Impact of inventory sharing on service availability and transportation levels in time-differentiated distribution

Mohsin Nasir Jat* and Raza Ali Rafique

Suleman Dawood School of Business,
Lahore University of Management Sciences,
DHA, Cantt, 54792, Lahore, Pakistan
Email: mohsin.nasir@lums.edu.pk
Email: raza.ali@lums.edu.pk
*Corresponding author

Abstract: This paper studies service availability and transportation performance measures considering a service system with two delivery time-window options. The performance measures are examined through a simulation analysis comparing inventory sharing and non-sharing scenarios under a centralised and a decentralised setup of service facilities. The analysis considers varying demand composition in terms of demand percentages of different service time options. Unlike the existing work in this area, the paper takes into account the spatial factor and argues that considering a uniform cost of alternate shipments may not be appropriate. The analysis highlights that there might be instances where introducing inventory sharing does not implicate an inventory-transportation performance trade off.

Keywords: inventory sharing; service logistics; distribution; service time differentiation.

Reference to this paper should be made as follows: Jat, M.N. and Rafique, R.A. (2017) ‘Impact of inventory sharing on service availability and transportation levels in time-differentiated distribution’, Int. J. Inventory Research, Vol. 4, No. 1, pp. 23–43.

Biographical notes: Mohsin Nasir Jat is an Assistant Professor and the Convener of Transportation and Logistics Research at the Suleman Dawood School of Business, Lahore University of Management Sciences (LUMS). He received his MSc in Operations Management and PhD in Manufacturing Engineering and Operations Management from the Nottingham University Business School, England. He had been engaged with IBM, Fujitsu, and the Department of Transport (England) to study the time aspect of service and its implications on logistics. He teaches courses on supply chain management, optimisation, and simulation, while his research focus is on after-sales, online retail, and disaster relief logistics. He also has a quality management related experience in automotive and construction sectors.

Raza Ali Rafique is an Assistant Professor at Lahore University of Management Sciences (LUMS). He obtained his PhD in Supply Chain Management and Marketing from Rutgers, the State University of New Jersey, USA, where he focused on logistics, supply chain management, and optimisation research methods. He has eight years of practical experience in manufacturing and service firms. His current research interests focus on the development of energy supply chains, optimisation and energy systems.

Copyright © 2017 Inderscience Enterprises Ltd.
1 Introduction

It has now been known for long that buyers of high investment equipment do not want to take their chances with their purchases. They expect a range of after sales services from installations to upgrades. One of the forms of after sales service is to provide service parts at customer sites within different contracted time windows as per different service requirements. Commonly offered by large-scale IT sector companies, such services are facilitated by service parts logistics (SPL) systems through a network of parts stocking facilities.

SPL is a well-researched area in management science with studies mainly employing quantitative modelling for service parts inventory management. Within this area, there is a significant stream of research focusing on inventory sharing, such that if there is a stock-out at a facility when a customer requests a part, the required part is transshipped to the out-of-stock facility from another facility in the system. This risk sharing mechanism allows reduction in the overall inventory levels. It is a common assumption in the existing models that the transportation cost for an alternate supply due to a stock-out is higher and fixed. Hence, an inventory and transportation cost trade-off is directly forced in the models. This paper essentially questions this assumption within a setting where different customers have different service time requirements, and, due to the service time constraints, customers are directly served by an alternate source rather than through transshipment at the facility that originally received the request. The effect of inventory sharing on service cost related performance under a centralised and a decentralised distribution setup is investigated over changes in demand fractions for different time-based service types. The objective is to understand the trade-off between inventory and transportation performance, which is as such not explicitly explored in the existing SPL literature.

The existing service logistics literature can also be criticised for overlooking the spatial factor. As noted by Yang et al. (2013), though the transshipment time for service parts is not negligible, this aspect is hardly considered. Besides, studies considering time-based service differentiation consider customer waiting-times at warehouses, i.e., not at customer sites. The traveling distance and time to reach customers, however, can be an important factor when service times are short, e.g., typically few hours for IT equipment support services. Unlike the existing studies, the paper only considers inventory sharing if an alternate supply can be made to the requesting customer site within the required service time window.

The investigation is based on stylised settings in a Cartesian system and employs a discrete event simulation model. However, several of the characteristics incorporated in the stylised model are influenced by the study of the SPL systems of IBM and Fujitsu in the UK. The service availability level within the requested service time window and the average service distance are studied as the system’s performance measures. The
Simulations under the decentralised setup confirm that inventory sharing among facilities positively impacts the service availability level but increases the transportation cost. A higher fraction of demand for the service in the longer time window increases the inventory sharing opportunity. This, in turn, increases the service availability levels (which may allow stock reductions) and the average distance to serve a demand. The case of inventory sharing under the centralised setup is interesting as there are instances that do not reveal a trade-off between the service availability and average service distance performances. Under the centralised setup, sharing can result in both higher service availability and lower average service distance, particularly when the demand fraction for the service in the longer time window is high.

The rest of the paper is organised as follows. Section 2 provides a brief literature review focusing on the studies considering inventory sharing in the context of SPL. Section 3 provides a formal description of the distribution system, the distribution strategies studied, and the performance measures. Sections 4 and 5 present the simulation model and analysis respectively. The paper concludes in Section 6.

2 Literature review

A considerable amount of research has been conducted on service parts inventory management specifically considering lateral transshipments or inventory sharing. Lateral transshipments allow members of an inventory system to pool their inventories, which can allow them to lower inventory levels and costs whilst still achieving the required service levels. Paterson et al. (2011) provide a comprehensive review of the literature on inventory sharing in general. Patriarca et al. (2016) outline the more recent developments in the area. The earliest service parts inventory sharing study reported by Paterson et al. (2011) is Lee (1987). Grounded on a basic model for repairable service parts, Lee’s study shows that lateral transshipment results in savings due to less stock being needed at the stock bases for a service availability level. A number of studies then extend these findings by relaxing the assumptions of the basic model by using un-identical locations, a different demand process, a different order policy, or by considering echelons. Other extensions in this area are brought by considering transshipments in determining the optimum stocking levels, ordering policy, and network design (see Paterson et al., 2011). A simplification considered by the pre 2011 studies is that a transshipment happens instantaneously in SPL, i.e., there is no transshipment lead-time. The recent studies by Axsiöter (2014), Olsson (2015), and Yang et al. (2013) question this assumption and incorporate transshipment lead-time in their models. There are though few studies that take the time factor into account, be it the transshipment lead-time at a warehouse, or the total time to serve a customer.

As far as SPL is concerned, along with the fill-rate level, i.e., the percentage of demand satisfied from stock on hand, the time taken by a service provider to provide a service part is also important. To the best of our knowledge, there are only five studies that explicitly consider time-based service levels with inventory sharing. Considering a stylised system, Kutanoglu (2008) analyses the impact of stock level on service cost and the percentage of total demand met within certain time windows. Kutanoglu and Mahajan (2009) present a model to determine the minimum-cost stock levels at facilities considering system wide constraints on the percentage of demand that is met within
certain time windows. Iyoob and Kutanoglu (2013) and Yang et al. (2013) present models with the objective similar to that of Kutanoglu and Mahajan (2009) while considering demand-facility allocation decision and stocks in replenishment pipeline respectively. Jalil et al. (2011), adapting the model by Kutanoglu and Mahajan (2009), analyse the planning performance in relation to the quality of installed-base data.

Though the studies described above consider time-based service levels, they do not consider different service time requirements within the customer-base. There is a dearth of literature that focuses on time-based differentiation amongst customers while also incorporating inventory sharing. The three studies to do so are by Kranenburg and van Houtum (2007, 2008, 2009). These studies look at determining the optimal stock level that minimises the service cost while meeting the average waiting time target of each customer group. An important distinction is that these studies only consider customer waiting-times at warehouses. Hence, the alternate shipment is made to warehouses and not to customer locations. The travel distance and the time to reach customers are not considered, which can be important factors that determine whether the desired service level is achieved.

The research on inventory sharing in the context of SPL primarily focuses on inventory management in isolation of the customer location aspect in service availability. Studies that do incorporate this feature are limited because they do not differentiate between groups of customers and their service requirements. Moreover, the literature assumes an inventory-transportation cost trade-off wherein inventory sharing leads to a higher transportation cost and a lower inventory cost (Kutanoglu, 2008). This cost trade-off is commonly either implicitly assumed or explicitly stated in inventory sharing literature. Examples of the SPL studies that assume this trade-off are Kranenburg and van Houtum (2007, 2008, 2009), Kutanoglu (2008), Kutanoglu and Mahajan (2009), Olsson (2015) and Yang et al. (2013). There lacks an effort on understanding the nature of the inventory and transportation performance trade-off under inventory sharing.

3 System definition

In this section, we formally introduce the system and performance measurement assumptions on which the investigation is based. Consider a convex bounded plane in which customers with identical demand are uniformly distributed. The customers are supplied with service parts within two different service time commitments. The service that ensures supply within the shorter time window is referred to as the ‘strict’ service, while the other as the ‘relaxed’ service. Travelling distances inside the area are Euclidean and are proportional to time. The service time windows implicate the maximum distances, referred as ‘service distance constraints’, which can be travelled from a facility to provide a service. There exists a distribution system in the area comprising facilities which, considering the strict service distance constraint (relaxed service distance constraint being redundant), cover the entire area efficiently by being located in the middle of their catchment regions. A catchment region comprises all the points that are closer to the respective facility than to any other facility (i.e., a Voronoi cell). Each facility in the system operates a one-for-one ($S–1$, $S$) inventory policy, typically considered appropriate for service parts. Under this policy, $S$ is kept as the base-stock level and a replenishment order is triggered as soon as a part is extracted from the inventory. Note that the Euclidean distance assumption is a common assumption for
problem simplification. The works by Simchi-Levi (1992) and Drèze et al. (2008) are examples of service system studies that consider Euclidean distances and efficient service area packing with facilities located in the middle of their service areas.

3.1 Distribution strategies

A non-hierarchical system and a hierarchical system are studied as a decentralised and a centralised distribution alternative respectively (Figures 1 and 2). Under the decentralised (non-hierarchical) setup, each facility provides both strict and relaxed services in its catchment area (region comprising all locations that are closer to this facility than to any other facility). To set-up a centralised (hierarchical) system, a subset of facilities is designated as higher level facilities, which are just sufficient to cover the entire area considering the relaxed service distance constraint. Under this setup, all facilities provide the strict service, but, only the higher level facilities provide the relaxed service to achieve centralisation or demand consolidation. Facilities other than the higher level facilities are lower level facilities (only providing the strict service). A higher level facility has two service areas; a smaller service area where it provides both relaxed and strict services (Voronoi region considering all the facility points) and a larger service area where it provides only the relaxed service (Voronoi region only considering the higher level facility points). Note that centralisation is known to reduce the required inventory level at the cost of higher transportation costs due to increased average service distance. Our analysis confirms that, compared to the decentralised system, the inventory level under the centralised system is lower while the transportation level is higher. A centralised system can be considered favourable in situations where inventory costs dominate (see Jat and Muyldermans, 2013).

Note that the service regions are polygonal while the service ranges are circular and inevitably overlap. The overlapping ranges provide opportunities for alternate service within a service time window.

Let the local facility for a service request be the nearest facility offering the requested service type. That is, under the centralised setup, in case of a relaxed service request, the local facility is the nearest higher-level facility, and, in the case of a strict service request, the local facility is the nearest facility, be it a higher or a lower level facility. Under the decentralised setup, for either a relaxed or a strict service request, the local facility is the nearest facility (as all facilities provide the full range of service options). Both centralised and decentralised setups are investigated with and without inventory sharing considering the following configurations:

- **Non-sharing:** A service request is only met by the local facility. In case of a stock-out at the local facility at the time of a request, the service request is back ordered. Hence, facilities only provide the service in associated Voronoi cells. This makes it fairly straightforward to estimate the average service distance and the service availability level as discussed later.

- **Sharing:** When a service request is registered, stocks are checked at the local facility providing the required service type. In case of a stock-out at the local facility, service can be met from the closest facility (higher-level or lower-level) with stocks in the distance range. If none of the facilities in the distance range has stocks at the time of the request, a backorder is recorded at the local facility.
Hence, under the decentralised setup without inventory sharing, service by each facility is restricted within its Voronoi region (marked by straight lines in Figure 1). With inventory sharing under the decentralised setup, if a facility cannot fulfill a customer request within its Voronoi region, the customer can get the demand fulfilled from the nearest facility which has the stock and has the customer within the service range. Note that there is a limited opportunity for inventory sharing in cases of strict service requests, as the overlapping of the strict service ranges is small and only on the borders of the catchment areas. For the relaxed service requests, the inventory sharing opportunity is significantly more, as the relaxed service ranges are larger. For example, facility $F_1$ in Figure 1 covers large parts of the adjacent facilities’ catchment areas under its range for the relaxed service.

Under the centralised setup with no inventory sharing, strict service requests are fulfilled by facilities only within their Voronoi regions considering all facility points, while relaxed service requests are fulfilled by higher level facilities only within their Voronoi regions considering the higher level facility points (marked by the dotted straight lines in Figure 2). For example, in Figure 2, customers $C_1$ and $C_2$ can respectively get the strict and the relaxed service only from facility $F_1$. With inventory sharing, $F_2$ can be an alternate facility for providing strict service to $C_1$ and relaxed service to $C_2$. 

Figure 1  Decentralised distribution

![Decentralised distribution](image)
3.2 Performance measures

The distribution strategies are compared on the basis of the service availability and transportation levels measured in terms of the fill-rate and average service distance. The terms ‘availability level’ and ‘fill-rate’ are used interchangeably and refer to the percentage of demand met from stocks on hand. The term ‘average service distance’ refers to the average distance to reach a demand point.

Let,

- \( A \) rectangular area with uniform geographical distribution of customers
- \( \lambda \) overall demand in the area over unit time generated through a Poisson process (resulting in time between two service requests having an exponential distribution)
- \( L \) replenishment lead time
- \( f \) fraction of demand for the strict service
\[ 1 - f \] fraction of demand for the relaxed service  
\[ Fr \] minimum service availability (fill-rate) level  
\[ n \] total number of facilities  
\[ \lambda_i \] demand served by facility \( i \) over unit time  
\[ S_i \] base stock level at facility \( i \)  
\[ A_{si} \] size of service area covered by facility \( i \) for the strict service  
\[ A_{ri} \] size of service area covered by facility \( i \) for the relaxed service  
\[ T_{si} \] average distance from facility \( i \) to reach a demand point for the strict service in \( A_{si} \)  
\[ T_{ri} \] average distance from facility \( i \) to reach a demand point for the relaxed service in \( A_{ri} \).  

Note that \( T_{ri} = T_{si} \) and \( A_{ri} = A_{si} \) under the decentralised (non-hierarchical) setup.  
\[ T_s \] average service distance for the strict service  
\[ T_r \] average service distance for the relaxed service.

Considering no inventory sharing, the base stock level at facility \( i \) (\( S_i \)) under the \((S-1, S)\) policy with backorders can be determined considering the steady state probability of the quantity of units in resupply. Note that considering the Poisson demand process, there is a well-known relationship between stock-out probability and fill-rate under the \((S-1, S)\) inventory policy (Muckstadt, 2005, 2010; Zipkin, 2000). Fill-rate at facility \( i \) is the probability that demand at the facility is less than \( S_i \) over the lead time \( P(\lambda_i < S_i) \).

\[
\sum_{x=S_i} e^{-\lambda_i} \frac{\lambda_i^x}{x!},
\]

where \( e^{-\lambda_i} \frac{\lambda_i^x}{x!} \) is the unconditional probability that \( x \) units remain in the resupply at facility \( i \) and \( \lambda_i \) is the sum of the relaxed and strict service demand faced by facility \( i \). Due to demand being uniform in the area, under both centralised and decentralised systems with no inventory sharing, \( \lambda_i \) equals to \((1 - f)\lambda_{if} \left( \frac{A_{ri}}{A} \right) + \beta \left( \frac{A_{si}}{A} \right) \), where \( \frac{A_{ri}}{A} \) and \( \frac{A_{si}}{A} \) are the fractions of relaxed and strict service demand faced by facility \( i \) respectively. \( \lambda_i \) reduces to \( \beta \left( \frac{A_{si}}{A} \right) \) in case facility \( i \) is a lower level facility as \( A_{ri} \) equals to zero. \( S_i \) can be set through an iterative procedure, starting with \( S_i = 1 \) and incrementing \( S_i \) by 1 each time the fill-rate is found to be less than \( F_r \). Summing base stock levels at all facilities gives the total inventory (base-stock) in the system \( \left( \sum_{i=1}^{n} S_i \right) \).

The overall average service distance can also be determined computationally when considering no inventory sharing. The average distance from facility \( i \) to reach a demand point for the relaxed service (\( T_{ri} \)) and to reach a demand point for the strict service (\( T_{si} \)), i.e., the average distance from facility \( i \) inside \( A_{ri} \) and \( A_{si} \) respectively, can be computed ...
through the procedure outlined by Stone (1991). The average service distance for the relaxed service in the overall area \( (T_r) \) can be determined as

\[
\sum_{i=1}^{n} T_r \left( \frac{A_{ri}}{A} \right)
\]

where \( (A_{ri}/A) \) is the proportion of demand for which the average service distance is \( T_r \). Similarly, the average service distance for the strict service \( (T_s) \) can be determined as

\[
\sum_{i=1}^{n} T_s \left( \frac{A_{si}}{A} \right)
\]

The weighted average of \( T_r \) and \( T_s \) considering the fractions of demand for the strict and relaxed services then gives:

\[
\text{Overall average service distance} = (1 - f) (T_r) + f (T_s)
\]  

Note that equations (1) and (2) apply to both decentralised and centralised setups with no inventory sharing. In equation (1), \( \lambda_i \) can represent demand at a facility in a decentralised or a centralised setup. Of course the value of \( \lambda_i \) differs depending on whether facility \( i \) is a facility in a decentralised system, a higher level facility in a centralised system, or a lower level facility in a centralised system. Similarly, \( T_r \) in equation (2) varies depending on whether the setup is decentralised or centralised.

When considering inventory sharing between facilities, estimating fill-rates becomes challenging even with strict assumptions on maximum stock levels and the number of facilities in an inventory sharing pool (e.g., base-stock \( \leq 1 \) and two sharing facilities) (Iyoob and Kutanoglu, 2013). Under the considered settings, there can be several overlapping areas of different forms covered by multiple facilities, making the estimation of both fill-rates and average service distance complex. For this reason, discrete event simulations are run to estimate the performance measures. The scenarios are simulated with the starting inventory positions (base stock level at facilities) computed considering no inventory sharing (1).

4 The simulation model

Figures 3 and 4 present simulation flowcharts for the non-sharing and sharing configurations respectively.

Inter-arrival times between two requests are generated according to the exponential distribution. The type of service requested is randomly determined considering the proportions of demand for both service types. Demand locations (x and y coordinates in the overall area) are generated considering the uniform distribution such that there is an equal probability of any point in the plane being selected. Service availability levels at each facility and the overall average service distance are recorded as the model’s performance measures.

Let \( R_i \) be the number of service requests received at facility \( i \), and \( b_i \) be the number of instances in which a service request at facility \( i \) is backordered. The service availability level at facility \( i \) is then equal to \( 1 - (b_i / R_i) \).

If \( d_j \) is the Euclidean distance between the location from where the service request \( j \) is registered and the location of facility from where the request is met, such that \( j = 1 \ldots k \) \((k \text{ being the total number of requests})\), then the overall average service distance equals to \( (\Sigma d_j) / k \).
The simulation model is implemented through a program coded in C++. Each facility is represented as an object, storing information about its x and y coordinates on the plane, base stock level, current stock level, number of requests received at the facility, number of requests not fulfilled by the facility from stock on hand (denial counter), and orders in the pipeline and their arrival times (order list). To verify that the program components perform as anticipated, information stored in the variables and the data members of interest are observed for a process walkthrough. Under the non-sharing configuration, comparison of simulation and computational results also provides a verification of the model’s performance. Common random numbers are used as a variance reduction technique to compare the alternative system configurations. The configurations (sharing and non-sharing) are simulated in a sequence. Before starting the simulation runs for each configuration, the random number streams are reinitialised so that same random numbers are generated for each configuration. As a result, under each configuration, the sequence of service request arrival times and the associated service types and demand locations are the same. This provides more confidence that any observed differences in the performance are due to the differences in the system configuration rather than due to the fluctuations of the experimental conditions.
Figure 4  Simulation flow chart – sharing mechanism

- Determine the location of request, i.e., demand point (random).
- Determine the type of service request (random).
- Update order lists and current stock levels at all facilities as per the request arrival time.
- Determine candidate facilities – all the facilities, of any type, in the range.
- Determine the closest facility (Local Facility) from the candidate facilities.

- Determine current stocks at the Local facility.
  - No Stock:
    - Reduce current stock level at the Local facility by 1.
    - Place replenishment order.
    - Set order arrival time to 'current time + replenishment time'.
    - Push order arrival time at the back of Local facility’s order list.
  - Stock Available:
    - Reduce current stock level at the closest candidate facility with stocks by 1.
    - Place replenishment order.
    - Set order arrival time to 'current time + replenishment time'.
    - Push order arrival time at the back of the order list of the closest candidate facility with stocks.
    - Consider service distance as the distance between the closest candidate facility with stocks and the demand point.

- Increment the Local facility's denial counter by 1.
- Reduce the current stock level at the Local facility by 1.
- Place replenishment order.
- Place replenishment order (setting order arrival time to 'current time + replenishment time').
- Push order arrival time at the back of Local facility’s order list.
- Consider service distance as the distance between the Local facility and the demand point.
5 Analysis

As with most simulation models, a number of parameters need to be estimated to perform the analysis. We are not aware that a variation in the following experimental settings will change the conclusions. The overall area is considered to be a unit square defined by corner points \((0, 0), (0, 1), (1, 1)\) and \((1, 0)\), with a distribution system with nine facilities. The Voronoi diagram generated by these facility points is a regular square packing, with the facilities located at the centre of their respective regions (Figure 5). Note that squares are one of the three regular polygons (others being hexagons and triangles) that can tile a plane to provide a complete coverage. A square partitioning provides a simple setup. As noted by Lösch (1954), a seminal contributor to regional science, as a region, square is better than triangle, not much inferior to the hexagon, and has an added advantage of simply drawn boundaries.

The strict service distance constraint is considered to be the maximum distance inside these catchment areas, i.e., approximately 0.236 unit length. Under the decentralised setup, all of the facilities provide both strict and relaxed services within these catchment areas. The relaxed service distance constraint is considered to be twice the strict service distance constraint, i.e., 0.472 unit length. The SPL system that we have studied and those reported in literature provide service time options such that the ratio between the service times is 1/2. With this ratio between the strict and relaxed service times, a hierarchy of facilities and service areas can be setup by designating the first and the last facilities in the top and bottom rows of the facilities as higher-level (Figure 6). The resulting service areas for the relaxed service have the maximum distance that is twice that of the strict service distance constraint.

Figure 5  Facility locations (decentralised setup)
The following results are based on the overall demand of four parts per day, one-day replenishment lead-time, and the minimum service availability level (fill-rate) as 95%. These parameter values are reasonable estimates in view of the UK SPL systems of IBM and Fujitsu. Similar results are found when 25 facilities are considered with square partitioning of the area and when the demand rate is considered to be three parts per day. One simulation is run for six years (the average life cycle of products reported by the case companies) and simulations are run 100 times for each set of service availability and average service distance performance measures. Increasing the number of runs does not change the outputs significantly. For example, in the following experiments, increasing the runs from 100 to 150 provides the same overall fill-rate at three decimal places and the same average distance at four decimal places. The output is presented in the Appendix, which also shows that the simulation results with the non-sharing configuration (entries under ‘FrN–s’ and ‘ADN–s’) closely match the computational results (entries under ‘Fr’ and ‘AD’). Note that the base-stock level at a facility (S in Tables A1 and A3 in Appendix) is set considering the total demand in the facility’s service area (λ in Tables A1 and A3) and the above mentioned minimum service availability level (95%). Also, note that the demand, and hence the base stock level, is the same at each facility in the decentralised system, as all facilities serve in equal service areas for both service types. In the centralised system, where only the higher level facilities serve the relaxed service demand, with the increase in (1–f), more demand gets consolidated at the higher level facilities while the demand at lower level facilities reduces (see Tables A1 and A3).

Figures 7 and 8 show the service availability levels and the average service distances under the decentralised setup.

**Figure 6** Centralised setup

![Centralised setup diagram](image-url)
Figure 7  (a–h) performance under the decentralised setup (see online version for colours)

<table>
<thead>
<tr>
<th>(1–f)*</th>
<th>Service availability level (fill-rates)</th>
<th>Average service distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td><img src="image1.png" alt="Graph" /></td>
<td><img src="image2.png" alt="Graph" /></td>
</tr>
<tr>
<td>0.4</td>
<td><img src="image3.png" alt="Graph" /></td>
<td><img src="image4.png" alt="Graph" /></td>
</tr>
<tr>
<td>0.6</td>
<td><img src="image5.png" alt="Graph" /></td>
<td><img src="image6.png" alt="Graph" /></td>
</tr>
<tr>
<td>0.8</td>
<td><img src="image7.png" alt="Graph" /></td>
<td><img src="image8.png" alt="Graph" /></td>
</tr>
</tbody>
</table>

Note: *Demand fraction for relaxed services.

Under the decentralised setup with the sharing configuration, an increase in the demand fraction for the relaxed service increases both service availability level and the average
service distance (Figures 7 and 8). Inventory sharing in this case can allow a reduction in the total stock in the system from 27 (three units at each facility (Table A1) to 18 (two units at each facility) while still meeting the minimum availability level of 95%; a 33% reduction in the total inventory. Note that without inventory sharing, the service availability level at all facilities is uniform because all facilities face equal demand and have the same base stock level. It can also be observed that the highest increase in the service availability level due to inventory sharing is at facility five, which has the highest number of neighbouring facilities [Figures 7(a), 7(c), 7(e) and 7(g)].

Figures 9(a)–9(h) present the performance measures under the centralised setup.

**Figure 8** (a and b) system wide performance under the decentralised setup (see online version for colours)
Figure 9  (a–h) performance under the centralised setup (see online version for colours)

<table>
<thead>
<tr>
<th>(1–f)*</th>
<th>Service availability level (fill-rates)</th>
<th>Average service distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td><img src="image1.png" alt="Chart" /></td>
<td><img src="image2.png" alt="Chart" /></td>
</tr>
<tr>
<td>0.4</td>
<td><img src="image3.png" alt="Chart" /></td>
<td><img src="image4.png" alt="Chart" /></td>
</tr>
<tr>
<td>0.6</td>
<td><img src="image5.png" alt="Chart" /></td>
<td><img src="image6.png" alt="Chart" /></td>
</tr>
<tr>
<td>0.8</td>
<td><img src="image7.png" alt="Chart" /></td>
<td><img src="image8.png" alt="Chart" /></td>
</tr>
</tbody>
</table>

Note: *Demand fraction for relaxed services.
The output again shows that the sharing configuration results in higher service availability levels compared to the non-sharing configuration. The sharing configuration has greater impact on the higher-level facilities (facilities 1, 3, 7 and 9) in terms of service availability. Because the higher-level facilities provide the relaxed service, having larger service radius, there is more opportunity for an alternate service for demand at these facilities. The output also shows that the sharing configuration can even result in an improved average service distance performance [Figures 9(f) and 9(h)]. With the sharing configuration under the centralised setup, in case of a stock-out when a strict service request is received, compared to the local facility, an alternate facility to meet the demand is at a longer distance. However, in case of a stock-out when a relaxed service request is received, the first alternative to meet the demand is a lower level facility, which is likely to be at a shorter distance from the demand point than the higher-level facility originally assigned to meet the request (Figure 10).

With a higher demand fraction for the relaxed service, the shorter distances to meet the relaxed service requests from alternate facilities can offset the longer distances for alternative service for the strict service requests. Hence with a higher fraction of demand for the relaxed service, the alternate services for the relaxed service requests can not only positively impact the service availability levels, but also reduce the average service distance.

**Figure 10** Inventory sharing under the centralised setup (see online version for colours)
6 Conclusions

Characteristically, inventory sharing among facilities should positively impact on the availability level while increasing the transportation cost as sharing results in shipments from non-local facilities. This is confirmed under the decentralised (non-hierarchical) setup, where a higher fraction of demand for the service in the longer time window increases the inventory sharing opportunity, which in turn increases the service availability levels and the average distance to serve demand. The analysis highlights that considering a uniform transportation cost for an alternate service in an inventory sharing system may be a strict assumption. The cost depends on the distance from where an alternate supply arrives, and, when the fraction of demand for the relaxed service becomes higher, the overall transportation cost due to inventory sharing increases. The case of sharing under the centralised setup on the other hand does not reveal a clear trade-off between the service availability and average service distance performances. Under the centralised setup, sharing can perform better in terms of both service availability and average service distance performance measures when the demand fraction for the service in the longer time window is high.

Though based on basic settings, the work in this paper has critically examined the notion of inventory and transportation performance trade-off and has provided some novel insights. The following managerial insights for time-differentiated distribution systems can be derived from the analysis:

- If a distribution system is setup as decentralised owing to the transportation costs dominating the inventory costs, introducing an inventory sharing mechanism can result in higher overall costs due to an increase in average service distance.
- If a distribution system is setup as centralised due to inventory costs being dominating and the fraction of demand for the service in the longer time window is high, an inventory sharing mechanism can not only increase the service availability level (allowing stock reduction) but also reduce the average service distance.

Acknowledgements

The authors would like to thank Mr. Wajih Shafiq, Research Associate at the School of Business, LUMS, for his assistance on this work.

References


Impact of inventory sharing on service availability and transportation


Appendix

Computational and simulation output

**Notations**

- $F#$: Facility number.
- $\lambda$: Local demand over lead-time at a facility.
- $l-f$: Fraction of demand for relaxed service.
- $S$: Base stock level (computed through the procedure in Section 3.2).
- $Fr$: Fill-rate (availability level) (computed through the formula in Section 3.2) considering no inventory sharing.
- $Fr_{NS}$: Fill-rate determined through simulation considering no inventory sharing.
- $Fr_S$: Fill-rate determined through simulation considering inventory sharing.
- $AD$: Average distance to reach a demand location for service (computed numerically through the procedure explained in Section 3.2) considering no inventory sharing.
- $AD_{NS}$: Average distance to reach a demand location for service determined through simulation considering no inventory sharing.
- $AD_S$: Average distance to reach a demand location for service determined through simulation considering inventory sharing.

**Decentralised setup**

Table A1 Fill-rates under the decentralised setup

<table>
<thead>
<tr>
<th>$l-f$</th>
<th>$\lambda$</th>
<th>$S$</th>
<th>$Fr$</th>
<th>$Fr_{NS}$</th>
<th>$Fr_S$</th>
<th>$l-f$</th>
<th>$\lambda$</th>
<th>$S$</th>
<th>$Fr$</th>
<th>$Fr_{NS}$</th>
<th>$Fr_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98974</td>
<td>0.994432</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98974</td>
<td>0.995764</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98946</td>
<td>0.995478</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98946</td>
<td>0.99673</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98950</td>
<td>0.993955</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98950</td>
<td>0.995653</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98930</td>
<td>0.995191</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98930</td>
<td>0.996277</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98915</td>
<td>0.996654</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98915</td>
<td>0.997621</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98921</td>
<td>0.995542</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98921</td>
<td>0.999667</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98964</td>
<td>0.994332</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98964</td>
<td>0.995668</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98905</td>
<td>0.995398</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98905</td>
<td>0.996666</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98929</td>
<td>0.994081</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98929</td>
<td>0.99565</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>$l-f$</th>
<th>$\lambda$</th>
<th>$S$</th>
<th>$Fr$</th>
<th>$Fr_{NS}$</th>
<th>$Fr_S$</th>
<th>$l-f$</th>
<th>$\lambda$</th>
<th>$S$</th>
<th>$Fr$</th>
<th>$Fr_{NS}$</th>
<th>$Fr_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98974</td>
<td>0.997179</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98974</td>
<td>0.998566</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98946</td>
<td>0.997831</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98946</td>
<td>0.999027</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98950</td>
<td>0.997021</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98950</td>
<td>0.998216</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98930</td>
<td>0.99747</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98930</td>
<td>0.998726</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98915</td>
<td>0.998321</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98915</td>
<td>0.999277</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98921</td>
<td>0.997909</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98921</td>
<td>0.999894</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98964</td>
<td>0.997204</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98964</td>
<td>0.998624</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98905</td>
<td>0.997973</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98905</td>
<td>0.998913</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98929</td>
<td>0.996877</td>
<td>0.44</td>
<td>3</td>
<td>0.98948</td>
<td>0.98929</td>
<td>0.998339</td>
<td></td>
</tr>
</tbody>
</table>

Note: Maximum 90% confidence interval for $Fr_{NS}$ and $Fr_S$ ($t_{90.0.5}$): ± 0.00196.
### Table A2  Average distance to reach a demand location for service under the decentralised setup

<table>
<thead>
<tr>
<th>$(1-f)$</th>
<th>$AD$</th>
<th>$AD_{NS}$</th>
<th>$AD_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.127533</td>
<td>0.127548</td>
<td>0.128003</td>
</tr>
<tr>
<td>0.4</td>
<td>0.127533</td>
<td>0.127548</td>
<td>0.128236</td>
</tr>
<tr>
<td>0.6</td>
<td>0.127533</td>
<td>0.127548</td>
<td>0.128459</td>
</tr>
<tr>
<td>0.8</td>
<td>0.127533</td>
<td>0.127548</td>
<td>0.128722</td>
</tr>
</tbody>
</table>

Note: Maximum 90% confidence interval for $AD_{NS}$ and $AD_S$ ($t_{99,0.5}$): ± 0.000091.

### Centralised setup

### Table A3  Fill-rates under the centralised setup

<table>
<thead>
<tr>
<th>$(1-f)$</th>
<th>0.2</th>
<th>0.4</th>
<th>0.6</th>
<th>0.8</th>
</tr>
</thead>
<tbody>
<tr>
<td>$F#$</td>
<td>$\lambda$</td>
<td>$S$</td>
<td>$Fr$</td>
<td>$Fr_{NS}$</td>
</tr>
<tr>
<td>1</td>
<td>0.56</td>
<td>0.56</td>
<td>0.98105</td>
<td>0.98180</td>
</tr>
<tr>
<td>2</td>
<td>0.36</td>
<td>0.36</td>
<td>0.99425</td>
<td>0.99410</td>
</tr>
<tr>
<td>3</td>
<td>0.56</td>
<td>0.56</td>
<td>0.98105</td>
<td>0.98122</td>
</tr>
<tr>
<td>4</td>
<td>0.36</td>
<td>0.36</td>
<td>0.99425</td>
<td>0.99427</td>
</tr>
<tr>
<td>5</td>
<td>0.36</td>
<td>0.36</td>
<td>0.99425</td>
<td>0.99412</td>
</tr>
<tr>
<td>6</td>
<td>0.36</td>
<td>0.36</td>
<td>0.99425</td>
<td>0.99444</td>
</tr>
<tr>
<td>7</td>
<td>0.56</td>
<td>0.56</td>
<td>0.98105</td>
<td>0.98129</td>
</tr>
<tr>
<td>8</td>
<td>0.36</td>
<td>0.36</td>
<td>0.99425</td>
<td>0.99420</td>
</tr>
<tr>
<td>9</td>
<td>0.56</td>
<td>0.56</td>
<td>0.98105</td>
<td>0.98073</td>
</tr>
</tbody>
</table>

Note: Maximum 90% confidence interval for $Fr_{NS}$ and $Fr_S$ ($t_{99,0.5}$): ± 0.00353

### Table A4  Average distance to reach a demand location for service under the centralised setup

<table>
<thead>
<tr>
<th>$(1-f)$</th>
<th>$AD$</th>
<th>$AD_{NS}$</th>
<th>$AD_S$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2</td>
<td>0.145110</td>
<td>0.145113</td>
<td>0.145231</td>
</tr>
<tr>
<td>0.4</td>
<td>0.162687</td>
<td>0.162655</td>
<td>0.162886</td>
</tr>
<tr>
<td>0.6</td>
<td>0.180265</td>
<td>0.180338</td>
<td>0.180218</td>
</tr>
<tr>
<td>0.8</td>
<td>0.197842</td>
<td>0.197973</td>
<td>0.19785</td>
</tr>
</tbody>
</table>

Note: Maximum 90% confidence interval for $AD_{NS}$ and $AD_S$ ($t_{99,0.5}$): ± 0.00019.