1. Introduction

With increasing concern over the environment, many companies worldwide are improving their supply chain sustainability under pressure from the government or from their own shareholders [1]. In logistics systems, transportation is considered to be the largest source of environmental hazards [2]. The smart use of logistics systems has much to offer in greening the supply chain by means of various efficient distribution and transportation strategies. Shipment consolidation is one of a main initiative toward environmental sustainability among the logistics service providers [3].

Shipment consolidation is a transportation strategy that can reduce environmental hazards. “Shipment consolidation is a logistics strategy that combines two or more orders or shipments so that a larger quantity can be dispatched on the same vehicle” [4]. The main motivation for shipment consolidation is decreased unit dispatch cost due to economies of scale in transportation. Shipment consolidation can also reduce adverse environmental impacts of business processes, as it can reduce harmful emissions affecting air quality, such as CO₂, emitted from delivery vehicle exhausts [5]. Moreover, shipment consolidation could accelerate corporate social responsibility performance [6].

Shipment consolidation strategy should be chosen in state of supply chain coordination [7]. To keep up with this trend, integrated policies with transportation decision and inventory control have been studied. Shipment consolidation does not always decrease a firm’s cost. Delivery time and inventory holding time increase while several small orders are consolidated into a larger shipment.
Many recent e-commerce providers such as 11 Street, the largest e-commerce provider in South Korea, allow order cancellation if order is not yet shipped for the sake of better customer service. Since better customer service results in higher customer satisfaction, and since customer satisfaction leads to customer loyalty and profitability [8], allowing order cancellation, whether from a flood of information about goods and prices shared among customers or from simple changes of customers’ opinion, is beneficial for companies in the long term, even at the cost of increased order uncertainty. Since uncertainty is caused by many factors including inventory management, vendor selection, transport planning, production planning, distribution planning, and procurement planning, it thus is important to understand, manage, and reduce uncertainty in supply chain to improve performance [9]. Hence, such trade-offs must be considered when making decisions about shipment consolidation [10].

Previous studies on shipment consolidation do not consider order cancellation due to longer delivery time. Thus, it is assumed that all customers are willing to wait during shipment consolidation, and all orders would be met. However, impatient customers may cancel their orders when delivery time increases [11], and delivery time uncertainty decreases customer satisfaction [12]. There is a trade-off between minimizing the cost of having a customer to wait and the cost of providing service [13]. Thus, the order cancellation needs to be considered in shipment consolidation policy.

In this paper, we consider order cancellation scheme and develop both quantity-based shipment consolidation policy and time-based shipment consolidation policy to minimize the total cost incurred and to reduce environmental hazards such as CO$_2$ emission caused during transportation. In addition, we analyze the results of both policies and assess the effects of dispatch cost and order cancellation cost on total cost for each increment to provide basis for managerial decision making under each of both policies.

This paper is organized as follows: Section 2 provides an overview of shipment consolidation and inventory policy under order cancellation literature. In Section 3, we developed quantity-based and time-based policies by incorporating order cancellation to minimize the total cost with consideration of environmental cost. Mathematical models are developed and efficient algorithms are provided to obtain the optimal parameters for the proposed policies. To gain further insight, in Section 4, we conducted computational experiments and analyze the sensitivity of the optimal decisions with respect to the model parameters. In addition, to compare the performances of the quantity-based policy with the time-based policy, extensive numerical experiments are conducted, and the total cost is compared. Our conclusions are provided in Section 5.

2. Literature Review

Over the last two decades, environmental or “green” factors have become increasingly incorporated into analyses in supply chain management [14]. In addition, recent global markets are greatly affected by swift changes in technology and customer demand, which cause shorter product life cycle that result in increased uncertainty of supply chain, requiring adequate control of risks to minimize uncertainty [15]. Such situation puts more emphasis on “green” factor than ever before, resulting many logistics companies to realize the value of greening their operations. For example, Wal-Mart, the world’s biggest retailer, undertook a green supply chain management (GSCM) project and asked its 60,000 suppliers worldwide to reduce their use of packaging by 5%, which amounts to removing 667,000 m$^2$ of CO$_2$ from the air and 213,000 trucks from the road, resulting in a huge savings of $3.4 billion [16]. In the scope of environmental sustainability, much focus has been given to greenhouse gas (GHG) emission, particularly carbon dioxide (CO$_2$) emissions [17]. Radio Frequency Identification (RFID) technology is a good example that can be applied to such case for more information regarding sustainability since introducing RFID benefits supply chain for all of its echelons by effectively controlling uncertainties and complexities, which in this case is greenhouse gas emission [18]. However, adopting RFID technology does not necessarily result in improvement since feasibility study may turn out to be negative in some cases [19]. Thus, different approach is needed to acquire sustainability in transportation.
Motivated by the apparent importance of reducing environmental damage caused by transportation, many studies explored the impact of shipment consolidation. McKinnon [20] surveyed opportunities to improve the use of road vehicles and suggested that the impact of freight transportation on the environment can be reduced by “increasing the share of freight moved in less environmentally damaging ways, by increasing vehicle load factors”. He reported that increasing truck load, such as shipment consolidation, could yield greater environmental benefits than a modal shift. He also showed that increasing the maximum truck weight can further yield economic and environmental benefits. Merrick and Bookbinder [21] reported the reduction of CO$_2$ emissions caused by the shipment consolidation. The finding showed that, for any given speed, the quantity of CO$_2$ emissions is a concave increasing function of consolidated load’s weight. Furthermore, Pan et al. [22] show that the consolidation of truck freight among supply chains (namely pooling supply chains) can be a solution to significantly reduce CO$_2$ emissions and transportation cost.

In industry, two different types of shipment consolidation policies are commonly used: quantity-based and time-based [23]. In addition, in general, consolidation policies are differentiated into these two policies by many existing literature, and each of the two policies has distinctive characteristics [24]. The quantity-based policy ships accumulated loads when a predetermined economical dispatch quantity, is accumulated, whereas the time-based policy ships accumulated loads (all outstanding orders) every period. Under the time-based policy, each order is dispatched on pre-specified shipment release dates, even if the dispatch quantity does not necessarily satisfy transportation scale economies. On the other hand, under the quantity-based policy, the dispatch quantity assures transportation scale economies, but a specific dispatch time cannot be guaranteed. An alternative to these two policies is a hybrid policy aimed at balancing the trade-offs between the timely delivery of the time-based policy and the transportation cost savings associated with the quantity-based policy. Under the hybrid policy, a dispatch decision is made either when the size of a consolidated load exceeds pre-specified dispatch quantity, or when the time since the last dispatch exceeds pre-specified dispatch time.

Cetinkaya and Bookbinder [25], Chen et al. [26], Cetinkaya and Lee [27], Moon et al. [28], Ching and Tai [29], and Cetinkaya et al. [23] have developed the optimization models for shipment consolidation. For the demand arrival following a Poisson process, Cetinkaya and Bookbinder [25], and Chen et al. [26] have developed the optimal quantity based policy. Cetinkaya and Lee [27] present an optimization model for coordinating inventory and transportation decisions at an outbound distribution warehouse that serves a group of customers located in a given market. Moon et al. [28] developed joint replenishment and consolidated freight delivery policies for a TPW that handles multiple items. They extended the results of Cetinkaya and Lee [27] to consider the joint replenishment of multiple items and introduce two time-based policies (stationary policy and non-stationary policy) for the warehouse. The optimal hybrid dispatch policy with stochastic demand is studied by Ching and Tai [29] and Cetinkaya et al. [23]. They analyzed the advantages and the disadvantages of the quantity-based policy and the time-based policy, and they proposed hybrid policies since combination of the two most popular policies (quantity-based and time-based) in logistics literature may yield another option for consolidation policy. However, it is found that hybrid policies are not superior to quantity-based policies when resulting cost is the variable on comparison [23]. Günther and Seiler [30] investigated an operational transportation planning problem based on a real industry case on shipment consolidation.

Recently, ÜLKÜ and Bookbinder [31] investigated the effects of different pricing schemes for a Third Party Logistics (3PL) provider who tenders a consolidated load to a carrier. They present an optimization model for integrating pricing and transportation decisions, and they derive the optimal quotations that should be made for price and delivery time with the objective of maximizing the profit. Mutlu et al. [32] investigated pure consolidation problem that can be confronted by a 3PL company and found a special case of time-and-quantity-based policy, which is quantity-based policy, is the optimal policy in terms of cost. Hong and Lee [33] considered a single-item inventory
system where shipments are consolidated to reduce the transportation cost using a time-based consolidation policy. They showed that additional profit can be obtained using shipment consolidation policy. Centinkaya et al. [34] examined the trade-off between average order delay and expected delivery frequency, as measured by the expected shipment consolidation cycle length. They proposed service-based performance criteria such as maximum waiting time and average order delay and compared the performance of shipment consolidation policies. Cetinkaya and Lee [35] also extensively computed time-based consolidation policy by incorporating replenishment quantity, which is presented as a basis for future analytical work. However, these previous studies on shipment consolidation have assumed that all customers are willing to wait during shipment consolidation and that all orders would be met. However, in real life, customers may cancel their orders when delivery time increases. Thus, the order cancellation should be considered to determine the optimal shipment consolidation policy. Unlike the existing research on shipment consolidation, we develop the optimization models to determine the optimal shipment consolidation policies with consideration of order cancellation.

Generally, there is a dearth of literature considering the impact of demand cancellation on the optimal ordering policy of the inventory model. Cheung and Zhang [36] study the impact of cancellation of customer orders via assuming an (s, S) policy and Poisson demands. They develop a Bernoulli type cancellation behavior in which a reservation will be canceled with probability \( p \). In addition, the timing to cancellation is considered. In particular, they show that a stochastically larger elapsed time from reservation to cancellation increases the system’s penalty and holding costs. Yuan and Cheung [37] consider a periodic review inventory model in which all demands are reserved with one-period lead time, but orders can be canceled during the reservation period. They formulated a dynamic programming model and show that the order-up-to policy is optimal. You [38] investigates a joint ordering and pricing problem for a single period model in which the system sells perishable products over a short sales season. He proves that the optimal ordering policy has an order-up-to structure. You and Hsieh [39] develop a continuous time model to determine the production level and pricing decision by considering constant rate of demand cancellation. They formulate a system of differential equations for inventory level so that holding cost and penalty cost can be calculated. Recently, Yeo and Yuan [40] consider a periodic review model where the firm manages its inventory under supply uncertainty and demand cancellation, and they explore the structure of the optimal inventory policy in the presence of demand cancellation and supply uncertainty in the multiple period framework. However, they did not address the impact of cancellation on the optimal cost of managing the system. These models did not consider outbound shipment scheduling. Our study differs from these models in that we consider shipment consolidation and obtain the optimal shipment policy.

In this paper, we consider a single-item inventory system where shipments are consolidated to reduce the transportation cost, and we develop optimization models where the shipment and ordering policies are optimized all together. In addition, we conduct sensitivity analysis for a couple of variables to find optimal zone where total cost is not heavily impacted by variability.

3. Mathematical Models

3.1. Quantity-Based Policy to Consideration of Order Cancellation

In this study, we consider an inventory system operated by a shipment consolidation policy. The time between two successive dispatch decisions is called a dispatch cycle, and all orders arriving during a dispatch cycle are combined to form a large outbound load. In order to employ a shipment consolidation, we assume that the customer order fulfillment may be postponed during a dispatch cycle. However, the postponement of order may result in customer waiting. This postponement has negative impact on customer demand. Thus, the customer may cancel the order.

In this paper, we assume that each customer requests one unit of the product, and the demand arrives according to a Poisson process with mean \( \lambda \). We also assume that the shipment cost is irrespective of the customer location (transportation distance). In this paper, we assume that the
delivery lead time is negligible (i.e., customers are located in a relatively close proximity). Under this assumption, the shipment cost consists of a fixed cost of hiring trucks (or other transportation means) and a variable cost that is determined by volume, not by distance.

The following notations are employed in this study:

- $F_R$: fixed cost of replenish inventory
- $C_R$: unit replenish cost
- $F_D$: fixed cost of dispatching shipment to customer
- $C_D$: unit dispatch cost
- $h$: holding cost per unit per unit time
- $\lambda$: Poisson demand rate
- $O_c$: order cancellation cost per unit
- $\gamma$: order cancellation rate per unit time ($0 < \gamma < 1$)
- $C_E$: unit environmental cost during dispatch

In this section, we present a mathematical model for optimal quantity-based dispatch policy. Figure 1 shows the inventory level under the quantity-based dispatch policy. Note that accumulated loads are shipped after economic freight quantity is accumulated in a quantity-based policy.

![Inventory level under the quantity-based policy.](image)

The following additional notations are employed in the quantity-based dispatch policy:

- $q$: dispatch quantity (integer, decision variable, $1 \leq q$)
- $n$: number of dispatch cycles within an inventory replenishment cycle (integer, decision variable, $1 \leq n$)
- $Q$: replenishment quantity ($Q = nq$)

During a dispatch cycle, customers may cancel their orders (we assume that order cancellation occurs in proportion to remaining order quantity for every unit time). The remaining order quantity is $q(1 - \gamma)$ at the first unit time after the beginning of a dispatch cycle, $q(1 - \gamma)(1 - \gamma)$ at the second unit time, and $q(1 - \gamma)^{n/\lambda}$ at the end of dispatch cycle (note that $q/\lambda$ is dispatch cycle time). Thus, when the sum of received order’s quantity reaches the dispatch quantity ($q$), units that a vendor should ship to customer are $q(1 - \gamma)^{n/\lambda}$. When the inventory level reaches zero, stock is replenished.

As a result, the problem is to compute the optimal number of dispatch cycles within a replenishment cycle, $n$, and the optimal dispatch quantity, $q$, in order to minimize the total cost.
As shown in Figure 1, the process under consideration is a renewal process. Thus, using the Renewal Reward Theorem, the long-run average cost, $TC(n, q)$, is determined by dividing $E(\text{Replenishment Cycle Cost})$ by $E(\text{Replenishment Cycle Length})$.

Since demand process is a Poisson process, the expectation of the dispatch cycle length is $q/\lambda$. Since the number of dispatch cycles within an inventory replenishment cycle is $n$, the expectation of the replenishment cycle length is:

$$E(\text{Replenishment Cycle Length}) = nq/\lambda \quad (1)$$

We now compute four different cost elements (replenishment, dispatch, inventory holding, and order cancellation) during a replenishment cycle:

- **Order Cancellation Cost**: Since the company will not dispatch its products until $q$ units of demand accumulate, the order cancellation rate is $[1 - (1 - \gamma)^{q/\lambda}]$. Thus, the order cancellation cost per dispatch cycle is $O_c \cdot q[1 - (1 - \gamma)^{q/\lambda}]$. Since there are $n$ dispatch cycles in a replenishment cycle, the customer waiting cost per replenishment cycle is:

$$\text{OrderCancellationCost} = n \cdot O_c \cdot q \left[ 1 - (1 - \gamma)^{q/\lambda} \right] \quad (2)$$

- **Replenishment Cost**: Since the replenishment quantity, $Q$, is equal to $nq(1 - \gamma)^{q/\lambda}$ (see Figure 1), the replenishment cost is:

$$\text{ReplenishmentCost} = F_R + C_R \cdot nq(1 - \gamma)^{q/\lambda} \quad (3)$$

- **Dispatch Cost**: Since the dispatch quantity is $q(1 - \gamma)^{q/\lambda}$, the dispatch cost in a dispatch cycle is $F_D + C_D \cdot q(1 - \gamma)^{q/\lambda}$. There are $n$ dispatch cycles during a replenishment cycle, and, thus, the dispatch cost during the cycle is:

$$\text{DispatchCost} = n \left[ F_D + C_D \cdot q(1 - \gamma)^{q/\lambda} \right] \quad (4)$$

- **Inventory Holding Cost**: At the beginning of a replenishment cycle, the inventory level is $(n - 1)q(1 - \gamma)^{q/\lambda}$. This implies that the inventory level is kept at $(n - 1)q(1 - \gamma)^{q/\lambda}$ throughout the first dispatch cycle, and incurs expected holding cost of $h \cdot (n - 1)q(1 - \gamma)^{q/\lambda} \cdot q/\lambda$. For the $i$th dispatch cycle, the expected holding cost is $h \cdot (n - i)q(1 - \gamma)^{q/\lambda} \cdot q/\lambda$. Hence, the total expected inventory holding cost is given by:

$$\text{InventoryHoldingCost} = \frac{\sum_{i=1}^{n-1} h \cdot (n - i)q(1 - \gamma)^{q/\lambda} \cdot q/\lambda}{2\lambda} \quad (5)$$

Using the above results, the expected cost during a replenishment cycle is computed by:

$$E(\text{ReplenishmentCycleCost}) = \left[ F_R + C_R \cdot nq(1 - \gamma)^{q/\lambda} \right] + n \left[ F_D + C_D \cdot q(1 - \gamma)^{q/\lambda} \right] + \frac{h \cdot n(n-1)q^2(1-\gamma)^{q/\lambda}}{2\lambda} + n \cdot O_c \cdot q \left[ 1 - (1 - \gamma)^{q/\lambda} \right] \quad (6)$$
Conversely, the expression for the long-run average profit, $TP(n, q)$, is given by:

$$
TC(n, q) = \frac{\left(F_R + C_B \cdot nq(1 - \gamma)^{q/\lambda}\right) + n\left(F_D + C_D \cdot q(1 - \gamma)^{q/\lambda}\right)}{2\lambda} + n \cdot O_c \cdot q \cdot \gamma^{q/\lambda} = \frac{F_R + C_B \cdot (1 - \gamma)^{q/\lambda} + F_D \cdot q(1 - \gamma)^{q/\lambda} + C_D \cdot (1 - \gamma)^{q/\lambda} + C_E \cdot q\cdot \gamma^{q/\lambda}}{2} + O_c \cdot \lambda \cdot \left[1 - (1 - \gamma)^{q/\lambda}\right]
$$

(7)

We assume that the environmental cost mainly depends on carbon emission during transportation, and we assume carbon emission is linear to loads transported. Since the dispatch quantity is $q(1 - \gamma)^{q/\lambda}$ and the dispatch cycle is $q/\lambda$, the long-run average environmental cost is $C_E \cdot q(1 - \gamma)^{q/\lambda} / (q/\lambda)$. Hence, the long-run average cost with consideration of environmental cost, $TC_e(n, q)$, is given by:

$$
TC_e(n, q) = \frac{F_R \cdot nq(1 - \gamma)^{q/\lambda}}{2\lambda} + n \cdot O_c \cdot q \cdot \gamma^{q/\lambda} = \frac{F_R \cdot (1 - \gamma)^{q/\lambda} + F_D \cdot q(1 - \gamma)^{q/\lambda} + (C_E + C_D) \cdot (1 - \gamma)^{q/\lambda} + C_D \cdot (1 - \gamma)^{q/\lambda} + C_E \cdot q\cdot \gamma^{q/\lambda}}{2} + O_c \cdot \lambda \cdot \left[1 - (1 - \gamma)^{q/\lambda}\right]
$$

(8)

For the sake of simplicity, we substitute $C_{D'} = C_E + C_D$, and the expected long-run average cost with the consideration of environmental cost is:

$$
TC_e(n, q) = \frac{F_R \cdot nq(1 - \gamma)^{q/\lambda}}{2\lambda} + C_D \cdot (1 - \gamma)^{q/\lambda} + C_D \cdot (1 - \gamma)^{q/\lambda} + C_D \cdot (1 - \gamma)^{q/\lambda} + C_E \cdot q\cdot \gamma^{q/\lambda}\}
$$

(9)

The value of $n$ and $q$ that minimizes the total cost per unit time follows the optimality conditions below.

**Lemma 1.** For a given value of $q$, the optimal value of $n$ always satisfies the following condition:

$$
n^*(n^* - 1) \leq \frac{2FR}{h\gamma^{q/(1 - \gamma)^{q/\lambda}}} \leq n^*(n^* + 1)
$$

(10)

**Proof.** For given values of $q$, the optimal value of $n$ always satisfies the following:

$$
TC_e(n^* - 1) \geq TC_e(n^*) \text{ and } TC_e(n^* + 1) \geq TC_e(n^*)
$$

Using Equation (7), an optimality condition for $n$ is:

$$
n^*(n^* - 1) \leq \frac{2FR}{h\gamma^{q/(1 - \gamma)^{q/\lambda}}} \leq n^*(n^* + 1)
$$

**Lemma 2.** The upper bound of $n$ satisfies the following condition:

$$
n_{max}(n_{max} - 1) \leq \frac{2FR}{h(1 - \gamma)^{1/q}} \leq n_{max}(n_{max} + 1)
$$

(11)

where $n_{max}$ denotes the upper bound of $n$.

**Proof.** The value $\frac{2FR}{h\gamma^{q/(1 - \gamma)^{q/\lambda}}}$ in Equation (10) is a non-increasing function of $q$. Thus, the maximum value of possible $n$ is determined when $q = 1$. 
Lemma 3. For a given value \( n \), the total cost function is a convex function of \( q \). Thus, the optimal dispatch quantity, \( q \), is obtained by taking the first order derivative of the total profit function.

Proof. Taking the first order and the second order partial derivatives of Equation (7) with respect to \( q \), we have:

\[
\begin{align*}
\frac{\partial TC_c(n,q)}{\partial q} &= -\frac{F_q \lambda}{n} + \frac{F_q \gamma}{q^2} + \frac{h(n-1)(1-\gamma)^{\gamma/\lambda}}{2} \\
&\quad + \frac{h(n-1)(1-\gamma)^{\gamma/\lambda}}{2} \ln (1 - \gamma)^{1/\lambda} + (C_R + C_D + O_c) \cdot \lambda \cdot (1 - \gamma)^{\gamma/\lambda} \ln (1 - \gamma)^{1/\lambda} \\
\frac{\partial^2 TC_c(n,q)}{\partial q^2} &= \frac{2F_q \lambda}{nq^2} + \frac{2F_q \gamma}{q^3} + h(n-1)(1-\gamma)^{\gamma/\lambda} \ln (1 - \gamma)^{1/\lambda} \left[ 1 + \frac{q \ln (1 - \gamma)^{1/\lambda}}{2} \right] \\
&\quad + (C_R + C_D - O_c) \cdot \lambda \cdot (1 - \gamma)^{\gamma/\lambda} \ln (1 - \gamma)^{1/\lambda}^2
\end{align*}
\]

Since the second order derivative is always larger than zero, \( TC_c(n,q) \) is convex with respect to \( q \) for a given value of \( n \).

Using the above optimality conditions, we develop a simple enumeration algorithm to obtain the optimal parameters for the proposed policy. The simple enumeration algorithm always guarantees the optimal solution. The procedure is as follows.

- The simple enumeration algorithm (SEA_Q)

Step 1: Compute the upper bound of \( n \) using Equation (11).
Step 2: For all \( n \in \{1, \ldots, n_{\text{max}}\} \), compute the optimal \( q \) using Lemma 3.
Step 3: For given combination of \( n \) and \( q \), compute the total cost \( TC_c(n,q) \) using Equation (9).
Step 4: Select the \( (n,q) \) with the minimum \( TC_c(n,q) \).

3.2. Time-Based Policy to Consideration of Order Cancellation

In this section, we present a mathematical model for optimal time-based dispatch policy. Figure 2 shows the inventory level under the time-based dispatch policy. Unlike quantity-based policy, time-based policy ships out accumulated load in every period that is predetermined and can guarantee delivery time. Thus, the amount of accumulation is not fixed.

![Figure 2. Inventory level under the time-based policy.](image)

The following additional notations are employed in the time-based dispatch policy:
Figure 2 shows the vendor’s inventory level. Under the time-based policy, a new dispatch cycle starts at every $T$ time unit. During the dispatch cycle ($T$), customers may cancel their orders (we assume that order cancellation occurs in proportion to remaining order quantity for every unit time). The remaining order quantity in $i$th $T$ period is $D_i(T)(1 - \gamma)$ at the first unit time after the beginning of a dispatch cycle, $D_i(T)(1 - \gamma)(1 - \gamma)$ at the second unit time, and $D_i(T)(1 - \gamma)^T$ at the end of a dispatch cycle (note that $T$ is dispatch cycle time). Thus, when time reaches the end of a dispatch cycle ($T$), units that a vendor should ship to customers are $D_i(T)(1 - \gamma)^T$.

Let $K$ denote the number of dispatch cycles within a given inventory replenishment cycle. When the inventory level reaches zero, $K$ is computed by:

$$K = \inf \left\{ k : \sum_{i=1}^{k} D_i(T)(1 - \gamma)^T > Q \right\}$$  \hspace{1cm} (14)

The objective is to develop an optimization model to jointly determine the optimal replenishment quantity, $Q$, and the optimal dispatch cycle, $T$, in order to minimize the total cost.

We first compute the expected replenishment cycle length. By the definition, $K$ is a random variable representing the number of dispatch cycles within an inventory replenishment cycle. Thus, the expected replenishment cycle length is:

$$E(ReplenishmentCycleLength) = E[K]T$$  \hspace{1cm} (15)

We now compute four different cost elements (replenishment, dispatch, inventory holding, and order cancellation) during a replenishment cycle:

- **Order Cancellation Cost**: Since the company will not dispatch its products until $T$ units of time lapses, the order cancellation rate is $[1 - (1 - \gamma)^T]$. Thus, the order cancellation cost per dispatch cycle is $O_c \cdot E\left[D_i(T) \cdot (1 - (1 - \gamma)^T)\right]$. Since there are $n$ dispatch cycles in a replenishment cycle, the customer waiting cost per replenishment cycle is:

$$\text{Order Cancellation Cost} = E[K] \cdot O_c \cdot E[D_i(T)] \left[1 - (1 - \gamma)^T\right] = E[K] \cdot O_c \cdot \lambda T \left[1 - (1 - \gamma)^T\right]$$  \hspace{1cm} (16)

- **Replenishment Cost**: The expected replenishment quantity is equal to the expected total demand within a replenishment cycle (see Figure 2). The expected replenishment cost is computed by:

$$\text{Replenishment Cost} = F_R + C_R \cdot E[K] \cdot E[D_i(T)](1 - \gamma)^T = F_R + C_R \cdot E[K] \cdot \lambda T (1 - \gamma)^T$$  \hspace{1cm} (17)

- **Dispatch Cost**: The expected dispatch quantity in a dispatch cycle is $E[D_i(T)](1 - \gamma)^T$, and the expected dispatch cost in a dispatch cycle is $F_D + C_D \cdot E[D_i(T)](1 - \gamma)^T$. There are $E[K]$ dispatch cycles during a replenishment cycle, and thus, the expected dispatch cost during a replenishment cycle is computed by:

$$\text{Dispatch Cost} = F_D \cdot E[K] + C_D \cdot E[K] \cdot E[D_i(T)](1 - \gamma)^T = F_D \cdot E[K] + C_D \cdot E[K] \cdot \lambda T (1 - \gamma)^T$$  \hspace{1cm} (18)
• **Inventory Holding Cost:** Let \( I(t) \) denote the inventory level at time \( t \).

\[
I(t) = \begin{cases} 
Q & \text{if } 0 \leq t \leq T \\
Q - D_1(T)(1 - \gamma)^T & \text{if } T < t \leq 2T \\
\cdots \\
Q - \sum_{j=1}^{K-1} D_j(T)(1 - \gamma)^T & \text{if } (K-1)T < t \leq KT 
\end{cases}
\]

Since holding cost is \( h \), the inventory holding cost is given by:

\[
\text{Inventory Holding Cost} = h \times \left\{ E(K)QT - \lambda T^2(1 - \gamma)^T \frac{E(K)[E(K) - 1]}{2} \right\} \tag{19}
\]

Using the above results, the expected cost during a replenishment cycle is computed by:

\[
E(\text{Replenishment Cycle Cost}) = F_R + C_R \cdot E(K) \cdot \lambda T(1 - \gamma)^T + F_D \cdot E(K) + C_D \cdot E(K) \cdot \lambda T(1 - \gamma)^T \\
+ h \times \left\{ E(K)QT - \lambda T^2(1 - \gamma)^T \frac{E(K)[E(K) - 1]}{2} \right\} + E(K) \cdot O_c \cdot \lambda T \left[ 1 - (1 - \gamma)^T \right] 
\]

(20)

Conversely, the expression for the long-run average profit, \( TP(Q,T) \), is given by:

\[
TC(Q,T) = F_R \frac{E(K)}{Q} + C_R \lambda (1 - \gamma)^T + F_D + C_D \lambda (1 - \gamma)^T \\
+ h \times \left\{ Q - \lambda T(1 - \gamma)^T \frac{E(K)[E(K) - 1]}{2} \right\} + O_c \lambda \left[ 1 - (1 - \gamma)^T \right] \tag{21}
\]

**Lemma 4.** A continuous approximation for \( K \) is provided by an Erlang random variable with a scale parameter \( \lambda T(1 - \gamma)^T \) and a shape parameter \( Q \).

\[
E(K) \approx \frac{Q + 1}{\lambda T(1 - \gamma)^T} \tag{22}
\]

**Proof.** Let \( f(\cdot) \) denote the distribution function of \( D_i(T)(1 - \gamma)^T \), and \( f^{(k)}(\cdot) \) denote the \( k \)-fold convolution of \( f(\cdot) \). From \( K = \inf \left\{ k : \sum_{i=1}^{k} D_i(T)(1 - \gamma)^T > Q \right\} \), we have \( P[K \geq k + 1] = f^{(k)}(Q) \), and thus \( P[K \leq k] = 1 - f^{(k)}(Q) \). Since \( f(\cdot) \) is a Poisson distribution with parameter \( \lambda T(1 - \gamma)^T \), \( k \)-fold convolution of \( f(\cdot) \) is a Poisson distribution with parameter \( \lambda T(1 - \gamma)^T \).

\[
f^{(k)}(Q) = \sum_{i=0}^{Q} \frac{[\lambda T(1 - \gamma)^T]^i}{i!} e^{\lambda T(1 - \gamma)^T} \tag{23}
\]

From Equation (23), we can obtain:

\[
P[K \leq k] = 1 - \sum_{i=0}^{Q} \frac{[\lambda T(1 - \gamma)^T]^i}{i!} e^{\lambda T(1 - \gamma)^T}, k = 1, 2, \ldots
\]

Treating \( k \) as a continuous variable, the right hand side of the above expression is a \( Q \)-stage Erlang distribution function with parameter \( \lambda T(1 - \gamma)^T \) and mean \( (Q + 1) / \lambda T(1 - \gamma)^T \) (refer to [35]).

From Equations (21) and (22), we obtain:

\[
TC(Q,T) = F_R \lambda (1 - \gamma)^T + C_R \lambda (1 - \gamma)^T + F_D + C_D \lambda (1 - \gamma)^T \\
+ h \times \left\{ Q - 1 + \lambda T(1 - \gamma)^T \right\} + O_c \lambda \left[ 1 - (1 - \gamma)^T \right] \tag{24}
\]
For the sake of simplicity, we substitute $Q = Q + 1$, and the expected long-run average cost is:

$$
TC(Q, T) = \frac{F_R(1 - \gamma)^T}{Q} + C_R(1 - \gamma)^T + \frac{F_D}{T} + C_D(1 - \gamma)^T \\
+ h \times \left\{ \frac{Q - 2 + \lambda T(1 - \gamma)^T}{2} \right\} + O_c\left[ 1 - (1 - \gamma)^T \right]
$$

(25)

Since the dispatch quantity is $\lambda T(1 - \gamma)^T$ and the dispatch cycle is $T$, the long-run average environmental cost is $C_E \cdot \lambda T(1 - \gamma)^T / T$. Hence, the long-run average cost with the consideration of environmental cost, $TC_e(n, q)$, is given by:

$$
TC_e(Q, T) = C_E \cdot \lambda T(1 - \gamma)^T / T + TC(Q, T) \\
= \frac{F_R(1 - \gamma)^T}{Q} + C_R(1 - \gamma)^T + \frac{F_D}{T} + (C_E + C_D)\lambda(1 - \gamma)^T \\
+ h \times \left\{ \frac{Q - 2 + \lambda T(1 - \gamma)^T}{2} \right\} + O_c\left[ 1 - (1 - \gamma)^T \right]
$$

(26)

For the sake of simplicity, we substitute $C_D = C_E + C_D$, and the expected long-run average cost with the consideration of environmental cost is:

$$
TC_e(Q, T) = \frac{F_R(1 - \gamma)^T}{Q} + C_R(1 - \gamma)^T + \frac{F_D}{T} + C_D\lambda(1 - \gamma)^T \\
+ h \times \left\{ \frac{Q - 2 + \lambda T(1 - \gamma)^T}{2} \right\} + O_c\left[ 1 - (1 - \gamma)^T \right]
$$

(27)

**Lemma 5.** For a given value of $T$, the optimal value of $Q$ satisfies the following condition:

$$
\overline{Q}'(\overline{Q}' - 1) \leq \frac{2F_R(1 - \gamma)^T}{h} \leq \overline{Q}'(\overline{Q}' + 1)
$$

(28)

**Proof.** For a given value of $T$, the optimal value of $Q$ follows:

$$
TC_e(Q' - 1) \geq TC_e(Q') \quad \text{and} \quad TC_e(Q' + 1) \geq TC_e(Q')
$$

Similarly, from Equation (10), an optimality condition for $Q$ is:

$$
\overline{Q}'(\overline{Q}' - 1) \leq \frac{2F_R(1 - \gamma)^T}{h} \leq \overline{Q}'(\overline{Q}' + 1)
$$

**Lemma 6.** The upper bound of $Q$ satisfies the following condition:

$$
\overline{Q}_{\max}(\overline{Q}_{\max} - 1) \leq \frac{2F_R(1 - \gamma)^T}{h} \leq \overline{Q}_{\max}(\overline{Q}_{\max} + 1)
$$

(29)

where $\overline{Q}_{\max}$ denotes the upper bound of $Q$.

**Proof.** The value $\frac{2F_R(1 - \gamma)^T}{h}$ in Equation (28) is a non-increasing function of $T$. Thus, the maximum value of possible $Q$ is determined when $T = 0$.

**Lemma 7.** For a given value of $Q$, the total cost function is a convex function of $T$. Thus, the optimal dispatch cycle, $T$, is obtained by taking the first order derivative of the total cost function.

**Proof.** Taking the first order and the second order partial derivatives of Equation (23) with respect to $T$, we have:
Since the second order derivative is always larger than zero, \( T_C(\bar{Q}, T) \) is convex with respect to \( T \) for a given value of \( \bar{Q} \).

Using the above optimality conditions, we develop a simple enumeration algorithm to obtain the optimal parameters for the proposed policy. The simple enumeration algorithm always guarantees the optimal solution. The procedure is as follows.

- The simple enumeration algorithm (SEA_T)
  
  Step 1: Compute the upper bound of \( \bar{Q} \) using Equation (29).
  
  Step 2: For all \( \bar{Q} (\bar{Q} = 1, \ldots, \bar{Q}_{\text{max}}) \), compute the optimal \( T \) using Lemma 7.
  
  Step 3: For given combination of \( \bar{Q} \) and \( T \), compute the total cost \( T_C(\bar{Q}, T) \) using Equation (27).
  
  Step 4: Select the \( (\bar{Q}, T) \) with the minimum \( T_C(\bar{Q}, T) \).

4. Numerical Results

4.1. Sensitivity Analysis

In this section, we conduct extensive numerical experiments, and examine the effects of the parameters on the optimal solutions. For the experiments, the same data set used in Cetinkaya et al. (2006) and Hong et al. (2012) is employed. The data set consists of 1024 problem instances, which is a full factorial design of \( C_R = 1; C_D' = 1; \gamma = 0.1; F_R = 40, 80, 160, 320; F_D = 5, 10, 20, 40; h = 1, 2, 4, 8; \) and \( O_c = 2, 4, 8, 16. \) Table 1 summarizes the results of sensitivity analysis.

<table>
<thead>
<tr>
<th>Quantity-Based Policy</th>
<th>Time-Based Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td>( n ) increases ( \bar{Q} ) increases ( T ) increases</td>
<td>( n ) increases ( \bar{Q} ) increases ( T ) increases</td>
</tr>
<tr>
<td>( h ) decreases ( \bar{Q} ) decreases ( T ) decreases</td>
<td>( h ) decreases ( \bar{Q} ) decreases ( T ) decreases</td>
</tr>
<tr>
<td>( O_c ) No impact ( \bar{Q} ) increases ( T ) decreases</td>
<td>( O_c ) No impact ( \bar{Q} ) increases ( T ) decreases</td>
</tr>
</tbody>
</table>

Table 1 (left side) shows the performance of the optimal quantity-based shipment consolidation policy with different parameter values. The replenishment quantity \( (Q = nq) \) and cost increase as \( F_R \) increases from 40 to 320. This agrees with the intuition that if fixed cost of replenishing inventory is high, we order more products to reduce the replenishment cost. In addition, the shipment quantity \( (q) \) increases as \( F_D \) increases from 5 to 40. This agrees with the intuition that if the fixed cost of dispatch shipment is high, we dispatch more orders to reduce the transportation cost. Table 1 also shows that the replenishment quantity decrease as the unit holding cost increases from one to eight. This agrees with the intuition that if the unit holding cost is high, we keep fewer inventories to reduce the holding cost while the shipment quantity decrease as the unit order cancellation cost increases from 2 to 16. This agrees with the intuition that, if the unit order cancellation cost is high, we dispatch fewer orders to reduce the total order cancellation cost because the order cancellation cost increases as shipment quantity increases.
In Table 1 (right side), we observe that the replenishment quantity for optimal time-based shipment consolidation policy and the total cost increase as $F_R$ increases from 40 to 320. This trend is observed for all choices of $F_D$, $h$, and $O_c$. This agrees with the intuition that if fixed cost of replenishing inventory is high, we order more products to reduce the replenishment cost.

As $F_D$ increases, we observe that the dispatch cycle time and the total cost increase. This agrees with the intuition that if $F_D$ is high, we dispatch more orders to reduce the transportation cost. The same trend is also observed for all choices of $F_R$, $h$ and $O_c$.

Given a fixed $F_R$, $F_D$ and $O_c$, we observe that the replenishment quantity decreases as $h$ increases from one to eight. This trend is observed for all choices of $F_R$, $F_D$ and $O_c$. This agrees with the intuition that if the unit holding cost is high, we keep fewer inventories to reduce the holding cost.

As $O_c$ increases, we observe that the dispatch cycle time decreases. This agrees with the intuition that if $O_c$ is high, we dispatch fewer orders to reduce the total order cancellation cost because the order cancellation cost increases as dispatch cycle increases. The same trend is also observed for all choices of $F_R$, $F_D$ and $h$.

Since the effect of changes in $F_D$ is the main motivation of shipment consolidation, and since order cancellation is the main theme in our paper, we analyzed the performance of the consolidation policy with $F_D$ and $O_c$ separately. Table 2 shows the increase of total cost ($TC$) and the decrease of $\Delta TC$ as $F_D$ increases from 5 to 75. Table 3 shows the impact made to the same variables ($TC$ and $\Delta TC$) as $O_c$ increases from 2 to 30.

**Table 2. Result with different $F_D$ ($F_R = 40$, $h = 1$, $\lambda = 2$, $O_c = 2$).**

<table>
<thead>
<tr>
<th>$i$</th>
<th>$F_D$</th>
<th>Quantity-Based Policy</th>
<th>Time-Based Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n$</td>
<td>$q$</td>
</tr>
<tr>
<td>1</td>
<td>5</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>2</td>
<td>10</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>3</td>
<td>15</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>20</td>
<td>8</td>
<td>6</td>
</tr>
<tr>
<td>5</td>
<td>25</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>6</td>
<td>30</td>
<td>8</td>
<td>8</td>
</tr>
<tr>
<td>7</td>
<td>35</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>8</td>
<td>40</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>9</td>
<td>45</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>10</td>
<td>50</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>11</td>
<td>55</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>12</td>
<td>60</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>13</td>
<td>65</td>
<td>13</td>
<td>9</td>
</tr>
<tr>
<td>14</td>
<td>70</td>
<td>14</td>
<td>10</td>
</tr>
<tr>
<td>15</td>
<td>75</td>
<td>14</td>
<td>11</td>
</tr>
</tbody>
</table>

**Table 3. Result with different $O_c$ ($F_D = 5$, $F_R = 40$, $h = 1$, $\lambda = 2$).**

<table>
<thead>
<tr>
<th>$i$</th>
<th>$F_D$</th>
<th>Quantity-Based Policy</th>
<th>Time-Based Policy</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$n$</td>
<td>$q$</td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>3</td>
<td>6</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>8</td>
<td>4</td>
</tr>
<tr>
<td>5</td>
<td>10</td>
<td>9</td>
<td>3</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>10</td>
<td>3</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>11</td>
<td>3</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>12</td>
<td>3</td>
</tr>
<tr>
<td>9</td>
<td>18</td>
<td>13</td>
<td>2</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>14</td>
<td>2</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>15</td>
<td>2</td>
</tr>
<tr>
<td>12</td>
<td>24</td>
<td>16</td>
<td>2</td>
</tr>
<tr>
<td>13</td>
<td>26</td>
<td>17</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>28</td>
<td>18</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>30</td>
<td>19</td>
<td>1</td>
</tr>
</tbody>
</table>
Table 3 shows that optimal dispatch quantity decreases to 1 (i.e., no-consolidation) as order cancellation cost increases. This result can be interpreted as high order cancellation cost deteriorates benefit of shipment consolidation.

In Tables 2 and 3, $\Delta TC / \Delta F_D$ and $\Delta TC / \Delta O_c$ show the amount of changes made to total cost, which is induced by the changes made to $F_D$ and $O_c$. $\Delta TC / \Delta F_D$ and $\Delta TC / \Delta O_c$ are tracked at each value of $F_D$ and $O_c$ in Figure 3. To compare the effects of $F_D$ and $O_c$ on $\Delta TC / \Delta F_D$ and $\Delta TC / \Delta O_c$, respectively, the two variables ($F_D$ and $O_c$) are normalized to (0, 100).

![Figure 3. Decrease in total cost induced by decrease of $F_D$ and $O_c$ in quantity-based policy and time-based policy.](image)

As shown in Figure 3, total cost is more sensitive to increase in order cancellation cost ($O_c$) than fixed dispatching cost ($F_D$). Other things being equal, reducing the order cancellation cost is more advantageous than reducing the fixed dispatching cost in this situation. This finding can help logistics service providers, who are performing shipment consolidation policy, to make decisions to select the cost they should reduce first to minimize total cost.

4.2. Comparison of the Quantity-Based and the Time-Based Policies

To compare the performances of the optimal quantity-based policy with those of the optimal time-based policy, numerical experiments are conducted, and the total costs are compared. For this comparison, we will use the same data in Section 4.1. In total, 1024 problems are generated and solved using both the quantity-based policy with time-based policy. The cost difference is computed by

$$\text{Cost Difference} = \text{Total cost of time-based policy} - \text{Total cost of quantity-based policy}$$

Figure 4 shows the total cost difference between quantity-based and time-based policies, and the total cost difference decreases as the order cancellation rate increases. As shown in Figure 3, if the customer is less sensitive to the waiting time, i.e., the order cancellation rate is small, the quantity-based consolidation policy shows the better performance in terms of the total cost compare with time-based consolidation policy. However, if the order cancellation rate increases, the performance of time-based consolidation policy is better than that of quantity-based policy.
As shown in Figure 3, total cost is more sensitive to increase in order cancellation cost ($c_O$) than fixed dispatching cost ($DF$). Other things being equal, reducing the order cancellation cost is more advantageous than reducing the fixed dispatching cost in this situation. This finding can help logistics service providers, who are performing shipment consolidation policy, to make decisions to select the cost they should reduce first to minimize total cost.

Figure 4. Total cost difference between quantity-based and time-based policies.

5. Conclusions

In this paper, we considered a single-item inventory system where shipments are consolidated to reduce the transportation cost using quantity-based and time-based consolidation policies for the purpose of the sustainability enhancement. We developed mathematical models for quantity-based and time-based policies with order cancellation to minimize the total cost, and optimality properties for the models are then obtained. Efficient algorithms using optimal properties are provided to compute the optimal parameters for ordering and shipment decision. To compare the performances of the quantity-based policy with the time-based policy, extensive numerical experiments are conducted, and the total cost is compared. Numerical results show that the performance of time-based consolidation policy is better than that of quantity-based policy when the order cancellation rate increases.

Acknowledgments: This research was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2015R1D1A1A09058818).

Author Contributions: Chulung Lee came up with the research ideas and initiated the research project as a corresponding author. Kyunghoon Kang developed the research ideas for the results and drew conclusions. Ki-sung Hong developed the mathematical models. Ki Hong Kim proofread and revised the manuscript.

Conflicts of Interest: The authors declare no conflict of interest.

References

5. Ülkü, M.A. Dare to care: Shipment consolidation reduces not only costs, but also environmental damage. Int. J. Prod. Econ. 2012, 139, 438–446. [CrossRef]


© 2017 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).