CONTINUOUS-REVIEW REPLENISHMENT WITH CAPACITY CONSTRAINTS AND BREAKDOWNS

S. T. Enns

Department of Mechanical and Manufacturing Engineering
Schulich School of Engineering
University of Calgary
Calgary, Alberta, Canada, T2N IN4

enns@ucalgary.ca

Abstract

This study examines the logic and performance of continuous-review replenishment systems. Specifically, the two-bin, single-card Kanban and reorder point systems are compared. A simple capacity-constrained supply chain with independent, time-varying customer demand is modeled. Material inputs come from a supplier, based on replenishment orders. These inputs are processed in batches at a machine that is subject to breakdowns, allowing the effects of potential inventory buildup to be examined. Experiments are conducted using discrete-event simulation and performance is monitored in terms of inventory levels and delivery performance. Performance tradeoff curves are generated to compare the replenishment systems over a range of service levels. These curves are generated on the basis of near optimal decision variable settings, such as lot sizes, Kanban cards and reorder points. Results show that the two-bin system performs poorly. Large lot sizes are required to provide sufficient inventory to meet demand during the replenishment cycle. The performance differences between the Kanban and reorder point systems are small. For the scenario modeled, the Kanban system benefits incurred through restricting the maximum inventory during machine breakdowns appears to approximately match the reorder point system benefits incurred as a result of using backorder information.

Keywords: Inventory Replenishment, Supply Chain, Two-Bin, Kanban, Reorder Point

Presenting Author's biography

SILVANUS T. ENNS is an Associate Professor at the University of Calgary, Canada. His research interests lie in job shop, batch production and supply chain modeling and analysis.

ISBN 978-3-901608-32-2

1 Introduction

This study examines the logic and performance of three common continuous-review systems for inventory replenishment. These are the two-bin, the single-card Kanban and the reorder point systems. Each of these systems has similarities and differences that affect behavior.

The two-bin system is the simplest. Decision variables include the reorder point, or small bin size, and the replenishment lot size. The main consideration is that the reorder point is high enough to cover demand during the replenishment cycle time, since a maximum of one order is allowed in the replenishment loop. If demand is seasonal, it is the maximum demand rate that is relevant. Under capacity constraints, increasing the lot size will increase the amount of additional product made available during a given replenishment cycle. However this will also increase the replenishment cycle time since the lot processing times will also increase. As well, increases in queue times may further increase cycle times, especially if machine utilization levels approach 100 percent. Therefore proper selection of both lot sizes and reorder points are critical to achieving desired performance.

Kanban systems, like two-bin systems, restrict the maximum inventory in the replenishment loop but in this case the number of orders outstanding is dictated by the number of Kanban cards. However, a key restriction is that the “reorder point” must be a multiple of the lot size since Kanban cards are only allowed to re-circulate when customer demand leads to the depletion of the lot associated with a card. This restriction can become very important when the number of Kanban cards used is low since performance changes become very granular. In other words, adding or deleting a card can cause large changes in performance. This can be alleviated if lot sizes or the frequency of delivery can also be changed.

The reorder point system is the most robust. The reorder point can be set at any discrete number of parts, like for the two-bin system, and an unlimited number of orders are allowed in the replenishment loop. In addition, the reorder point system keeps track of backorders and takes this information into account when placing orders. Therefore, there is no ceiling on inventory levels in the replenishment loop.

The merits of the Kanban system with respect to simplicity and transparency are well understood. It would also seem that under scenarios with machine breakdowns the ability to limit total replenishment inventory would benefit performance. However, the granular nature of performance might hinder performance relative to reorder point systems under some conditions. As well, under time-varying demand the ability of reorder point systems to use backorder information would appear to be useful. This study investigates these issues, using the two-bin system as a benchmark. Simulation experiments with all three replenishment systems are conducted using a common test bed and supply chain scenario.

The next section provides a brief literature review. This is followed by a detail discussion of the replenishment logic and experimental scenario used in this research. Finally, the experimental design, results and conclusions are presented.

2 Literature Background

Studies examining the relative performance of replenishment systems have usually relied on the use of discrete-event simulation. Krajewski et. al [1] concluded that there was not much difference between Kanban and reorder point performance. Other factors, like scrap rates, were found to have a greater impact on performance than the choice of replenishment system. Yang [2] concluded a Kanban system was superior. However, the Kanban logic was modified to essentially allow lot sizes of one and dispatching facilitated by setup time reduction. Since the lot sizing and dispatching policies were not consistently applied, it cannot be stated that the replenishment logic was responsible for the inferior performance of the reorder point system. Suwanruji and Enns [3] concluded that reorder point systems are generally superior unless demand patterns are level, in which case it is possible for a Kanban system to perform slightly better due to decreased lot interarrival time variability.

Evaluating replenishment systems raises the issue of performance measurement. One approach is to simply measure lot flowtimes and compare systems on this basis. From Little’s Law it is known that minimizing lot flowtimes also minimizes inventory. Under capacity-constrained lot sizing, attempts can be made to make such comparisons with optimal lot sizes based on queuing relationships. Lambrecht and Vandaele [4] dealt with this problem for a single product type. However, this research used a lot arrival process that was independent of any specific replenishment logic. Enns and Choi [5] investigated lot flowtime minimization in a material requirements planning (MRP) environment by adjusted the queuing relationships used in optimization for the effects of auto-correlation. Enns and Li [6] also addressed lot flowtime minimization but used a dynamic feedback mechanism for adjustment. None of these studies was designed to compare replenishment systems based on different types of logic and therefore delivery performance issues were ignored.

When comparing different types of replenishment systems, some measure of delivery performance, such as mean tardiness or proportion of deliveries from stock, should be taken into account. However, delivery performance is dependent on the level of inventory carried within the system. A tradeoff exists between delivery performance and inventory levels. Therefore the problem becomes one of dealing with two performance measures simultaneously. One approach is to set the performance level for one measure the same across all replenishment systems and then make comparisons on the basis of the other measure. For example, inventory levels could be set the same across all replenishment systems and comparisons made on the basis of delivery performance. The main challenge is then to determine decision variable settings, such as lot sizes, Kanban cards or reorder points, that will result in equal inventory levels. This generally requires extensive experimentation. As well, conclusions are limited to results obtained at one particular inventory level. Jacobs and Whybark [7] illustrate this approach in a study comparing MRP and reorder point systems.

Another approach is to develop tradeoff curves. This requires obtaining inventory and delivery performance results over a range of relevant values. If the curve for one replenishment system dominates another, it can be concluded that this replenishment system is superior. An advantage of this approach is that conclusions are based on a range of inventory or delivery service levels, not one particular point. Whybark and Williams [8] used this approach in an early study on safety leadtimes and safety stocks in MRP systems.

However, even with the use of tradeoff curves there is the problem of which decision variables to change in generating the tradeoff curves and what settings to use for the decision variables that remain fixed. For example, in Kanban systems a particular lot size may be selected and then the number of Kanban cards varied to generate a tradeoff curve. Reorder point experiments could then be run by using the same lot size and varying the reorder point. This approach was used by Suwanruji and Enns [3].

This research takes the approach of using tradeoff curves further by also seeking to select optimal decision variables prior to making comparisons. In other words, systems are compared when each is being operated under near optimal conditions.

3 Experimental Scenario

A basic illustration of the scenario used in this research is shown in Figure 1. There are two part types that come from suppliers and are then processed on the same capacity-constrained machine. The completed lots of processed parts become finished goods that are consumed by individual customers taking single items. The customer demand for each part type is Poisson but the demand rate is adjusted every time unit according to a seasonal demand pattern. The expected demand for both part types follows a sinusoidal pattern. The mean demand is 40 items per time unit and the demand pattern has an amplitude of 4. The cycle length of the demand pattern is 250 time units. However, the demand patterns for the two part types are offset by 125 time units. Therefore the aggregate workload requirement at the machine is fairly constant through time. The actual demand rate during each time unit is based on sampling from a Normal distribution centered around the expected demand rate indicated by the sinusoidal demand curves. The standard deviation of this distribution is specified to be 0.1 times the expected demand rate. Figure 2 illustrates an example of the demand patterns for the two part types, *P1* and *P2*, through one demand cycle. In this diagram a time unit is assumed equal to one day.

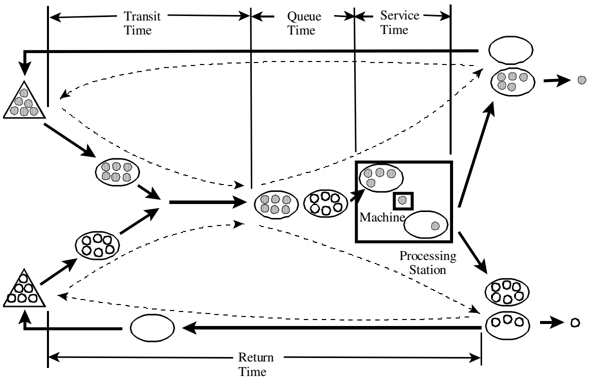

Figure 1: Configuration of replenishment scenario

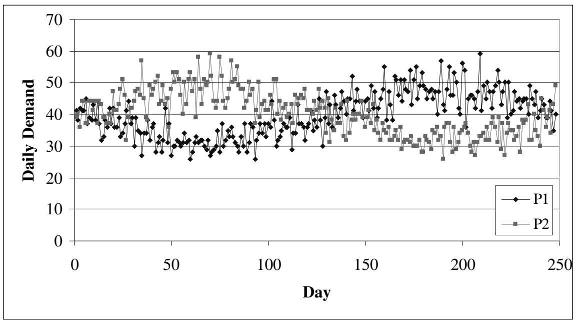

Figure 2: Demand pattern

Replenishment is controlled by a two-bin, (*2Bn*), single-card Kanban (*Kbn*) or continuous-review reorder point (*ROP*) system. Performance is measured in terms of average total inventory counts, *TI*, and the proportion customer demand filled from stock, *SL*. If the desired part type is not in stock, a customer backorder is placed and then filled as soon as finished goods inventory is replenished.

The logic for placing a finished goods replenishment order for the two-bin system is given by Equation 1. The order quantity, $Q_{i,t}$, for part type i at time t may be equal to a multiple of the replenishment lot size, LS_i.

$$Q_{i,t} = \max\left(0, \left\lceil \frac{OP_i - (FG_{i,t} + OR_{i,t} + OT_{i,t} + OQ_{i,t}) + 1}{LS_i} \right\rceil LS_i \right) \quad (1)$$

where:

OP_i - Order point for part type i

$FG_{i,t}$ - Qty. of part type *i* finished goods in stock
$OR_{i,t}$ - Qty. of part type *i* orders released to supplier but not yet filled
$OT_{i,t}$ - Qty. of part type *i* in transit
$OQ_{i,t}$ - Qty. of part type *i* in queue or on machine

The logic for placing a replenishment order for the single-card Kanban system (*Kbn*) is given by Equation 2. This equation assumes the Kanban card is released upstream for recirculation when a lot-size container is completely depleted.

$$Q_{i,t} = \max\left(0, \left\lceil \frac{(KC_i - 1)Q_i - (FG_{i,t} + OR_{i,t} + OT_{i,t} + OQ_{i,t}) + 1}{LS_i} \right\rceil LS_i \right) \quad (2)$$

where:
KC_i - Number of Kanban cards for part type *i*

The implementation for reorder point replenishment is different in that customer backorders are considered in the replenishment decision. The logic for the continuous-review reorder point (*ROP*) system is given by Equation 3.

$$Q_{i,t} = \max\left(0, \left\lceil \frac{OP_i - (FG_{i,t} + OR_{i,t} + OT_{i,t} + OQ_{i,t} - BO_{i,t}) + 1}{LS_i} \right\rceil LS_i \right) \quad (3)$$

where:
$BO_{i,t}$ - Quantity of part type *i* backordered

It is assumed that the lot size, Kanban card and order point decision variables remain constant through time despite the seasonality of demand. When a replenishment order is triggered, it goes to the supplier instantaneously for all three systems. For the Kanban system this means orders are transmitted electronically as opposed to having the cards travel with a transporter.

The supplier is always assumed to have inventory available to fill the order. However, shipping must wait until a transporter is available to pick up the order and move it to the capacity-constrained machine. There are 24 transporters in continuous circulation in each of the replenishment loops. The circuit for these is shown as dashed lines in Figure 1. It is assumed these transporters circulate even if there are no orders to convey for the given part types. The expected travel time to the workstation is described by a Gamma distribution with a mean of 0.5 and a standard deviation of 0.05 time units. The travel time back to the supplier follows the same distribution. Therefore, the expected cycle time for each transporter is 1 time unit.

The lots arriving at the capacity-constrained machine join a queue if the machine is busy. Lots are processed in first-come-first-serve (FCFS) order, with each lot incurring a setup. The two part types are assumed to be identical in terms of processing and replenishment time requirements. However a unique setup is required for each part type and they are not interchangeable with respect to the supply source or customer demand. This assumption simplifies the number of decision variables that must be dealt with the since lot sizes and the number of Kanban cards or reorder points for both part types are assumed to be equal.

The lot setup time for both part types is 0.23 time units. Each part in the lot requires a processing time of 0.004 time units. There is no uncertainty associated with these times except for machine breakdowns. Breakdown times follow a negative exponential distribution with a mean equal to 0.1 times the lot processing time. When a lot has been completed, it is immediately made available as finished goods to meet customer demand.

4 Experimental Design

The main experimental design factor was the replenishment system, run at three levels. These levels were the two-bin, the Kanban, and reorder point systems. Additional factors, such as lot sizes, reorder points and Kanban cards, were used as appropriate to create performance tradeoff curves.

Two-bin experiments were performed using various combinations of lot sizes and reorder points. Performance was found to be very sensitive to the lot sizes selected, especially if these were too small. The optimal lot size was found to be within the range of 50 to 70. For each lot size, various levels of the reorder point were run so that performance tradeoff curves could be generated.

Kanban experiments using three or more Kanban cards produced very granular results. In other words, it was not possible to produce tradeoff curves exhibiting control of delivery performance. Therefore, the main experimental analysis was performed on the basis of using two Kanban cards and then varying lot sizes to generate the tradeoff curves. Lot sizes of between 30 and 60 were found suitable. Changing lot sizes to control delivery performance may not be desirable in practise. However, the alternative of changing delivery frequency was not judged to be feasible in this research due to the complexity involved when adding another decision variable.

Reorder point experiments were performed in a similar manner to those for the two-bin system, with separate tradeoff curves generated for various lot sizes. The optimal lot size was found to be within the range of 30 to 50.

All experiments were performed using Arena® 5.0 discrete-event simulation software [9]. Each simulation run was 2750 time units in length, with the first 250 time units used for initialisation. Each tradeoff curve was replicated three times. Common random numbers were used as a variance reduction technique.

5 Experimental Results

The experimental results are illustrated by averaging the three replications made for each tradeoff curve. Results for the two-bin, Kanban and reorder point systems are first presented individually to help identify decision variable settings resulting in the best tradeoff curves. The tradeoff curves associated with near optimal performance for each system are then compared.

Figure 3 shows the results for the two-bin system. Lines along each curve represent increasing reorder points. In this case a lot size, *LS*, of 60 appears to be approximately optimal. This lot size results in a higher delivery service level, *SL*, for a given level of inventory, *TI*, than a lot size of 70. The machine utilization level at this setting was approximately 66%.

As illustrated in Figure 4, a lot size of 50 is too small for the two-bin system since it will not allow a 100% service level to be achieved regardless of the reorder point, *OP*. The reason high service levels cannot be achieved is that peak demand during the replenishment cycle time exceeds the lot size.

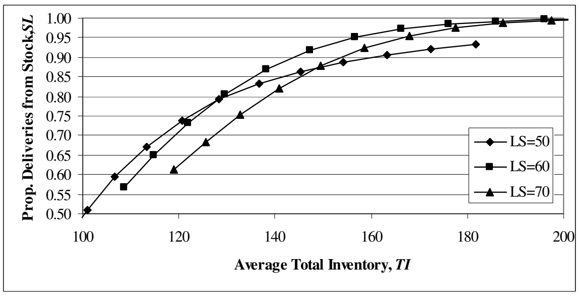

Figure 3: Two Bin performance results

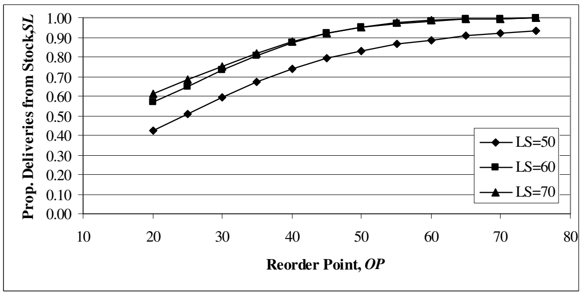

Figure 4: Two Bin service level versus reorder points

Figure 5 shows the results for the Kanban system. Delivery service levels in this case increase with increasing lot sizes. The use of two Kanban cards (*KC*=2) was observed to be best since this allowed control of delivery service levels, *SL*, over the range of interest. Machine utilization levels with *KC*=2 varied from over 90% with small lot sizes, resulting in delivery service levels around 50%, to under 65% with service levels approaching 100%.

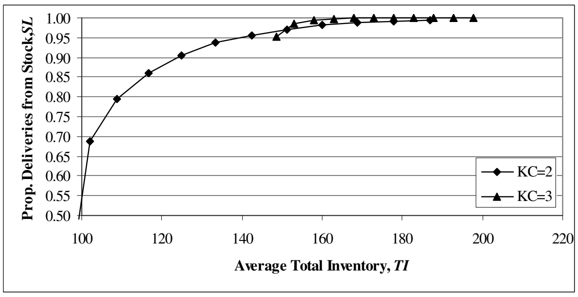

Figure 5: Kanban performance results

As illustrated in Figure 5, the use of three Kanban cards (*KC*=3) could also be used to achieve very high service levels. However, lower service levels could not be achieved without loss of system stability, as illustrated in Figure 6. At lot sizes of around 30 the utilization level at the capacity-constrained machine approached 100%. In other words, too many setups were incurred with lot sizes below 30 to prevent inventory building up in queue behind the machine. Therefore, there was no benefit to using more than two Kanban cards in this experimental scenario.

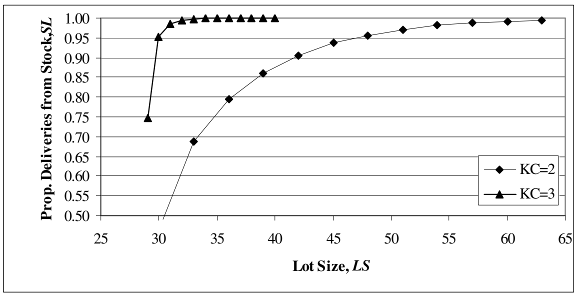

Figure 6: Kanban service levels versus lot sizes

Figure 7 shows the results for the reorder point system. Each tradeoff curve is for a specific lot size, *LS*. Delivery service levels increase with increasing reorder points. The optimal lot size in this case was found to be in the range of 40, since this tradeoff curve dominates the rest. Utilization levels with this lot size were around 81%.

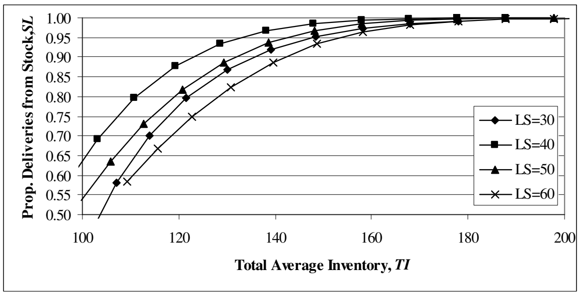

Figure 7: ROP performance results

It should be noted that with reorder point systems it is possible to achieve 100% service levels even lot sizes are smaller than optimal. Since the maximum inventory level is not controlled, increasing the reorder point simply allows more lot-size orders to be released into the replenishment loop. With the two-bin or Kanban systems the minimum feasible lot size is restricted by the ability of meet demand during the replenishment cycle given a maximum number of orders outstanding.

Finally, Figure 8 shows a comparison of the best results achieved with each of the three replenishment systems. As expected, the two-bin system results in inferior performance. The differences between the Kanban and reorder point systems are very small but it does appear that the reorder point system performs slightly better at high service levels for this scenario.

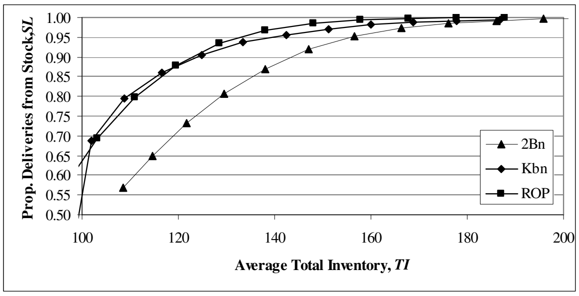

Figure 8: Comparison of performance results

Previous results, such as Suwanruji and Enns [3], have demonstrated that continuous-review reorder point systems tend to perform better than single-card Kanban systems under time-varying demand due to the use of backorder information. However, previous research has generally not assumed that there are machine breakdowns. In the case of breakdowns, it is often claimed that prevention of inventory pileups ahead of machines is a positive feature of Kanban systems. This may true but the advantages in this study would appear to be quite small. The fact that the reorder point and Kanban systems perform nearly the same may be due to their individual strengths offsetting each other. For example, the reorder point system advantage of using backorder information may be offset by the Kanban system advantage of restricting inventory buildups during breakdowns.

6 Conclusions

This research illustrates a methodology for comparing continuous-review replenishment systems under conditions where each is being operated near optimally. Performance comparisons consider the tradeoff between inventory levels and delivery performance. As well, the tradeoff curves generated allow comparisons to be made over a range of service levels so that it can be determined if one replenishment system generally dominates.

The tradeoff results show that, as expected, the two-bin system performs the poorest. Large lot sizes are required and the machine utilization is relatively low. The differences between the Kanban and reorder point systems are small and not of practical consequence. In the scenario evaluated, the advantage of the Kanban system in preventing inventory buildup during machine breakdowns may be offset by the advantage of the reorder point system in considering backorder information. Further investigation is necessary to demonstrate and confirm this behavior.

7 References

[1] Krajewski L.J., King B.E., Ritzman L.P., Wong D.S., 1987, Kanban, MRP and shaping the manufacturing environment. Management Science, 33, 39-57.

[2] Yang K.K., 1998, A comparison of reorder point and kanban policies for a single machine production system. Production Planning and Control, 9, 385-390.

[3] Suwanruji P., Enns S.T., 2006, Evaluating the effects of capacity constraints and demand patterns on supply chain replenishment strategies. International Journal of Production Research, 44, 4607-4629.

[4] Lambrecht M.R., Vandaele N.J., 1996, A general approximation for the single product lot sizing model with queueing delays, European Journal of Operational Research, 95, 73-88.

[5] Enns S.T., Choi S., 2002, Use of GI/G/1 queuing approximations to set tactical parameters for the simulation of MRP systems, Proceedings of the Winter Simulation Conference, Yücesan E., Chen C-H., Snowdon J.L. and Charnes J.M. (Editors), 1122-1129.

[6] Enns S.T., Li L., 2004, Optimal lot-sizing with capacity constraints and auto-correlated interarrival times, Proceedings of the Winter Simulation Conference, Ingalls R.G., Rossetti M.D., Smith J.S., and Peters B.A., (Editors), 1073-1078.

[7] Jacobs F.R., Whybark D.C., 1992, A comparison of reorder point and material requirements planning inventory control logic. Decision Sciences, 23, 332-342.

[8] Whybark D.C., Williams J.G., 1976, Material requirements planning under uncertainty. Decision Sciences, 7, 595-606.

[9] Kelton W.D., Sadowski R.P., Sturrock D.T., 2004, Simulation with Arena, 3rd Ed., McGraw-Hill, New York, NY.