to maintain the appropriate utilization. We assumed four waves per day, as in the operations at DDSP at the time of the study.

Using a stationary arrival stream, we determined the optimal release times for a single class, four wave system for each utilization level. Optimal release times suggested NSD would be 96.7%, 88.6%, and 78.1% for ρ = 0.5, 0.75, and 0.95, respectively. We inserted the release times into the SIMIO simulation software and ran the model for 30 simulated days, with three days of warm-up and 100 replications.The analytical model approximates the corresponding system's NSD within 1%. We refer the reader to Çeven and Gue (2013b) for the detailed simulation results.

Recall that the hourly arrival stream is non-stationary at DDSP and Çeven and Gue (2013a) only provided a closed-form optimal solution for stationary arrivals.However, the authors claimed that the same procedure can be applied to non-stationary streams to find optimal release times (not in closed form). We followed the procedure in Çeven and Gue (2013a) to determine optimal wave times using DDSP's non-stationary stream.

Monte Carlo Simulation

We attempted to extend the discrete simulation models to reflect real operations at DDSP, but for a number of reasons, we were unable to accomplish this goal directly. Note that DDSP's DTK operations constitute only a portion of each day's workload, so it was impossible to assess capacity devoted to these orders. Another complication for exact modeling was the number of classes that are processed in waves. DDSP has one set of release times, but several deadlines (UPS, FedEx, LTL carriers) each day, so their release times are defined to balance competing objectives.DDSP also receives orders days in advance of when they are scheduled to ship, so many orders remain in queue until near their deadline. Therefore, we built a substitute model—an NSD calculator—to duplicate the performance against NSD. (We used a spreadsheet to estimate the performance.) We developed this tool to serve as a black-box calculator. The tool is not, in fact, a simple calculator, but a Monte Carlo simulation model that can predict the service performance of a given set of wave release times. Our simulation procedure is illustrated in Figure 9.

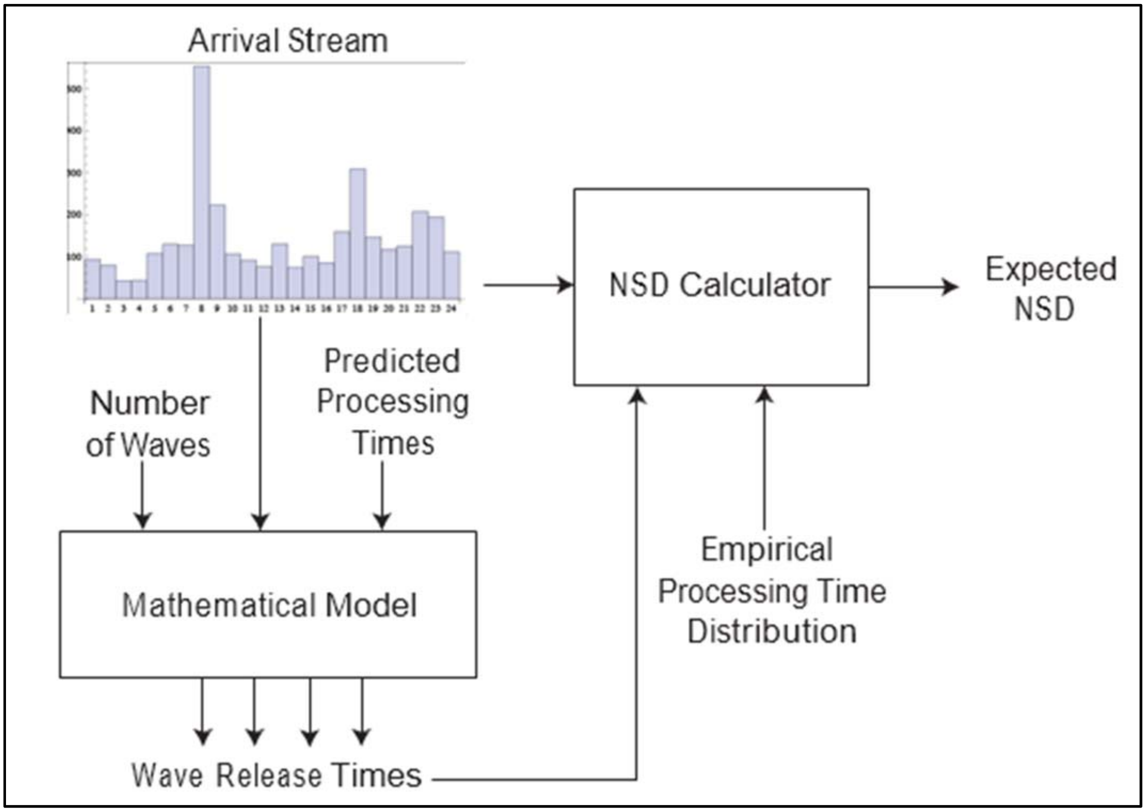

Figure 9. Conceptual Model of an NSD Calculator

The calculator works as follows. It pulls the order arrival information (arrival date and time) from the data set and generates a random sojourn time based on the empirical distribution. For a given number of waves and release times, we simulated the system and predicted the expected NSD. The model produces an expected NSD for a given set of release times and generates the maximum when the given set is optimal.

To verify the accuracy of the calculator, we inserted the current release times into the simulation and tested them. The calculator estimated the actual NSD as 68.7%; managers at DDSP reported that overall system performance on NSD was around 70% over the three-month period. The results confirmed that the calculator estimates the actual NSD with an acceptable level of error.

Summary and Recommendations

We have addressed order release problems in order fulfillment systems and shown that setting wave release times properly can improve NSD, and, thus, the operational availability of supported systems. In order fulfillment systems such as those operated by DLA, order releases should be timed to accommodate daily deadlines.

We believe that the outcomes of our research could provide important insights to improve the service performance of DLA depots. We showed that operations should begin the work as late as possible in order to allow as many orders as possible to make it into the wave. We showed that the server should be idle only between the deadline and the first release on the following day. This proposition implies workers will be more productive by disallowing idleness. Further, we showed that consecutive wave lengths should be smaller as the deadline approaches. The optimal policies also suggest that the last wave should finish exactly at the deadline.

This insight also suggests a way to determine optimal cutoff times. NSD can be approximately 100% if the cutoff time is set to the time of the last wave release time.This policy is easy to implement for operations having a single, daily deadline, but is inappropriate for more complex operations such as those at DDSP, which has multiple deadlines each day.

For stationary arrivals, optimal release times result in the same unworked inventory at the deadline each day, and, therefore, the workload for each wave is consistent from day today. By contrast, a set of arbitrary wave release times results in an unworked inventory at the deadline that is different each day.

In the presence of workload and workforce fluctuations, our procedure adjusts the release times to maximize expected NSD. For systems with multiple waves, release times should be earlier, which reduces the maximum possible NSD, but hedges against the risk of much lower NSD in case a high workload appears (details in Çeven and Gue, 2013a).

We validated our analytical models with simulation models. We showed that releasing waves optimally improves NSD and, thus, operational availability.Fortunately, changing the number and the timing of waves is as easy as making programming changes in the WMS. The results would be immediate after the change.

As we discuss throughout the report, instances of the problems we addressed are faced by the depots of the DLA, which includes more than 20 major depots and hundreds of distribution centers around the globe. Depots of the DLA are evaluated daily on their performance against the NSD metric, in addition to other measures. The largest depots, such as DDSP, operate waves in a way similar to the systems we studied. Therefore, the results we produced in this research should be of interest to dozens of commands in the DLA, and they have the potential to improve the service performance of order fulfillment systems.

Additional research is needed to deal with the complexities of current DLA operations, but the results of this research are, we believe, a helpful beginning to more complex models.

References

Çeven, E., & Gue, K. (2013a).*Optimal order release strategies and cutoff times for order fulfillment systems with deadlines* (Working paper).

Çeven, E., & Gue, K. (2013b). Wave release strategies to improve service in order fulfillment systems. In *Proceedings of the10th Annual AcquisitionResearch Symposium*. Retrieved from http://www.acquisitionresearch.net

Department of the Navy. (2003, June). *Operational availability handbook: A practical guide for military systems, sub-systems, and equipment*(OPNAVINST 3000.12A). Retrieved from http://doni.daps.dla.mil/directives/03000%20naval%20operations%20and%20readiness/03-00%20general%20operations%20and%20readiness%20support/3000.12a.pdf

Doerr, K., & Gue, K. R. (in press). A Performance Metric and Goal Setting Procedure for Order FulfillmentOperations. *Production and Operations Management.*

THIS PAGE INTENTIONALLY LEFT BLANK

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net