Capacity Management for Low Cost Storage

by
Omar Kahil

Bachelor of Engineering, Civil and Environmental Engineering, American University of Beirut, Lebanon, 2015

Submitted to the MIT Sloan School of Management and the MIT Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degrees of

MASTER OF BUSINESS ADMINISTRATION
AND
MASTER OF SCIENCE IN CIVIL AND ENVIRONMENTAL ENGINEERING

in conjunction with the
LEADERS FOR GLOBAL OPERATIONS PROGRAM

at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
June 2021

Signature of Author...
MIT Sloan School of Management
MIT Department of Civil and Environmental Engineering
May 14, 2021

Certified by...
Dr. Stephen C. Graves, Thesis Supervisor
Abraham J. Siegel Professor of Management Science, MIT Sloan School of Management

Certified by...
Dr. David Simchi-Levi, Thesis Supervisor
Professor of Civil and Environmental Engineering, Department of Civil and Environmental Engineering

Accepted by...
Maura Herson, Assistant Dean, MBA Program
MIT Sloan School of Management

Accepted by...
Colette L. Heald, Professor of Civil and Environmental Engineering
Chair, Graduate Program Committee

THIS PAGE IS INTENTIONALLY LEFT BLANK

Capacity Management for Low Cost Storage

by

Omar Kahil

Bachelor of Engineering, Civil and Environmental Engineering, American University of Beirut, Lebanon, 2015

Submitted to the MIT Sloan School of Management and the MIT Department of Civil and Environmental Engineering in partial fulfillment of the requirements for the degrees of

Master of Business Administration

and

Master of Science in Civil and Environmental Engineering

Abstract

The growth that Amazon is projecting brings implications on the supply chain network, mainly, increasing capacity needs that can be addressed by adding Amazon Robotics (AR) Sortable Fulfillment Centers (FC), which require significant investment. High per unit storage costs, driven by the technology used at the AR-Sortable FCs, along with stock keeping units (SKUs) with high dwell time have created an opportunity to leverage low cost storage (LCS) nodes upstream of the FCs. These nodes reduce the number of required future AR-Sortable FCs allowing for significant savings.

Amazon conceived its first LCS node to address the challenge of high safety stock requirements and costly holding overheads. This solution proved that pooling excess inventory upstream improved turns at the FCs and reduced storage related fixed cost. The LCS node is now established for all excess inventory across imports and domestic retailing businesses which would provide opportunities for additional free cash flow savings.

LCS receives inventory from three flows: Asia Pacific consolidation node, US consolidation node that processes overseas shipment, and domestic. LCS has been experiencing a high backlog meaning trailers waiting at LCS yards to have their freight processed into the sites for a prolonged period of time. A high backlog can cause added out of stock risk, carrier fees and disruptions, and for units to dwell at the LCS below the required period to breakeven with the added processing cost at the sites. The backlog is driven by the fact that LCS nodes have instances where the amount of freight arriving is higher than what can be processed into the facilities.

To support LCS in its capacity management efforts, this thesis explores the redirection of trailers from LCS nodes towards the fulfillment network during instances of high backlog. In addition, the effort will include balancing the backlog at LCS by setting processing capacity (i.e. mechanical and labor capability to transfer freight from trailers into facilities) constraints on incoming arcs into LCS nodes. This will contribute to achieving cost savings by prioritizing inventory that will spend the bigger portion of its dwell time at the LCS nodes, and support the mitigation of out of stock risk by redirecting inventory with low excess coverage in the fulfillment network.

Thesis Supervisor: Dr. Stephen Graves
Title: Abraham J. Siegel Professor of Management Science
Thesis Supervisor: Dr. David Simchi-Levi
Title: Professor of Civil and Environmental Engineering

THIS PAGE IS INTENTIONALLY LEFT BLANK

Acknowledgments

It is my pleasure to acknowledge the roles of numerous contributors who were essential for the completion of my thesis.

First, I would like to express gratitude to my advisors, Dr. Graves and Dr. Simchi-Levi. Their input was very valuable while making sense of all the intricacies in the study. The discussions we had helped me shape this project and develop solutions for the challenges faced by Amazon.

I would also like to thank my managers at Amazon, Siddhartha Trivedy and Sameer Devji. As much as Amazon is a data driven company, my managers' constant support and guidance was unmeasurable. I would like to appreciate the Supply Chain Innovation and LCS teams at Amazon for welcoming me and supporting me throughout my internship.

Sincere gratitude to Carl Berger for championing my project and for his thought leadership and mentorship on and off the job.

I would also like to thank Amazon for their sponsorship and to the Leaders for Global Operations program at MIT for their support. It was an incredible opportunity to work with a pioneering company and be exposed to exciting research.

I would like to thank my friends who supported me on this journey and were always there for me during the highs and lows. Dana, Marc, Daniella, Adil, Obada, Mostafa, Majd, Abdulhadi, Dima, Adam, and Jessica, I cannot thank you enough for being my second family.

I want to express my immense gratitude to my family. My parents, Adnan and Fadia, whom I owe all my accomplishments to. Thank you, for your unconditional love and encouragement. Thank you, for instilling in me values that I live by; values that push me to become the best version of myself. I hope I can give back a fraction of what you provided me and my siblings. Maher, Samer, and Samar, thank you for being my support system and for making me feel like, no matter where we are in this world, I am protected. Finally, Zeina, thank you, for being my guardian angel. I know you are watching over me and I hope I am making you proud.

THIS PAGE IS INTENTIONALLY LEFT BLANK

Table of Contents

List of Figures

List of Tables

Commentary on Amazon.com Proprietary Information

In order to protect information that is proprietary to Amazon.com, Inc., the data presented throughout this thesis has been modified and does not represent actual values. Data labels have been altered, converted or removed in order to protect competitive information, while still conveying the findings of this project.

1. Introduction

High per unit storage costs, driven by the technology used at the AR-Sortable FCs, along with stock keeping units (SKUs) with high dwell time have created an opportunity to leverage Low Cost Storage (LCS) nodes upstream of the FCs. These nodes reduce the number of required future AR-Sortable FCs allowing for significant savings. Amazon conceived LCS nodes to address the challenge of high safety stock requirements and costly holding overheads. This thesis seeks to highlight the operational challenges faced by inbounding inventory at LCS nodes and determining the positive business impacts of making point in time decisions to select inventory in a capacity-constrained environment to optimize cost savings and minimize FC out of stock. This study is based on a six-month research work carried out at Amazon, Seattle as part of MIT's Leaders for Global Operations (MIT LGO) curriculum. This chapter focuses on presenting an overview of the problem statement and the motivation to address it. In addition, the author also presents the research methodology used in the study.

1.1. Amazon.com Overview

In 1994, Jeff Bezos incorporated Amazon.com. Based on his research, he concluded that books would be the first product that best fits an online platform presence at the time. However, Bezos pointed from the start that this concept was more than just a retailer of consumer products. He reasoned that Amazon.com was a technology company with the aim of transforming online transactions for consumers and redesigning a product's supply chain between producers and customers. Bezos aimed for rapid growth and understood the immense potential of the Internet. In fact, Amazon reached more than 180,000 customers by the end of 1996 and jumped to 1,000,000 consumers by the end of 1997 [1]. The firm quickly targeted other spaces and began international operations with the acquisition of online booksellers in Europe. At the end of 1999, the marketplace that is Amazon.com expanded its offering to include an array of consumer products. The rapid growth and Amazon's acquisition strategy were funded by private investors and an IPO that raised $54 million on the NASDAQ market. As aforementioned, Bezos intended for Amazon.com to go beyond just being an online retailer and position itself as a technology company. It has expanded into a complex web of businesses and subsidiaries. To underline the point, the company ventured into cloud computing with Amazon Web Services (AWS) which has recently surpassed $40 billion in annual revenue with AWS's share of the cloud infrastructure market at 45%, more than double its closest competitor [2].

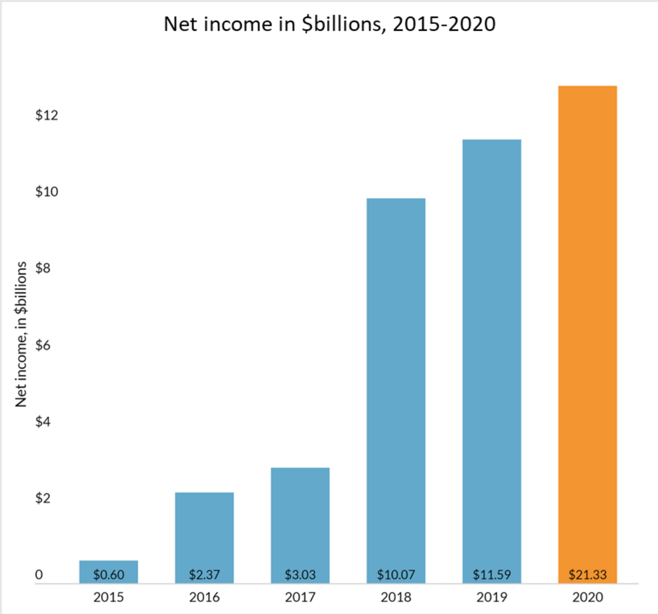


Figure 2. Amazon's growing profits

Amazon has witnessed an immense run over the past 27 years. 2020 produced blockbuster results for Amazon. The financial outcomes for Q4 of 2020 and the full year exceeded projections. The reported revenue for that quarter was more than $125 billion, an increase of nearly 44% from the prior-year quarter. (Figure 1).

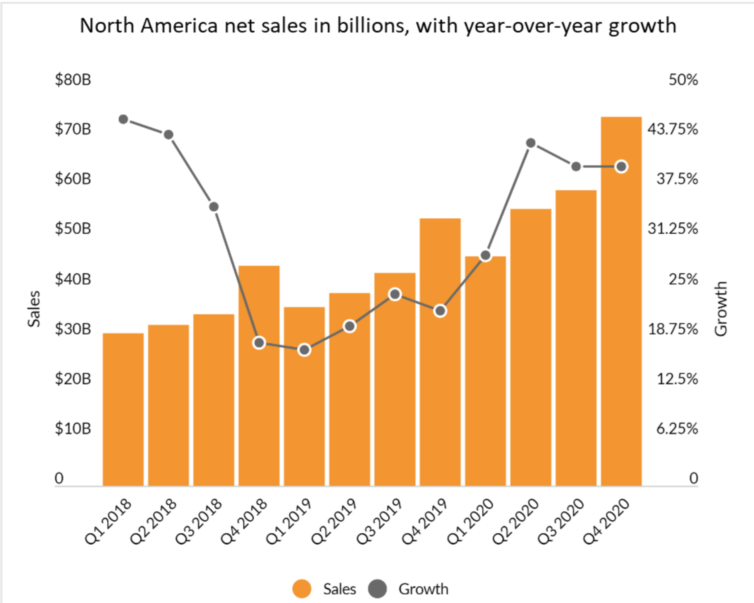


Figure 1. Amazon North America Net Sales and Growth

Amazon also managed to exceed its net income estimation (between $1 billion and $4.5 billion) given that it stated earnings of $7.22 billion for Q4 and $21.33 billion for the year, an 84.1% increase over 2019. (Figure 2) [3].

Through its exponential growth over the last decades, Amazon has persistently fixated on the three pillars of its flywheel model: (1) low cost, (2) unlimited selection, and (3) convenience and reliability of delivery. From a retail perspective, Amazon has managed to ensure customer satisfaction by focusing on advancement and innovation in its supply chain. The mixture of advanced IT, a wide network of warehouses, and outstanding transportation positions Amazon's supply chain as the most efficient at global level. The company operates more than 180 fulfillment centers worldwide. This international presence brings Amazon closer to the customer and allows it to complete orders in a cost-effective manner. Amazon also employs its transportation fleet of airplanes, trucks, etc. to enable its two-day delivery window for Prime members. Finally, Amazon's collaboration of innovative technology and automation with a

Figure 3. Amazon customer shopping frequency

deep network of employees allows the company to accomplish quick deliveries across the world. Through these supply chain related investments, Amazon managed to capture customer loyalty. (Figure 3) [4].

1.2. Amazon Supply Chain Network

Amazon acquires a significant portion of its imports products from its manufacturing sellers in Asia-Pacific. Similar to other Amazon items, this supply chain did not have a middle storage facility between the warehouse of the manufacturer and the fulfillment network in the US. There is a consolidation point in Asia Pacific, where SKUs from different sellers are received and then arranged into containers depending on their destination, and shipped across the world. The shipped items are received at inbound cross docks, where items are de-consolidated and shipped through trailers to different FCs depending on local demand forecasts [5]. Asia Pacific vendors that are geographically far from consolidation nodes arrange their products in containers and ship them to the US. These containers then get sent to local US consolidation nodes which sort products into trailers depending on their destination. These trailers get transported to FCs to fulfill customer demand. There are instances where these trailers pass by the cross docks for deconsolidation and reassortment depending on demand or operational considerations. With the introduction of an LCS node, items will be directed from consolidations nodes towards the LCS which will then feed the fulfillment network through FCs [5]. Inventory placement models determine which units need to be allocated to the LCS nodes given a set of purchase orders. The allocation runs at the consolidation nodes which place units into containers/trailers based on their corresponding destination.

As for domestic retailing, the majority of products come from domestic sellers. These sellers can send freight to LCS nodes which then replenish the network. Domestic retailing works with their own models to determine the placement of items into LCS nodes. Both imports and domestic retailing place items that are in excess into the LCS. At the moment, a portion of imports and domestic retailing SKUs leverage LCS.

Figure 4 showcases the inventory flow to the LCS nodes.

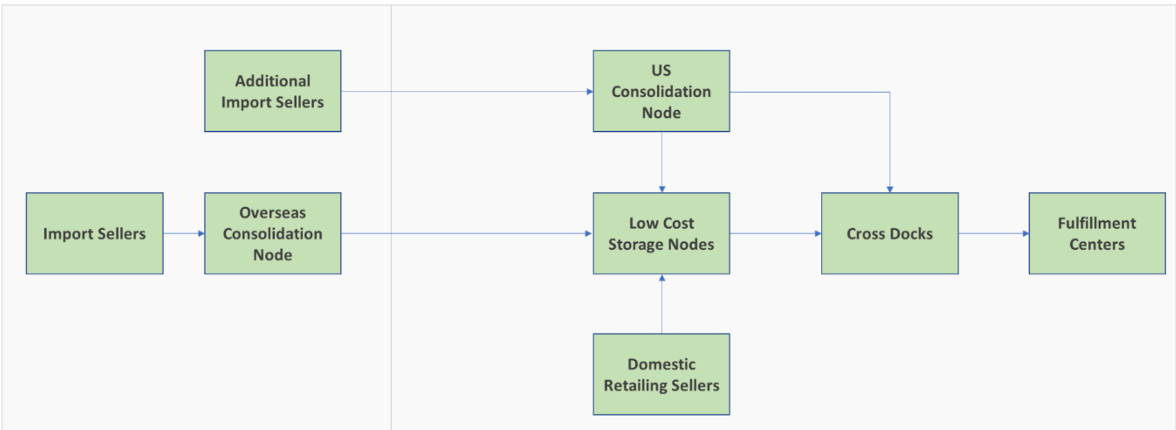


Figure 4. Supply Chain Network with LCS

1.3. Project Motivation

Across the board, incremental storage demand is growing faster than throughput demand. This is more so the case as Amazon transitions to a 1-day world. To support this logistical challenge, Amazon established the Low-Cost Storage (LCS) solution, which is a stand-alone storage network that complements its global and domestic supply chain and enables Amazon to increase storage capacity. LCS' driver of savings stems from a lower investment cost of LCS nodes vs that of AR- Sortable FCs (e.g. lower need for high tech equipment).

LCS was first conceived to support the growth of imports, which require storage for large amounts of inventory driven by variability in customer demand, long lead-time buying horizons, and large minimum order quantities (MoQs). Import units are mainly sourced out of Asia-Pacific. Import contracted manufacturers require Amazon to buy an economic manufacturing lot size, which is typically much larger than traditional retailing. As a result, inventory is purchased in excess, dwells for long periods of time, and is sub-optimally placed.

LCS stores sortable inventory upstream of Amazon's Inbound Cross Docks (IXD) network, which is leveraged to replenish and distribute to Fulfillment Centers (FCs) when required. Prior to LCS, Amazon's supply chain was a single-echelon network of FCs, which means that products flow from the seller's warehouse through IXDs to the FC network, which serves consumers. LCS expanded the businesses it serves to support segments of domestic retailing given that it has characteristics similar to imports with respect to dwell times. For domestic retailing's operations, sellers send inventory to storage facilities such as LCS. With the introduction of LCS, Amazon's multi-layer supply chain is positioned to introduce this new node type that can pool inventory in a low-cost environment and send it where and when it is needed.

Positioned upstream of the IXD network, LCS will be the first point of receive for freight intended for the domestic market. This implies that LCS nodes absorb the variability of arrivals for the network. In addition, the lengthy supply chain (freight source to FC receive lead time at an average of 140 days) introduces a variability in demand. This uncertainty introduces capacity challenges at LCS nodes that are translated into high inbounding backlog (i.e. long trail of trailers/containers waiting to be received at the nodes).

A high backlog is not encouraged because trailers waiting to be received might contain SKUs that have a low coverage in the network. This situation would increase out of stock risk and might drive economic losses. This thesis proposes a trailer redirection and capacity allocation solutions to effectively manage capacity challenges at the inbounding front of LCS. A clear understanding of how cost savings are achieved through current LCS nodes, how capacity is looked at, planned, and executed for these nodes, along with the challenges imposed by variability in lead time and demand motivated this thesis work.

1.4. Problem Statement

Today, LCS operates two sites and supports the sortable network with over 3 MM ft^3 of storage capacity. In 2021, LCS plans to launch additional nodes growing its storage footprint by more than a 100% to support the sortable network with storage capacity.

LCS absorbs the variability of arrivals for the network. This can lead to instances where arrivals are both below and above their weekly mechanical processing capacity (ability to take freight from trailers and containers into the sites). On occasions where weekly freight arrival exceeds available processing capacity for that week (more than 40% of the time in 2020), LCS operates with an inbounding backlog. An inbounding backlog of 3 days means that LCS needs 3 days to process the current freight (present in trailers and containers) in its yards into storage. Amazon has aligned with 3rd party operators of trailers and containers on a backlog threshold since this inventory is not immediately required in the network. However, 20% of the time this threshold is breached. A high backlog is not encouraged because freight waiting to be received might contain SKUs that have a low coverage in the network. In this situation, inventory to replenish the FCs will be unavailable given that it is in the yard. In addition, when backlog is above threshold, 3rd party operators will charge Amazon daily penalties for not returning their trailers and containers on time. In these instances, LCS does not have a methodology to select inventory that would optimize its entitlement and minimize out of stock risk in the FC network.

Over the past 35 weeks, LCS node #1 has experienced 90 days of inbounding backlog above the set threshold. This is despite LCS node #1 having sufficient space to store this inventory. This is due to the fact that weekly freight that arrived exceeds the corresponding inbounding capacity (i.e. transferring the freight into LCS storage facilities). Backlogs beyond the threshold cause (1) an added risk of out of stock (OOS) in the network, (2) negative fulfillment economics driven by units dwelling at LCS below the period required to breakeven, and (3) carrier fees and disruptions to trailer management.

With LCS expanding the businesses it supports and inherent variability in inventory arrivals from multiple sources, Amazon needs a mechanism to make a point in time decisions to minimize out of stock risk in the FCs by keeping the LCS inbounding backlog at a minimum.

Moreover, placement models do not take capacity at the sites into account (i.e. the weekly inbounding capacity at LCS nodes) . Placement models are run by Amazon at the freight supply source (e.g. Asia-Pacific) and determine from a pile of inventory there, the destination of each item (e.g. LCS node vs directly to the FC, which LCS node). This contributes to instances where the amount of freight arriving is higher than the site can process. At the moment, there is no current logic to allocate capacity between incoming volume from its different freight flows.

Today, when an inbounding backlog persists, LCS randomly redirects trailers present in its yard towards the fulfillment network through the IXDs to maintain a backlog below the threshold. In 2020, it has intentionally redirected over 3MM units to the fulfillment network. The absence of a redirection logic diminishes the cost savings potential that can be achieved by LCS and increases out of stock risk.

1.5. Project Hypothesis and Methodology

To address the capacity challenges currently faced by LCS nodes, the author's hypothesis suggests developing a logic for redirecting freight intended for the LCS network in instances where backlog is over the threshold, and introducing arc (freight sources and LCS nodes) capacity signals to current placement models, which assign inventory to the US network at the freight source (e.g. Asia-Pacific).

To support LCS in its inventory management efforts, the author will develop a model to identify trailers to redirect from LCS by minimizing supply chain costs across the network. In addition, this effort will include designing a logic to distribute available inbounding capacity at LCS, across the different freight flows, which will serve as a metric to be taken into consideration as a constraint by placement models. To make this capacity allocation, the author will analyze the freight characteristics of the different supply sources and dive into the existing operational and logistical conditions and determine their implications on capacity assignment. This work aims to balance inbounding backlogs by redirecting freight and setting capacity constraints on incomings arc. This will allow LCS to select inventory in a capacity-constrained environment to optimize cost savings and minimize FC out of stock. The effort also aims at selecting the optimal inventory that would ensure the backlog threshold is not breached but also store also store inventory best suited for LCS.

2. Background

This chapter provides the relevant background on how the LCS solution fits into Amazon's supply chain network, an overview of some of the financial and operational intricacies of this innovative concept, and the other factors that significantly affect this study.

2.1. Low Cost Storage Conception and Purpose

Amazon's traditional supply chain network is formed from the perspective of fast turning, high volume SKUs. However, not all imports and domestic retailing products have these characteristics. Explicitly, large minimum order quantities, and demand variability cause high inventory in the network. These products benefit from LCS to enable continuous investment and expand selection [6].

LCS' vision is to become the default storage solution for sortable excess inventory for both imports and domestic retailing. LCS nodes replenish inventory through multiple cross docks once inventory cover reaches a minimum threshold in the FC network according to demand profiles, transportation lead time, holding cost, and FC topology. With a reliable Service Level Agreement (SLA) to the cross-dock network, LCS will minimize out of stock risk and over time the AR-Sortable FCs will retain the right level of inventory, maintaining capacity to hold a wider range of selection. LCS' primary driver of savings stems from a low monthly operational cost per cube compared to an AR-Sortable FC [5].

Furthermore, LCS nodes make it possible to decrease the on-hand inventory and safety stock in the fulfillment network. In the future, LCS nodes will have an influence on current buying strategies. Instead of regularly having to re-order freight from an upstream seller, it becomes conceivable to procure larger quantities of some product at once, e.g. the supply of a full year, against bulk-rates, and to store this inventory in a cost-effective manner [7].

2.2. LCS Cost Savings Analysis

LCS' driver of savings stems from a lower investment cost of LCS nodes vs that of AR-Sortable FCs. To enable Amazon's cost accounting management, a monthly operational cost per cube compared is set for LCS and AR-Sortable FCs. Sending to LCS nodes adds transportation costs between LCS nodes and the cross-docks along with handling costs per unit at LCS nodes. As such, the difference in holding costs between LCS and the AR Sortable FCs needs to cover these costs. To breakeven with the additional per unit costs caused by sending a unit to LCS, a unit needs to remain at LCS for a specific duration. The duration is on average

20 days. A unit spending more time at LCS nodes than the breakeven period is what positions LCS as a cost-saving solution.

Example: Assume the difference in transportation/handling cost between LCS and the cross-docks and the Asia-Pacific consolidation node and the cross-docks through LCS is 0.2 $/unit. Storage at LCS must offset the 0.2 $ difference. Assume the storage cost difference between the FCs and LCS is 0.1 $/day/ft3. Assume a product volume of 0.1 ft^3. The breakeven period is then 20 days.

2.3. LCS Capacity Planning

Weekly processing capacity (number of units that can be received/sent out per week) for LCS nodes is estimated by the LCS team based on the site's mechanical capacity and labor planning. Processing capacity is limited by the mechanical capability of each site and is shared between inbounding and outbounding processes. Mechanical capacity is the hard constraint at these sites because even though labor planning can be flexible, processing units cannot exceed the mechanical capacity. Given that LCS nodes replenish the fulfillment network, outbounding is prioritized over inbounding to ensure customer demand is met. As such, the main contributor to estimating inbounding capacity is the forecasted units to be outbounded. Any variation in terms of outbound forecast based on real time demand updates will drive a high variability in inbounding capacity.

3. Current Situation Analysis and Problem Overview

We have discussed how the LCS fits in with Amazon's supply chain and went over the purpose of having these nodes and the benefit they contribute to the network and the company's profitability. In this chapter, we will go over an in-depth review of inbounding operations at the LCS nodes and discuss the challenges they are currently facing with respect to capacity, and formulate the problem this study aims to solve.

3.1. LCS Inbounding Inventory Flows

LCS addresses the high fixed cost and capital expenditure (e.g. building, equipment) headwinds arising from the storage of excess inventory in the core Sortable FCs. LCS node increase the storage capacity of the sortable network at a CAPEX per cube that is 20% of that of an AR-sortable FC.

Currently, LCS nodes support imports and domestic retailing businesses. They are well-positioned to serve these segments as they require low-cost storage to smooth the inbounding flow, optimize the number of FCs built based on outbound utilization versus storage requirement, and minimize out-of-stock risk by storing inventory ahead of demand [5]. The inventory flow into LCS nodes is split into three categories: Asia-Pacific Consolidation Nodes, US Consolidation Nodes, and Domestic Retailing. The split and additional spread metrics among these segments can be found in Table 1.

Table 1. LCS Inventory Flow Split Metrics (in Units/Week)

	Asia-Pacific Consolidation Node	US Consolidation Node	Domestic Retailing
Average Split	31%	54%	15%
Mean	388,430	676,620	187,950
Median	214,868	472,208	75,274
1st Quartile	114,795	227,632	-
3rd Quartile	354,900	870,250	164,306

The split between LCS' freight flows varies weekly. However, the US consolidation nodes capture the majority of the inventory arrival with 54%, followed by US consolidation nodes with 31%, and then domestic retailing with 15%.

3.2. LCS Inbounding Backlog Challenges

For import inbound (IB), Amazon leverages a placement system that allocates inventory to LCS at either consolidation nodes, by leveraging a model that minimizes total supply chain costs while ensuring customer demand is fulfilled. For domestic retailing Inbound, sellers can create a shipment in any quantity and send it to Amazon at any time. LCS sets an inbounding backlog threshold to meet carriers' service level agreements for returning trailers and containers. However, due to inherent variability in lead-times for import freight arrival, the backlog threshold at LCS is often breached. During the past seven months, the backlog at LCS sites was above the threshold 25% of the time and above 15 days 10% of the time. There are two main drivers for this backlog:

A. Inbounding capacity is dependent on outbound (OB), meaning the processing capacity at the site is split among inbound and outbound. Priority is given to outbound units to hold a 24-hour SLA to the cross-dock network and minimize out of stock risk. Hence, the forecast for inbound capacity also experiences low accuracy. The high variability in the forecast error limits the ability to send capacity signals to the placement model without risking low fullness metrics at LCS nodes. Table 2 details the different error metrics across all three forecasts.

Table 2. LCS Weekly Forecast Error Metrics – Error between actual numbers and the forecast a week ahead

Error between week-1 forecast and actual	Inbound (Incoming) Units	Outbound (Outgoing) Units	Inbound Capacity
Mean Percent Error	-3.20%	-63.89%	-56.25%
Mean Absolute Percent Error	57%	70%	81%
Mean Deviation (in Units)	65,154	(124,153)	(11,481)
Mean Absolute Deviation (in Units)	226,207	154,652	154,488

The LCS team is developing a machine learning (ML) model in 2021 to better predict arrivals, however, variability in lead time and demand will continue to drive an error in forecasting. Another reason why the placement model does not consider capacity is that LCS has not yet developed a logic to split weekly forecasted inbounding processing capacity at its sites among its three inventory flows (Asia-Pacific Consolidation Nodes, US Consolidation Nodes, and domestic retailing).

Figure 5 demonstrates the difference between the forecast of incoming units to LCS and the actual number of units that arrived at the nodes. The large difference in both metrics is a driver of the inbounding backlog at LCS nodes.

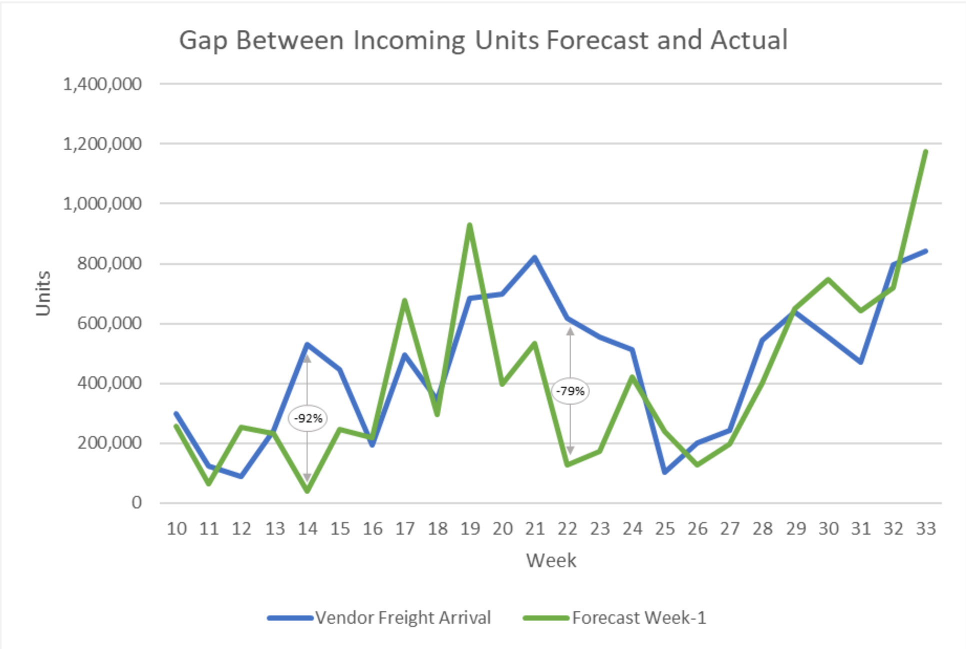


Figure 5. Weekly Forecast of Incoming Units vs the Actual Number of Units that Arrived

In the Figure 6, we notice the difference between the inbounding capacity forecast at LCS and the actual number of units that were received/processed into the site.

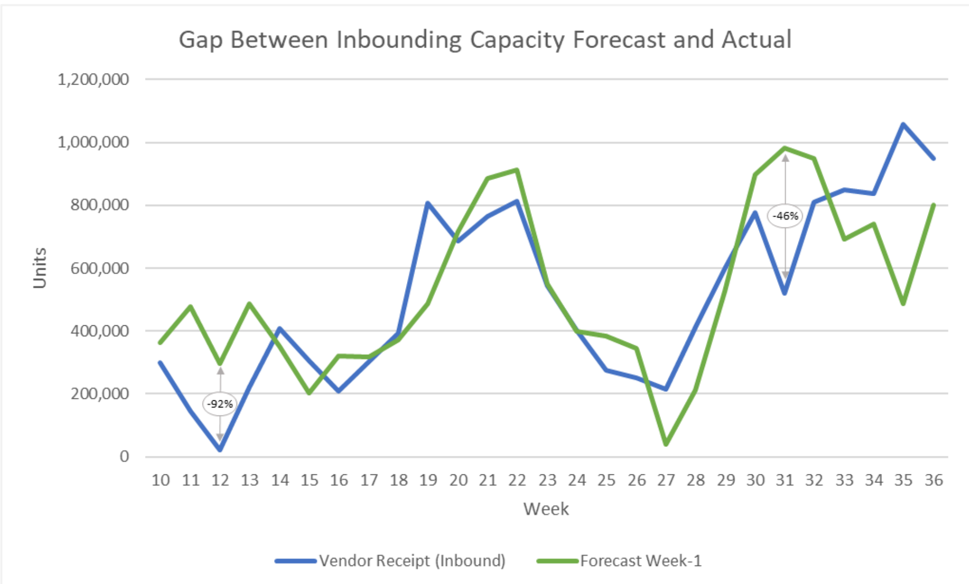


Figure 6. Weekly Inbounding Capacity Forecast and the Actual Capacity

B. The stochasticity in demand can drive LCS to outbound more units from its nodes than previously forecasted because of a decrease in coverage in the network. This unpredictability increases the backlog given that outbound is prioritized over inbound. Figure 7 represents the lead time distribution between supply origins and LCS and thus showcases the lead time uncertainty.

This long lead time also means that there is uncertainty in transit times. For example, the average lead time between Asia-Pacific consolidation nodes and LCS nodes is 41.7 days with a pooled standard deviation of 8.6 days. This uncertainty in lead time can contribute to a backlog at LCS sites, given that inventory might be received in a different period than the one estimated when the model ran. On the domestic retailing side, 80% of the domestic retailing shipments arrive within 30 days of the creation date, however, domestic retailing inbound does not have any mechanism to predict shipment arrivals from Amazon sellers. This translates into instances where LCS sites receive domestic retailing shipments beyond the expected freight, causing inbounding backlogs.

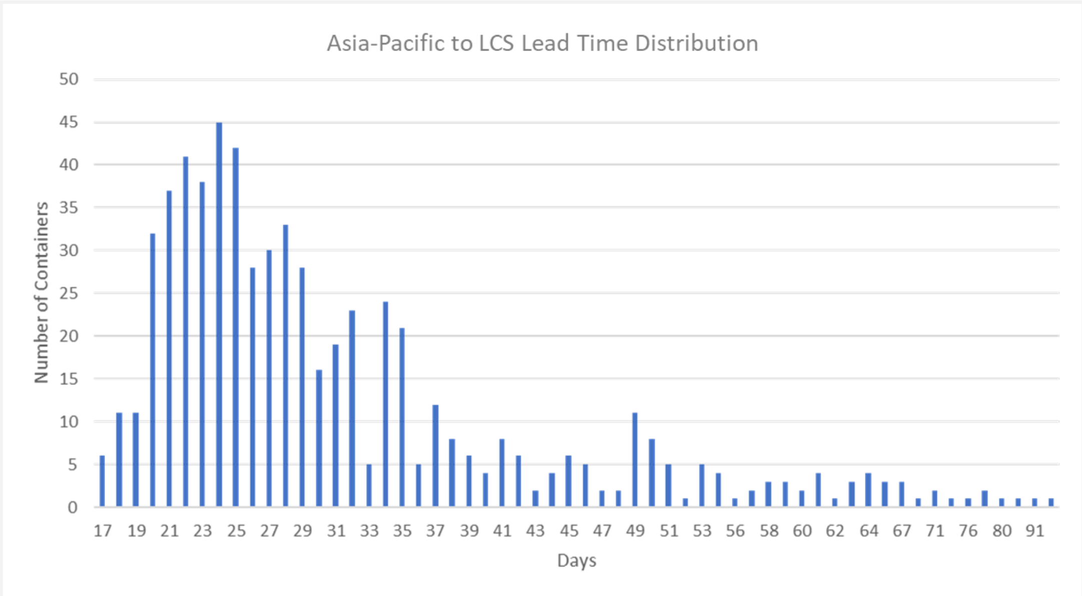


Figure 7. LCS Lead Time Distribution

3.3. Current LCS Backlog Reduction Lever

To reduce the backlog, LCS Operations can pull a lever to redirect trailers in its yard. The team communicates with S&OP and logistical teams and informs them of the quantity to be redirected into the fulfillment network or to other LCS sites. Logistical teams then follow a LIFO methodology to determine which trailers to be redirected. LIFO is used mainly for the logistical convenience of moving trailers. LCS has used this lever for seven weeks in the past seven months and redirected an estimated 3MM units towards the fulfillment network due to high backlog instances.

Given how costs are measured by Amazon, units with more excess weeks of cover in the network will drive more cost savings as they spend a higher portion of their dwell time at LCS nodes. However, using the current logic does not allow for the optimal cost savings to be achieved when redirecting trailers. Due to lead time and customer demand (forecast) uncertainty, SKUs arriving at LCS might experience lower than expected coverage in the fulfillment network. The current logic does not prevent keeping trailers, containing SKUs with low coverage, at LCS. This implies SKUs staying at LCS nodes below the breakeven period, which will drive economic losses in the network.

In addition, redirection from LCS' yard can only occur for trailers and not containers. Trailers head to LCS from the US consolidation nodes and domestic retailing while containers come from the Asia-Pacific consolidations nodes. This means that only US consolidation node and domestic retailing freight can be redirected from the yard. If containers are redirected, Amazon logistical teams face a difficulty in tracking them across the network which can cause containers to be lost or sent to wrong end destinations (e.g. FCs).

3.4. Problem Focus

This effort aims to build a capacity management process for LCS that will solve the backlog challenge. First, we will add a cost-based logic to LCS redirections. We will develop a holding cost-based model to capture maximum cost savings by identifying which trailers should be redirected from LCS' yard when the backlog is higher than the threshold. We will also set up a logic to distribute available inbounding capacity at LCS, across the different freight flows (Asia-Pacific Consolidation Nodes, US Consolidation Nodes, and domestic retailing), which will serve as a constraint to be taken into account by the imports and domestic retailing placement models. This will act as a measure to limit instances where more inventory is sent to sites that do not have the needed processing capacity. We will also develop a model for redirecting Asia-Pacific consolidation nodes containers prior to their arrival to LCS, given that operational constraints render it impossible to redirect them once they reach the yard.

With an increasing number of LCS eligible SKUs and more volume flowing to LCS, instances of high backlog are expected to increase [5]. A high backlog is not encouraged because trailers waiting to be received might contain SKUs that have a low coverage in the network. In this situation, an SKU might stay in LCS sites below the required breakeven duration and drive losses.

4. Literature Review

Businesses find managing inventory a difficult balancing act in times of volatility and uncertainty. Multi-echelon inventory systems have been identified as an innovative way to manage today’s complex, multi-stage, global supply chains. Uncertainty implies that organizations ought to continually review their capacity management methods in order to ensure customer demand fulfillment is not disrupted and that total supply chain costs are in check. In this chapter, the author will go over the benefits of multi-echelon systems, supply chain costs and network flow design and their influence on capacity management.

4.1. Multi-Echelon Inventory Systems

Controlling the supply flow from manufacturers to end customers has always been an essential problem. Management of all companies in any industry have acknowledged the strategic importance of supply chain management. Innovative leaders focused on effective inventory management as a key part of supply chain management. According to Forrester Research, the key differentiator between successful companies and ones that have sub-optimal performance is an ability to increase the inventory turnover. A multi-echelon inventory system has been identified to serve this purpose [8]. In fact, it drives the enablement to have shorter lead times to customers, making working capital available for profitability and supports in managing demand variability.

A multi-echelon system is a network composed of stages that are grouped into echelons. Stages can represent physical locations, items in stock or in processing activities [9]. Given the linkages among different stages, multi-echelon systems can be classified under several structures. Amazon is aiming to adopt a multi-echelon system. The current supply chain with respect to its LCS represents Amazon’s pursuit of a multi-echelon system. It can be said that this system follows an acyclic structure where every echelon is a possible site to store safety stocks of the transitional product produced at that stage.

The multi-echelon inventory system was discussed by Clark and Scarf [10]. Multi-echelon inventory management is setting rules or policies which drive the flows products through the network and which satisfy a set performance target. Multi-echelon inventory management is a cohesive approach that reflects the elements related to inventory optimization: cost, variability, service and complexity. It considers the intricate interdependencies between echelons as well as variables that cause excess of inventory [8].

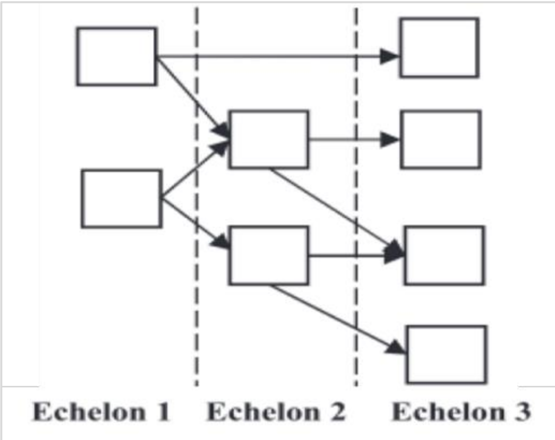


Figure 8. Acyclic Multi-Echelon System Topology

4.2. Supply Chain Costs

Inventory management is at the core of all supply chain and logistics management. There are many reasons for holding inventory. These include minimizing the cost of controlling a system, buffering against uncertainties in demand, supply, delivery and manufacturing, as well as covering the time required for any process. Having inventory allows for a smoother operation in most cases since it alleviates the need to create product from scratch for each individual demand. However, managing the total cost of the supply chain with respect to inventory (e.g. delivery) is key to drive profitability and achieve a competitive advantage [11].

Voordijk (2010) referred to the logistics costs of a supply chain as those involved in the management and storage of inventories, transportation and those incurred for physical distribution [12]. In their book, Silver, Peterson, and Pyke detail a Total Cost (TC) equation. It is typically used to make the decisions of how much inventory to hold and how to replenish. Total cost includes key cost components such as purchasing, holding, transportation, and shortage costs. We pursue decisions that minimize total costs and explicitly the total relevant costs. A cost element is considered relevant if it influences the decision at hand and it is controllable by some act [13]. As such, the exact shape of the total cost equation leveraged depends on the assumptions made in terms of the situation.

In this study, we focus on holding, transportation, and shortage costs given that the analysis is for products that have already been purchased. The holding costs are simply those costs that are required to keep inventory and include such things as storage costs, insurance, damage, obsolescence, and capital costs. The units are typically in terms of cost per unit of time. Transportation costs reflect the movement of products from one location to another and are typically in terms of cost per unit per arc. Finally, the shortage costs (also known as the penalty cost) are those costs associated with not having an item available when demanded.

4.3. Network Flow Model

Considerable research has been done in the field of distribution networks for multi-echelon multi-product supply chain. Among the original efforts targeting distribution systems was offered by Bowersox (1969) [14] [16]. where the author presents the idea of an integrated physical distribution system. Recently, mathematical models have been developed by researchers to address numerous sorts of problems that are frequently face by the practitioners. Tsiakis and Papageorgiou (2008) addressed the operation of multi-product, multi-echelon distribution networks [15] [16]. They developed a mathematical model that minimizes the total cost while taking into account the infrastructure and logistical conditions. Other research has also been conducted (Nikolopoulou and Ierapetritou – 2012) that formulates models aiming at the minimization of transportation, holding, and shortage costs subject to capacity and inventory balance constraints [16] [17]. Moreover, additional research was conducted to address variations of shipping and transshipment models. Federgruen and Zipkin (1984) developed a single period model. The model minimized the costs of transportation, shortage and storage [16] [18]. In this study, we will leverage the work that has already been done, along with Amazon's research advancement, to develop a time-index multi-commodity network flow model that can support redirection decisions and freight destination given Amazon's network design and inventory. We study the network flow and the corresponding costs given the various parameters of the problem (e.g. node demands, arc-flows).

5. Methodology

Given the aforementioned challenges and set hypotheses, this chapter contains an overview of the identified solutions to capacity management at LCS. The author describes the logic of each solution, its formulation and the corresponding outputs.

5.1. LCS Yard Trailer Redirection Model

5.1.1. Model Objective

The LCS network does not currently leverage an optimization methodology when undergoing the process of redirecting trailers from its yard and into the fulfillment network. The objective of this model is to provide a cost-based logic for trailer redirection selection. The model identifies trailers, from the backlog pool, that should be moved out of the yard to the fulfillment network to minimize total holding costs.

5.1.2. Model Rationale

The model considers holding cost at the LCS nodes and the FCs. Transportation is assumed to be instant between LCS node in question and the fulfillment network.

The model assigns a cost for each trailer in the backlog pool. To do so, the model picks a trailer to be sent to the FCs, computes its holding costs at the FCs, then computes the holding costs (LCS and FCs) for the other trailers that are assumed to remain at the LCS. We then repeat this process for every trailer. For each selected trailer, the model assumes that it is to be redirected to the FCs, through IXDs, while the remainder of the trailers are to stay in the yard and be received into the LCS node. As such, the model calculates the corresponding FC storage cost of the units that are part of the redirected trailer. Then, the model calculates the LCS and FC storage costs of the units that are part of the trailers that remain in the yard. For every selected trailer, the holding costs are summed up to determine a total holding cost.

The model then calculates a cost per unit for each selected trailer by dividing the cost assigned to that trailer by the number of items it includes. The LCS team sets the minimum number of units that need to be redirected. Using mixed integer programming, the model identifies which trailers should be redirected in order to minimize total supply chain costs while ensuring the required redirection volume (number of units) is met.

5.1.3. Holding Cost Calculation and Assumptions

The model calculates holding costs for the units involved in the computation by estimating the expected duration that each unit will spend in the LCS and the fulfillment network. To estimate the holding duration for each unit, the model takes information about the quantity in the backlogged trailers, the on-hand inventory at LCS sites and the fulfillment network, forecasted customer demand, future incoming purchase orders that will be received in the FCs, and the target weeks of cover in the FCs recommended by Amazon's inventory planning team.

The model encompasses the following assumptions and considerations in its logic: (1) replenishment between the FCs from LCS sites occurs in an equally distributed manner, which is the current method adopted by Amazon fulfillment. (2) Given that the model does not need to determine how units will be distributed across the FCs (this effort is conducted by the cross docks network team), the fulfillment network is considered as one cluster (3) The cost of moving trailers is not taken into consideration as that these costs are unavoidable given that every unit will eventually need to be moved to an FC (4) Consumption of units at LCS and the FCs is assumed to follow a First in First Out (FIFO) methodology.

With this adopted methodology, the cost-based model prioritizes the redirection of trailers containing units that have low coverage in the fulfillment network which would contribute to safeguarding against out of stock risk and minimizing economic losses.

5.1.4. Model Formulae

The model generates a total holding cost for each trailer assuming its redirected to the FCs by calculating the FC holding cost for this trailer and the LCS and FC holding cost for the trailers that remained in the yard to be received at the LCS node in question. The overall intent is to identify which trailers to redirect in order to reduce the backlog towards the required threshold. The author will go through the total cost calculation for a specific trailer.

5.1.4.1. Redirected Trailers Holding Cost

In this part, we calculate the holding cost of the selected trailer to be redirected to the FCs. Table 3 provides an overview of the parameters used in the cost calculation along with a description for each one.

Table 3. Redirected Trailer Holding Cost Calculation Parameters

Parameter	**Description**
C^{LCS}	*Holding cost at LCS expressed in dollars/ft^3/day*
C^{fc}	*Holding cost at the FCs expressed in dollars/ ft^3/day*
B_k^t	*Inventory level required in the fulfillment network (order up to level) for day t. It is the sum of forecasted demand for the next 4 weeks*
u_k^t	*Demand for day t*
R_k	*Redirected units from LCS' yard to the FCs for SKU $k \in K$*
V_k	*Volume of SKU $k \in K$ in ft^3*
$I_{fc\,k}^t$	*Inventory at the FCs for SKU $k \in K$ for day t*
$I_{LCS\,k}$	*Inventory at LCS for SKU $k \in K$ when redirection occurs*
x_k^t	*Number of units consumed for SKU $k \in K$ from the quantity that was redirected R_k*

Holding cost at Amazon is calculated by unit of time and volume. Let $I_{fc\,k}^t$be the combined inventory of SKU $k \in K$ in time (day) $t \in T$ at the FCs where T is 365 days. $I_{fc\,k}^0$is the inventory at the FCs excluding the redirected quantity from the LCS node. Daily ending inventory at the FCs is updated based on the following equation $I_{fc\,k}^{t+1}$=$I_{fc\,k}^t - u_k^t + x_k^t$ for all $k \in K$. Redirected units are assumed to be consumed once the existing inventory excluding the redirected units is no longer able to satisfy demand, i.e. if $I_{fc\,k}^t - u_k^t < 0$. x_k^t would equal that difference up until all the redirected units of an SKU $k \in K$ are consumed, i.e. when $\sum_{t \in T} x_k^t = R_k$. The corresponding holding cost of the redirected units, $RC1_{\,k}$, would be:

$$RC1_{\,k} = C^{fc} \times V_k \times \sum_{t \in T} x_k^t \times t \qquad \forall\, k \in K$$

Another cost component that we need to account for, is the additional storage period that this redirection implicates on the respective SKUs currently in LCS sites. Corresponding LCS units will remain in LCS sites for an additional time period up until the redirected units are consumed. The corresponding holding cost, $RC2_{k}$, would be:

$$RC2_{k} = C^{LCS} \times V_k \times I_{\text{LCS}\,k} \times \frac{\sum_{t \in T} u_k^t}{T} \qquad \forall\, k \in K$$

5.1.4.2. Non-Redirected Trailers Holding Cost

In this part, we calculate the holding cost for the trailers that remained at the LCS node given the selected trailer to be redirected. Table 4 provides an overview of the parameters used in the calculation along with a description for each one.

Table 4. Non-Redirected Trailers Holding Cost Calculation Parameters

Parameter	**Description**
C^{LCS}	*Holding cost at LCS expressed in dollars/ ft^3/day*
C^{fc}	*Holding cost at the FCs expressed in dollars/ ft^3/day*
VC^{LCS}	*Variable unit processing cost at LCS expressed in dollars/unit*
B_k^t	*Inventory level required in the fulfillment network (order up to level) for day t. It is the sum of forecasted demand for the next 4 weeks*
u_k^t	*Demand for day t*
P_k^t	*Purchase orders for SKU $k \in K$ arriving in day t*
V_k	*Volume of SKU $k \in K$ in ft^3*
$I_{fc\,k}^t$	*Inventory at the FCs for SKU $k \in K$ for day t*
$S_{fc\,k}^t$	*Inventory at FCs for SKU $k \in K$ for day t where day t is reset for each $x_k^t > 0$*
x_k^t	*Number of units consumed from LCS to FCs for SKU $k \in K$ from the quantity that is in the trailers that were not redirected*
y_k^t	*Number of units consumed from LCS to FCs for SKU $k \in K$ from the on-hand inventory excluding the trailers that were not redirected*
z_k^t	*Number of units consumed out of the x_k^t units. For this parameter, t is initialized at 0*

For the trailers that remained in LCS, the model calculates the duration of their contents' storage at the LCS node and then at the FCs and the corresponding holding cost. It will be assumed that the content of the redirected trailer is now part of the fulfillment network's inventory.

The ending inventory at the FCs is set as $I^{t+1}_{fc\,k}$=$I^{t}_{fc\,k} - u^t_k + P^t_k + x^t_k + y^t_k$. We are assuming that the network will be replenished by LCS whenever the inventory level at the FCs falls below the target inventory level of B^t_k and will be filled with units up until it reaches the target. It will also be assumed that the inventory that is already on hand at LCS nodes will be consumed before the inventory that is in the trailers at the yard. With this in mind, once the inventory at the FCs is below the target, LCS replenishes the network from the pre-existing inventory, $y^t_k = B^t_k - (I^t_{fc\,k} - u^t_k + P^t_k)$ while $y^t_k \geq 0$. The instance where y^t_k is negative, the model assumes that the inventory that is coming from the non-redirected trailers is leveraged. Once y^t_k is negative, it is zeroed out. $x^t_k = B^t_k - \left(I^t_{fc\,k} - u^t_k + P^t_k + y^t_k\right)$.

As such, the holding cost at LCS, $NRC1_{\,k}$, would be equal to:

$$NRC1_{\,k} = C^{LCS} \times V_k \times \sum_{t\,\in T} x^t_k \times t \quad + \quad VC^{LCS} \times \sum_{t\,\in T} x^t_k \qquad \forall\, k \in K$$

We also need to account for the FC holding cost of the freight that was part of the trailers. We will leverage the logic from the redirected trailers holding cost calculation. For every $x^t_k > 0$, we register the inventory in the FCs, excluding x^t_k, as $S^0_{fc\,k}$. Here *t* is initialized at 0 to serve as an indicator for how many days the incoming LCS units stayed in the FCs. The $S^t_{fc\,k}$ is updated based on demand using the following equation: $S^{t+1}_{fc\,k} = S^t_{fc\,k} - u^{t+1}_k$. When $S^t_{fc\,k} < 0$, the x^t_k will start to be consumed. $z^t_k = u^{t+1}_k - S^t_{fc\,k}$ represents the number of units consumed by demand out of the x^t_k units. $z^t_k \geq 0$.

With this mind, for each $x^t_k > 0$ the corresponding holding cost at the FCs would be equal to:

$$C^{fc} \times V_k \times \sum_{t\,\in T} z^t_k \times t$$

With the above equation in mind, assume that the total corresponding cost for each *k* is $NRC2_{\,k}$.

5.1.4.3. Determining Trailers to Redirect

Based on the above calculations, every trailer *i* has been assigned a total cost $C_i = \sum_{k\,\in K} RC1_{\,k} + RC2_{\,k} + NRC1_{\,k} + NRC2_{\,k}$, based on the holding cost its items would incur if it is redirected and the holding cost for the items that are part of the trailers that are assumed to remain at LCS. This simulation is run for each trailer. Let Q_i be the quantity of trailer *i*. Let *R* represent the volume that LCS wishes to redirect to the FCs in order to reduce the backlog (this parameter is a user input). Let x_i be a binary decision variable that determines whether trailer i should be redirected. The optimization mode formulation is as follows:

$$Min \sum_i C_i \times x_i$$

$$Subject\ to$$

$$\sum_i Q_i \times x_i \geq R \qquad \forall\ i \in I$$

$$x_i = \begin{Bmatrix} 0, if\ trailer\ should\ not\ be\ redirected \\ 1, if\ trailer\ should\ be\ redirected \end{Bmatrix}$$

$$Q_i, C_i, R \geq 0$$

5.1.5. Data Requirements

The model requires large set of data points as inputs. All relevant data that contribute to evaluating holding cost and storage duration is required. In this study, the data was received from published metrics from different Amazon teams or estimated based on the tracked metrics. Table 5 gives a high-level overview of key parameters used in the model.

Table 5. Data List Required as Input for the Model

Parameter	Importance	Estimation Technique
Customer Demand	Demand profiles inform how much inventory will be consumed	90th percentile weekly national demand for different SKUs is received from Amazon forecasts
Target Inventory Plan	Target Inventory Plan informs how much inventory should be kept on-hand in the FCs	Target Inventory Plan is set by Amazon inventory planning team. It is given in weeks of cover (sum of weekly demand). Usually the weeks of cover for SKUs is 4
Operational costs	Costs affect the inventory cost and thus the trailers recommended to redirect	Costs obtained from databases: - LCS Processing Cost: \$/unit - Inventory storage costs at FC and LCS nodes: \$/ft^3/day
Purchase Orders	Incoming freight to the FCs affects replenishment instances from LCS nodes	Obtained from Logistics team databases representing future quantities of freight for different SKUs along with arrival date

5.2. Capacity Pipeline Allocation Tool

5.2.1. Tool Objective

The capacity pipeline allocation tool enables LCS to split weekly inbounding capacity at its sites between multiple sources of inventory (Asia-Pacific Consolidation Nodes, US Consolidation Nodes, and domestic retailing). The objective is to reduce the backlog at LCS sites by constraining capacity.

5.2.2. Tool Rationale

Amazon plans capacity at its sites for the next 13 weeks and updates these metrics on a weekly basis. The tool will first consider the 13-week forecast of incoming units from the different freight flows along with the inbounding capacity that is set by the LCS team. The forecast of incoming units identifies freight that is in transit and freight that has not yet been assigned to LCS sites. From the planned inbounding capacity at LCS sites, the tool will estimate the capacity portion that can be allocated given the forecast of incoming units. To do so, the tool will also take into account on the current backlog at LCS sites along with the backlog threshold. Inbounding capacity at LCS sites will first be allocated to Asia-Pacific consolidation nodes inventory, the remaining capacity will then be allocated to the US consolidation nodes inventory and then domestic retailing.

Asia-Pacific consolidation inventory is prioritized over that of the US since the local consolidation node lead time into the LCS network is 14-20 days vs. that of Asia-Pacific which is 35-50 days. This implies that for a given week, Asia-Pacific consolidation node freight will be allocated to LCS before that of the US consolidation node. In this approach, the capacity allocated to the Us consolidation node will take into consideration the incoming Asia-Pacific consolidation nodes freight as fixed.

Having a tool for distributing capacity across the three LCS freight flows will allow Amazon placement models to consider the capacity signals when they are executed at the consolidation nodes and domestic retailing. When the placement models run at Asia-Pacific consolidation nodes, it will take into consideration the planned inbounding capacity forecast for the week that this freight is expected to arrive in. The available inbounding capacity will thus be updated to account for that freight. When the placement model runs at the US consolidation nodes, it will consider the incoming Asia-Pacific consolidation nodes inventory to be fixed.

Given that capacity is updated on a weekly basis, if the updated inbounding capacity decreases from what was previously projected, then US consolidation nodes capacity allocations decrease. If the capacity increases, then US consolidation nodes allocations would increase accordingly. The remaining capacity will be allocated to domestic retailing. Domestic retailing is deprioritized compared to imports (Asia-Pacific and US consolidation nodes) given that the latter benefits more from LCS solution as it requires a long lead time to arrive into the US and comes in higher excess inventory than domestic retailing.

5.2.3. Data Requirements

The tool requires key data inputs in order to be able to estimate an available capacity that can be allocated to the different supply origin arcs. In this study, the data was received from published metrics from S&OP, the Logistics team, and LCS. Amazon. Table 6 gives a high-level overview of key data leveraged by the tool.

Table 6. Data List Required as Input for the Capacity Allocation Tool

Parameter	Importance	Estimation Technique
Freight End Backlog	Determines the backlog currently at an LCS node and is leveraged to determine available processing capacity with respect to backlog threshold	Backlogged units are provided by S&OP's portal which are then transformed into backlog days given the daily average of the forecasted inbounding capacity for the next 7 days
Inbounding Capacity Plan	Provides a forecast of the total inbounding capacity that LCS nodes have from a mechanical and labor perspective for the next 13 weeks – updated weekly	Estimated by the LCS team given their labor planning efforts and forecasted units to outbound LCS, and is submitted to S&OP's portal
Incoming Units Forecast	Provides a 13-week forecast of the arrival of incoming units. It distinguishes between units currently in transit and those expected to be allocated to LCS across the different freight origins	The forecast is provided by LCS' engineering team. The team is working on a machine learning model to improve forecasting

5.3. Asia-Pacific Consolidation Node Container Redirection Model

Given that Amazon cannot redirect Asia-Pacific consolidation node containers from the LCS yard, the goal of this model is to determine how to redirect this inventory that is headed towards

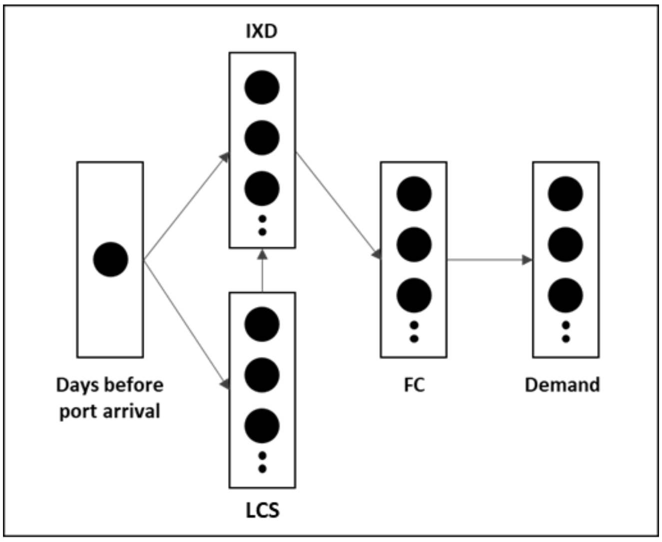
LCS, when inbounding capacity at LCS indicates the inability to receive the original number of containers due to backlog concerns.

5.3.1. Model Rationale

Given a given period and its corresponding capacity constraints, the incoming ocean containers will be redirected across LCS sites and the cross-dock network to head towards the FCs. The model draws out the entire network: incoming vessels have outgoing connections to the port then to the LCS sites and the IXD nodes, which have outgoing arcs to the FC nodes. The model is intended to determine which containers to redirect and their relocation destination.

All FC nodes have outgoing arcs to all demand nodes. The model will ensure that capacity among LCS nodes is taken into account and that inventory is placed closer to the FCs where it will be needed. This will be achieved by considering transportation costs as incentive for the model to store inventory as close as possible to the regions where inventory is needed. The redirection decision made by the model will be at a container level and will follow the freight flow in Figure 9. Containers will be redirected to other LCS nodes with available capacity or to IXD nodes. The model will be called when the LCS witnesses that it cannot receive some of the incoming freight based on updated capacity signals.

Figure 9. Redirection Flow Diagram



5.3.2. Model Formulae

This logic can be modeled as a time-index multi-commodity network flow optimization problem, which is the methodology that Amazon leverages in its allocation/ placement solutions. The model was based on the placement model developed by Amazon's research scientist and follows a similar logic [7]. Due to confidentiality reasons, only parts of the model will be portrayed. Table 7 represents an overview of some of the parameters and decision variables used.

Table 7. Model Parameters

Parameter	Description
$V = \{ V^{LCS} \cup V^{port} \cup V^{ixd} \cup V^{fc} \cup V^{dmd} \}$	*Set of nodes*
$A \subset V \times V$	*Arcs*
$T = \{ t_1, t_2, \dots, t_n, t_{n+1} \}$	*Consecutive set of non-overlapping time periods*
Δ_t	*Duration of time period $t \in T$ measured in days*
K	*Set of unique SKUs*
$d_{i\,k}^{t}$	*Demand of node $i \in V^{dmd}$ for SKU $k \in K$ in period $t \in T$, expressed in units*
C_a^{ship}	*Cost of shipping across arc $a \in A$*
$C_{i\,k}^{hold}$	*Cost of holding inventory at node $i \in V$, per unit of given SKU $k \in K$ per day*
C_t^{loss}	*Cost of lost demand of SKU $k \in K$, at node $i \in V^{dmd}$ in period $t \in T$*
$p_{k\,j}$	*The level of inventory in container $j \in J$ for $k \in K$*
$x_{a\,k}^{t}$	*Amount of SKU $k \in K$ shipped along arc $a \in A$. Shipment commences at period $t \in T$*
$I_{i\,k}^{t}$	*Inventory of node $i \in V^{fc} \cup V^{LCS}$ for SKU $k \in K$ at the beginning of period $t \in T$, expressed in units $i \in V^{dmd}$ in period $t \in T$*
$z_{i\,k}^{t}$	*Lost demand of SKU $k \in K$, at node $i \in V^{dmd}$ in period $t \in T$*
G_a^{j}	*Binary variable determining if container $j \in J$ is shipped along arc $a \in A'$ where $A' \subset V^{port} \times V^{LCS}$, $V^{port} \times V^{IXD}$*

The problem will follow a time-indexed multi-commodity network flow model where integer variables $x_{a\,k}^{t}$ represent the amount of SKU k transported across arc $a \in A$ in period $t \in T$, and where binary variable G_a^{j} represents if container j is shipped along the arcs connecting the US port with the cross docks or LCS sites.

The planning horizon H for this model is given in days. The ordered set of time periods $T = \{t_1, t_2, \dots, t_n, t_{n+1}\}$ is obtained by partitioning the planning horizon into n consecutive, non-overlapping time intervals which depend on the granularity at which the problem has to be solved.

In order to make the redirection decisions, the optimization model aims to minimize total supply chain cost, which includes holding, transportation, and loss of demand. It is formulated as follows:

$$\text{Min} \quad \sum_{t \in T} \sum_{a \in A} \sum_{k \in K} C_a^{ship} x_{a\,k}^t + \sum_{t \in \setminus T\{t_1\}} \sum_{i \in V} \sum_{k \in K} \Delta_{t-1} C_{i\,k}^{hold} I_{i\,k}^t + \sum_{t \in T} \sum_{i \in V^{dmd}} \sum_{k \in K} C_t^{loss} z_{i\,k}^t$$

This objective function is subject to a set of constraints. Below, we formulize two of these constraints. They model the fact that SKUs are in their corresponding containers and that SKUs cannot be individually routed between the port and the IXDs and LCS nodes

$$x_{a\,k}^1 = \sum_{j \in J} p_{k\,j} G_a^j \qquad \forall\, a \in A', k \in K$$

$$\sum_{a \in A'} G_a^j = 1 \qquad \forall\, j \in J$$

The remaining constraints cover the below:

- Implement flow preservation for IXD nodes
- Control inventory levels for a given SKU $k \in K$. The inventory level at period t+1 equals the inventory level at period t minus the quantity that has been shipped out plus the quantity that has arrived
- Ensure that customer demand is fulfilled
- Implement processing capacity constraints
- Define the variables and their domains

5.3.3. Data Requirements

The model requires large set of data points as inputs. All relevant data that contribute to evaluating supply chain costs is required. In this study, the data sources were identified through published metrics from different Amazon teams or estimated based on the tracked metrics. Table 8 gives a high-level overview of key parameters used in the model.

Table 8. Data List Required as Input for Asia-Pacific Consolidation Node Container Redirection

Parameter	Importance	Estimation Technique
Customer Demand	Demand profiles inform how much inventory will be consumed	90th percentile weekly demand across Amazon's demand nodes is received from their forecasts. Demand for a time period is then aggregated
Inventory Level	Informs on-hand inventory of different SKUs across all of Amazon's FCs and LCS sites	On-hand inventory is received by Amazon's logistics team and facilities' operational teams
Operational costs	Costs affect the redirection of freight	Costs obtained from databases: - LCS Processing Cost: \$/unit - Transportation Cost: \$/unit/arc - Demand Loss Cost: \$/unit - Inventory storage costs at FC and LCS nodes: $\$/ft^3/day$
Purchase Orders	Incoming freight to the FCs affects inventory at facilities and thus customer demand fulfillment	Obtained from Logistics team databases representing future quantities of freight for different SKUs along with arrival date

6. Results and Discussion

This chapter presents the results from the study. We highlight the cost savings and benefits that can arise from adopting our cost-based LCS yard trailer redirection model. We then discuss logistical implications and dependencies for both the Asia-Pacific consolidation node container redirection model, and the capacity pipeline allocation tool. We also identify the implementation plan required for this effort along with the teams involved to execute. Finally, we comment on the limitations and potential improvements.

6.1. LCS Yard Trailer Redirection Model

6.1.1. Testing Setup

To test the redirection model and understand savings potential, we ran redirection simulations by taking backlog snapshots across different weeks. We identified what trailers are part of the backlog at an LCS site along with the data regarding each SKU included in these trailers. The main data components include current inventory at LCS and the fulfillment network, SKU volume, SKU quantity and distribution across trailers, demand forecasts, future POs to be received in the FCs. Below are the results of one representative simulation.

6.1.2. Simulation Details

6.1.2.1. Cost Saving Potential

The simulation included eleven containers and five backlogged trailers at an LCS site (backlog above threshold). The trailers had a total quantity of 74,000 units and 111 SKUs. The quantity that needed to be redirected was set at 30,000 units. The model generated a combination that met the redirected target quantity (>= 30,000 units) with the lowest cost on the network. In Figure 10, we compare the optimal combination that the model identified with a combination that the current non-cost-based redirection process could have recommended. Our redirection model presents a large cost-saving opportunity for Amazon. Following the combination of trailers that the model recommended redirecting allows Amazon to benefit from significant savings in holding costs. The represented total cost is the expected holding cost of all the SKUs, included in the five analyzed trailers, from the instance they are received into the LCS node or the FCs up until they are consumed by customer demand. We notice that the total cost of the optimal solution is 62% less than that of the current method case.

Figure 10 demonstrates that the difference in total holding cost between the 2 options is driven by the split in FC to LCS holding costs. The difference between FC holding cost for the SKUs in the current methodology case and the optimal case is in the hundreds of thousands of dollars. Moreover, the FC holding cost in the current method solution represents 84% of the total cost vs. 29% in the optimal solution.

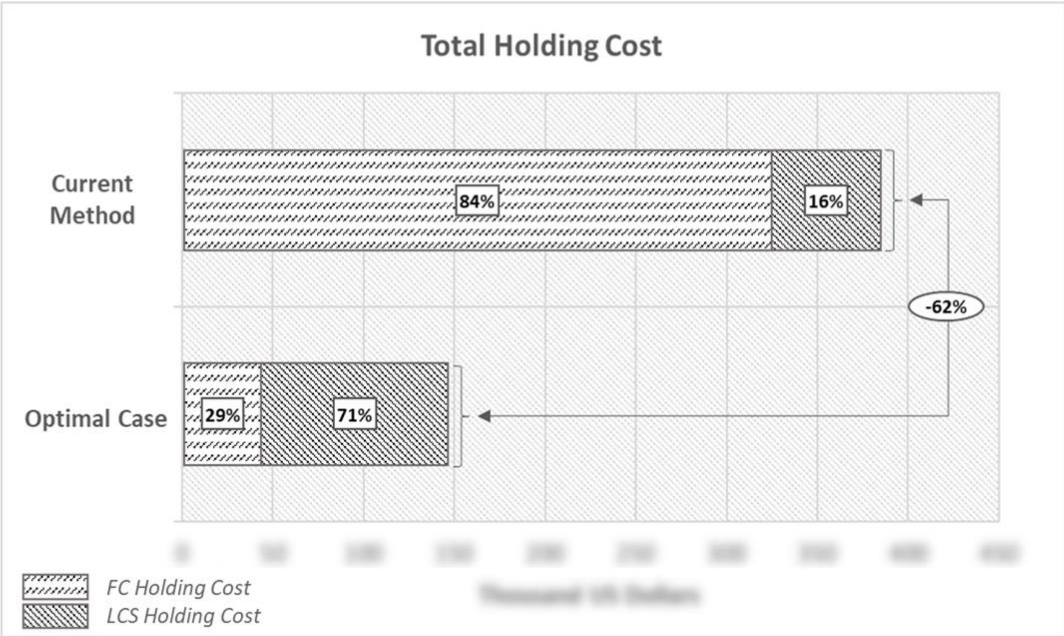


Figure 10. Total Holding Cost Comparison between the Current Method and the Optimal Case

6.1.2.2. Network Coverage

The metrics in Table 9 demonstrate that our optimization model is recommending the redirection of SKUs that have lower coverage in the network (4.23 vs 26.38 weeks). This is in line with LCS' value proposition. LCS is to be leveraged to hold excess inventory in a low-cost storage node. This implies that SKUs will spend a bigger portion of their dwell time at LCS vs. the fulfillment network, which achieves cost savings in the network. The SKUs that remained at LCS in the current method had lower coverage in the FCs compared to the ones that remained in the optimal case (4.1 weeks vs. 7.4 weeks). In fact, in the current method, one of the trailers that remained at LCS had SKUs with an average weeks of cover (WoC) of 2.3 in the FCs.

Table 9. Redirected SKUs Metrics for the Current Method and Optimal Case

	Number of Trailers	Avg. WoC in the FCs	Avg. WoC in LCS	Number of Units	Volume per SKU	Number of SKUs
Current Method	2	26.38	4.31	33,505	0.16	20
Optimal Case	2	4.23	4.83	36,861	0.09	11

The current methodology at LCS might put more weight on moving quantity that is closer to the suggested target quantity set by the LCS team, which in this case is 30,000. Our model aims instead at minimizing the total holding cost while considering the target as a constraint to be satisfied.

This method does not only support achieving cost savings at LCS and within the network but also supports lowering the out of stock risk. In our analysis, 19% of the SKUs in the backlogged trailers had coverage in the FCs lower than the target inventory plan of four weeks. This is driven by the stochastic nature of demand and the uncertainty in lead times (e.g. operational challenges at the receiving US port). If these SKUs were not redirected to fulfill customer demand through the FCs, they would not have spent the breakeven duration at LCS and would have driven losses at the site. With this in mind, the model can also be used to analyze the out of stock risk of inventory in LCS' yard.

6.1.3. Implementation Implications

Implementing the model would not disrupt current work procedures across the main stakeholders: LCS, S&OP, and Regional Logistics Team (RLT). The addition to LCS' part would be to run the model by inserting the target number of units to be redirected and inputting the aforementioned data attributes. The LCS team would need to exclude trailers that are currently being worked/received at the site. The model would output the trailer combination that minimizes total holding costs. When launching a request to redirect trailers, the LCS team would include the identified trailer IDs. Upon receiving the request, RLT would work to move the suggested trailers. There is a possibility that RLT faces limitations in moving the identified trailers. The causes might be issues related to 3rd party carriers, which hinders the ability to move them. In this case, RLT would notify the LCS team who can suggest other trailers based on the value per volume calculated for each trailer. There is no amendment to the role that S&OP would play in this process.

6.1.4. Model Execution

The model has been included in LCS' workflow and has been executed by the team. It is a standalone script with detailed simulation steps. To execute the model, LCS' business intelligence engineers will obtain the yard trailers data though queries, feed the data into the script, and then run it (estimated total runtime less than 1 hour). The model will output the optimal trailers to redirect away from the LCS site's yard. To execute trailer redirections, the selected trailers will be conveyed to the RTL through the current redirection queries.

6.2. Capacity Pipeline Allocation Tool

6.2.1. Capacity Dashboard

Having a model for distributing capacity across the three LCS freight flows will allow the Amazon allocation model to consider capacity signals when it is executed at consolidation nodes. Implementing this model when forecast accuracy improves will be the main driver for balancing the inbounding backlog at LCS nodes. In the methodology section, we highlighted that the capacity allocation priority will be the following: Asia-Pacific consolidation node, US consolidation node, and then domestic retailing.

We have developed a dashboard that follows the aforementioned logic. The model considers information about the current backlog, incoming units, and capacity at LCS then recommends to the LCS operational team how to distribute capacity to maintain a backlog of 10-days. The model's flexibility allows the user to adjust capacity allocation and understand how it affects backlog figures. The data needed is provided by S&OP's portal and LCS' forecasts. LCS will hold ownership of the model and share its capacity outputs with S&OP every week for the next 13 weeks.

6.2.2. Development and Implementation Dependencies

The implementation and usage of this logic and tool to start planning capacity at an arc level have a set of dependencies and considerations to be taken into account:

- The machine learning (ML) prediction model for trailer/container arrivals is needed as it provides needed input for the tool. The ML model is expected to improve forecast accuracy of incoming and outgoing LCS units which in hand will enable an accurate inbounding capacity plan at LCS nodes. An accurate forecast is needed to be able to leverage this tool and split capacity among freight arcs.

- Amazon's current placement model, that allocates freight to LCS nodes, ought to account for lead time uncertainty. Given that the placement model will consider weekly capacity at LCS, it needs to ensure that inventory will arrive to LCS nodes within the same time period it estimated.
- From a technology perspective, once the above model dependencies are solved for, LCS should inform placement teams that capacity signals should be considered while undergoing the placement decision. Placement models have the capability to account for capacity, however, the team should determine if there are system changes required to consider these signals.
- From an operational perspective, LCS needs to inform S&OP and Logistics teams that placement models will run in a constrained environment. These constraints will be shared weekly by LCS with the S&OP team who will transmit them to Logistics teams to execute on.
- LCS will lead the effort to implement the current excel dashboard onto Amazon systems to automate data input and output, and visualization. Once the dashboard is active and capacity signals are generated, they are to be transferred weekly to S&OP through the same procedure that the LCS team currently has in place.

6.3. Asia-Pacific Consolidation Node Container Redirection Model

6.3.1. Model Use

The redirection decision made by the model will be at an ocean container level. Containers will be redirected to other LCS nodes with available capacity or to cross dock nodes and then FCs. The model will be called when the LCS team realizes that it cannot receive some of the incoming freight based on updated capacity signals.

For a given week, the model will consider an inbounding capacity signal along the Asia-Pacific consolidation node – LCS node arc. Containers that are on vessels 3 or more days away from the port with an estimated LCS arrival time during the week in question will be taken into consideration and their content will be fed into the model. The model will then determine which containers should be redirected and where they should be redirected to. A redirection decision needs to be taken at least 3 days before a vessel is received at the port to give time for the Port Logistics Team (PLT) to update its systems with the final destination of these containers.

The model is expected to be leveraged when incoming Asia-Pacific consolidation node freight is significantly high and trailer LCS yard redirections does not reduce the backlog down to its threshold of 10 days. With this logic in mind, we conducted an analysis to determine the potential frequency of redirecting Asia-Pacific consolidation node containers. We noticed that over the past 30 weeks, the model would have to be run once, where the backlog rose to 22 days. The expected number of containers to be redirected was 19. Undergoing this redirection would have pushed the backlog closer to 10 days within a week. Not undergoing this decision would have entailed an above threshold backlog for three weeks.

In addition to capacity management, this model gives Amazon the resilience and cost-based logic to manage any emergencies that impose the need to redirect containers away from LCS and into the fulfillment network in order to reduce any potential of out of stock risk.

6.3.2. Model Execution

From an execution perspective, the main stakeholders involved are LCS, PLT, and the Placement Team (PT). PT would support LCS in running the model. LCS would be responsible in understanding what freight is heading to its sites and send corresponding capacity signals to both PT and PLT. In terms of sending a capacity signal, the LCS team needs to take into account the time for vessels to be received at the port and then at LCS nodes which is between 12 and 15 days during non-peak conditions.

Given the aforementioned logistical constraints, the LCS team would need to send redirection requests that accommodate the transfer period, which means 2 weeks in advance. Once PT runs the model and determines which containers should be redirected and their end destination, this information would be transferred to PLT for execution.

6.3.3. Development and Implementation Dependencies

Given the current forecasting error within LCS and that implementing this model is labor-intensive, this study only covered the mathematical formulation of this redirection model and the logistical and operational feasibility of undergoing a redirection for ocean containers. Going forward, PT would leverage this model to implement a prototype, run corresponding simulations, and include this redirection in its workflows.

In addition, the machine learning model that the LCS team is working on for better forecast accuracy is needed given that capacity is needed as a signal to be taken into account by this model. An accurate forecast will be essential to determine ocean container redirection needs and to provide the needed input.

7. Conclusion

The application of freight redirection and capacity management in this study establish the benefits they can drive on LCS cost savings and inventory selection that would optimize LCS entitlement and minimize out of stock risk in the FC network. This chapter provides a summary of the principal findings and highlights opportunities for future work and improvement.

7.1. Final Recommendation

The LCS yard trailer redirection model provides a cost-based logic for selecting trailers to redirect towards the fulfillment network. The model supports Amazon in achieving higher cost savings than the current method and supports mitigating out of stock risk by directing trailers that contain SKUs with lower excess coverage. Given that the holding cost at LCS sites depends on dwell time, the model also contributes to increasing fullness at LCS. In our example, we saw that redirecting trailers based on our model recommendations achieved 62% lower holding cost across the network.

Our work also determined a capacity split logic between LCS' freight flows. This split can be leveraged by the LCS team as its forecast improves. Leveraging this model would enable PT to account for capacity signals when it is determining inventory allocations and would be the main driver for balancing backlog. Also, setting capacity signals on placement models ensures that the optimal inventory is allocated to LCS and achieves the minimum supply chain costs.

Finally, this study also setup a redirection logic for ocean containers originating from Asia-Pacific consolidation nodes and heading towards LCS. The formulization minimizes total transportation, holding, and shortage cost and determines where containers need to be rerouted based on their final customer demand destination. The model does not only contribute to capacity management at LCS but can also be leveraged to redirect containers, closer to the US, and mitigate against out of stock risk when updated demand forecasts experience high variation.

7.2. Improvements and Future Work

The developed models and tools set the foundation for capacity management at LCS and have some opportunities to be improvement. Looking forward, below are some updates that can be incorporated:

LCS Yard Trailer Redirection Model: (1) The transportation lead time between LCS and the fulfillment network can be taken into account given that the model currently assumes that inventory from LCS is instantly transferred into the network. (2) In case the current replenishment logic from LCS changes, the model would need to be evaluated to account for how inventory is selected across LCS sites. (3) A feedback loop with PT that shares the proportion of trailer inventory in LCS' yard that has low coverage in the network.

Capacity Pipeline Allocation Tool: The model logic will need to be evaluated, if LCS starts influencing buying decisions, to determine if there is a need to account for bulk discounts. For example, when this inventory is set to head towards the network from a supply source (e.g. consolidation nodes), the arc capacity into LCS should be set to accommodate this inventory. Optimization logic can be explored to determine how to split capacity among sources. To undergo this effort, the LCS team will need to collaborate with the PT research team to develop a solution.

Asia-Pacific Consolidation Node Container Redirection Model: The redirection model was formulized based on the mathematical logic of the current PT allocation model and any improvement that the PT model makes can be benefited from for this model. The key improvements that can be brought on are: (1) Accounting for transportation lead time uncertainty. (2) Enforcing a minimum inventory threshold for the FC nodes as an additional constraint to take into account the target inventory plan. (3) Additional research to determine an optimal partitioning of the time horizon H into a set of n non-overlapping periods.

8. References

[1] M. Hall, "Amazon.com American Company", Britannica, 2018. [Online]. Available: https://www.britannica.com/topic/Amazoncom. [Accessed: 15-Jan-2021].

[2] *A. Jassy, Amazon Web Services re:Invent Conference.* 2020

[3] D. Davis, "Jeff Bezos sets his departure as CEO as Amazon reports a record year", Digital Commerce 360, 2021. [Online]. Available: https://www.digitalcommerce360.com/article/amazon-sales/. [Accessed: 05-Feb-2021].

[4] "Frequency with which Amazon shoppers in the United States purchase from Amazon as of February 2019, by membership status", Statista, 2019. [Online]. Available: https://www.statista.com/statistics/646987/us-amazon-shopping-frequency/. [Accessed: 15-Jan-2021].

[5] S. Devji, personal communication, July - August, 2020.

[6] S. Das, "Multi Echelon Supply Chain Design for Amazon Private Brands", Massachusetts Institute of Technology, 2020.

[7] J. Kinable, "Inbound Planning", Amazon Research, 2020.

[8] N. Sbai, A. Berrado, "A literature review on multi-echelon inventory management: the case of pharmaceutical supply chain", MATEC Web of Conferences, 2018, doi: https://doi.org/10.1051/matecconf/201820000013.

[9] A.J. Clark, "An informal survey of multi-echelon inventory theory", NRL, 1972.

[10] P. Li, "Optimization of (R,Q) policies for multi-echelon inventory systems with guaranteed service", LOSI - UTT, 2013.

[11] E. Ramos, S. Dien, A. Gonzales, M. Chavez, B. Hazen, “Supply chain cost research: a bibliometric mapping perspective”, Benchmarking: An International Journal, ISSN: 1463-5771, 2020.

[12] H. Voordijk, “Physical distribution costs in construction supply chains: a systems approach”, International Journal of Logistics Systems and Management, Vol. 7 No. 4, pp. 456-471, 2010.

[13] E.A. Silver, R. Peterson, D.F. Pyke, “Inventory Management and Production Planning and Scheduling”, ISBN: 978-0471119470, Chapter 1.

[14] D.J, Bowersox, “Physical distribution development, current status, and potential”, Journal of Marketing, Vol. 33, No. 1, pp.63–70, 1969.

[15] P. Tsiakis, L.G. Papageorgiou, “Optimal production allocation and distribution supply chain networks”, International Journal of Production Economics, Vol. 111, No. 2, pp.468–483, 2008.

[16] R. Rajkanth, G. Srinivasan, M. Gopalakrishnan, “Material flow optimisation in a multi-echelon and multi-product supply chain”, Int. J. Logistics Systems and Management, Vol. 26, No. 1, 2017.

[17] A. Nikolopoulou, M.G. Ierapetritou, “Hybrid simulation based optimization approach for supply chain management”, Computers & Chemical Engineering, Vol. 47, pp.183–193, 2012.

[18] A. Federgruen, P. Zipkin, “A combined vehicle routing and inventory allocation problem”, Operations Research, Vol. 32, No. 5, pp.1019–1037, 1984.