UNIVERSITÉ DE MONCTON
CAMPUS DE MONCTON

Faculté d'administration

CUWIP: A modified CONWIP approach to controlling WIP

Jules Comeau, professor at Université de Moncton, NB
Uday Venkatadri, professor at Dalhousie University, Halifax, N.S.

Cahier électronique de la Faculté d'administration
Faculté - © cahier 09-2014-Canada

CUWIP: A MODIFIED CONWIP APPROACH TO CONTROLLING WIP

Jules Comeau, Université de Moncton, Moncton, N.B.
Uday Venkatadri, Dalhousie University, Halifax, N.S.

Abstract: *CONWIP (Constant Work In Process) is a production control method that puts an upper bound on the number of lots in a production line or individual work center in order to control the level of WIP. In a make to order production environment with a first come, first serve sequencing rule, if customer demand varies greatly in terms of order size and product mix, the observed WIP level can fluctuate correspondingly. In the proposed method, CUWIP, all customer orders are defined using Units of Equivalence (UE). The proposed approach caps the number of UE's for the entire production line. The method is tested via an AweSim (SLAM based) simulation using actual data from a window plant production line. Results show that production throughput is maintained (as compared to a push system) while reducing average WIP levels and average cycle times. CUWIP also appears to give lower coefficients of variation for cycle times and WIP levels compared to those obtained with CONWIP. Hence, CUWIP seems to retain the advantages of CONWIP, while being suited for a greater variety of environments.*

Keywords: *CONWIP, Pull production systems, variable demand, Little's law*

1. Introduction

Production control systems can be categorized into two major types: push and pull (Spearman et al., 1990). Push systems schedule releases, while pull systems authorize them. As a result, push systems control throughput and observe work in process (WIP), while pull systems control WIP and observe throughput (Hopp and Roof 1998). Push systems are very popular in the form of Materials Requirements Planning (MRP). The best known pull system is commonly known as Kanban. For a complete description of pull systems, see Hopp and Spearman (2001).

Kanban works by authorizing production rather than by scheduling it. Spearman and Zazanis (1992) point out that pull systems control WIP and then measure throughput against required demand. Adjusting WIP levels on the production line will adjust throughput. There are significant advantages to this approach. There is less congestion in pull systems and WIP is easier to control than throughput because it can be observed directly.

Spearman et al. (1990) point out that Kanban is difficult, or impossible to use when there are job orders with short production runs, or significant set-ups, or scrap loss, or large unpredictable fluctuations in demand. This is an important observation as Takahashi and Nakamura (2002) point out: "product life cycles have become shorter and shorter due to diversification of customer needs, and the duration of stationary demand has also shortened". Furthermore, Bonvik et al. (1997) explain that Kanban is only suitable for high volume production environments with relatively few part types.

2. CONWIP

Push and pull production systems can work together. CONWIP, CONstant Work In Process, in essence, is an example of such a combination. Work is scheduled and put on the backlog list but cannot be started without a Kanban authorization (Spearman et al. 1990).

The primary difference between CONWIP and Kanban systems is that CONWIP pulls jobs into the front of the line and pushes them between stations elsewhere in the line, while Kanban pulls jobs between all stations (Hopp and Roof. 1998). This ensures a constant amount of WIP in the entire production line. Contrary to Kanban, a CONWIP system does not keep track of card distribution in the system but only of the total count (Gstettner and Kuhn (1996), Roderick et al. (1994)). In a CONWIP system, downstream work centers pull stock from previous operations as needed (Spearman and Zazanis (1992) and Ryan and Vorasayan (2005)). All operations then perform work to replenish outgoing stock. Ryan and Choobineh (2003) state that it can also be implemented on a large set of work stations within a production line where jobs are pulled into the set of stations and pushed between them. CONWIP appears to share the benefits of Kanban while being applicable to a wider variety of production environments.

CONWIP assumes that parts are moved in standard containers or lots, each of which contains roughly the same amount of "work" (Hopp and Spearman, 2001). Such isn't always the case. When implementing CONWIP where lot sizes are not identical but vary little, tight control of WIP levels can be accomplished by using techniques such as Statistical Throughput Control (STC) credited to Hopp and Roof (1998). Leu and Chang (2001) show the need, in certain circumstances, to vary the size and number of lots to maintain acceptable finished goods inventory and customer service levels. WIP level adjustment techniques, such as STC and lot size management, all strive to accomplish one thing: adjust the allowable card count (upper and lower limits on number of lots on the production line) to keep constant WIP levels. For example, if lot sizes become smaller (fewer products in each order), the total WIP on the production line will drop for the same card count. In this case, the maximum card count should be raised to reestablish the WIP level to a predefined level (between minimum and maximum targets). Methods like STC and lot size management have been shown to work well in cases where demand has a low rate of change (seasonal variations for example) in both product mix and lot sizes. When these two demand variables change rapidly, a different approach must be taken.

Several studies have developed methodologies that facilitate the implementation of CONWIP in production environments that have the characteristics stated above. Researchers have developed sophisticated methods for setting and controlling card counts. See Takahashi and Nakamura (2002), Framinan et al. (2003) and Ryan and Vorasayan (2005), for example. Ryan and Varasayan (2005) discuss the use of dedicated Kanbans that can be assigned to a given product type. In this scenario, the number of cards is controlled and optimized for each family type. There may or may not be an upper and lower limit on each card count. The sum of card counts for all product types is then capped in the traditional CONWIP style. This approach is used when routings are different for distinct product families (Hopp and Roof (1998), and Ryan and Choobineh (2003)).

Although studies in the literature present very sound and logical adjustments to the traditional CONWIP ideas of Hopp and Spearman, many manufacturing facilities aren't conducive to the type of micro-management necessary to ensure successful implementation. There is clearly a need for an easily manageable production control system that retains the obvious advantages of pull systems. Therefore, the study presented in this paper had the following objectives: (i) develop a method of controlling WIP that can be implemented in a make to order production facility and that can effectively manage fluctuations in volume and product mix and (ii) compare it to existing production control systems.

3. CUWIP

In the cases where lot sizes and product mix vary greatly, Hopp and Spearman (2001) suggest defining the cap in terms of the capacity of production required for all lots in production. They argue that when products have similar processing times, assigning a card count that represents the total processing time for all products in a lot is equivalent to physically counting the products in the lot. This observation is valid if all the products that compose the lots are similar.The proposed approach in this research is to use the idea put forward by Hopp and Spearman but to define the cap as the maximum number of total hours of production on the entire production line. The manner of assigning the cards to each order is the major contribution of this work.

This study proposes a method that will transform all lots into time blocks that are called units of equivalence (UE). Constant Units of Work In Process (CUWIP) is a method by which each order is assigned a card count based on its proposed processing time. For example, an order that has twice the total processing time will have a UE count two times higher. The duration in length for a UE can be set to suit the need of the application.

The cap control is specified on the total count of UE rather than on the number of lots in circulation. When a product finishes production, it will authorize the release of a number of UEs equal to the number assigned to it before production began. Therefore, it is expected that the WIP level in a CUWIP implementation will not vary as much as a CONWIP implementation in the same variable demand environment.

In a similar fashion to Ryan and Choobineh (2003), no attempt is made to improve the system performance by optimizing the backlog list.

4. Case study

The ideas explained in Section 3 have been tested using data from a real application. Framinan et al. (2003) suggest that many of the methodologies discussed earlier should be tested in real applications; so far, few studies of this nature have been published.

4.1 Plant and demand

The company studied produces 31 different products that can be combined in any quantity to form a customer order. Each product has a different routing through the production line and therefore a different cycle time.

To compare the push method in place on the production line at the time of the study with the CONWIP and CUWIP methods, 148 customer orders broken down into 2408 UE were used. The UE are distributed as indicated in figure 1. There was no attempt to optimize the backlog list which uses a FIFO (first in, first out) discipline which creates a wildly fluctuating product mix for the production floor.

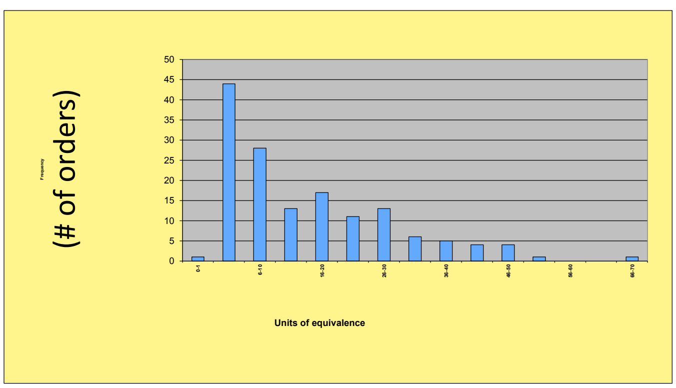


Figure 1 – Distribution of UE in customer orders

4.2 Product characteristics

The 31 products produced in the plant are all combinations of 5 different vinyl window components plus the hardware that operates the windows. Figure 2 shows a product (SH – X) broken down into two different vinyl components (BX200 and BX222) which will have hardware added to make the final product.

Each of those vinyl components will be considered one Unit of Equivalence because they each take approximately the same amount of production time to complete.

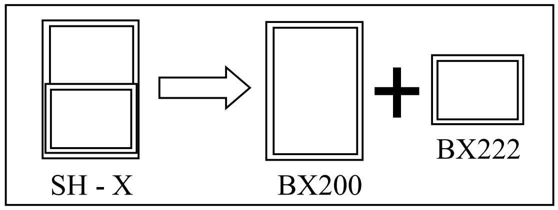


Figure 2 – A product shown being broken down into its basic components

Vinyl windows have dimensions varying anywhere from 25cm x 25cm to enormous steel reinforced units of 4m x 4m or more. They are built in increments of 1mm along both dimensions.

Because each vinyl window component requires the same amount of production time, each customer order will be assigned a card count equivalent to the number of vinyl components that make up its products.

4.3 Simulation of a production line

The plant in question has 13 work stations with two stations being operated by the same person. All products must go through the welding machine which is the bottleneck for the entire line. Figure 3 shows that components of a product can take different routes through the production line depending on the attributes assigned to them via a C++ user insert in a SLAM II / Awesim simulation.

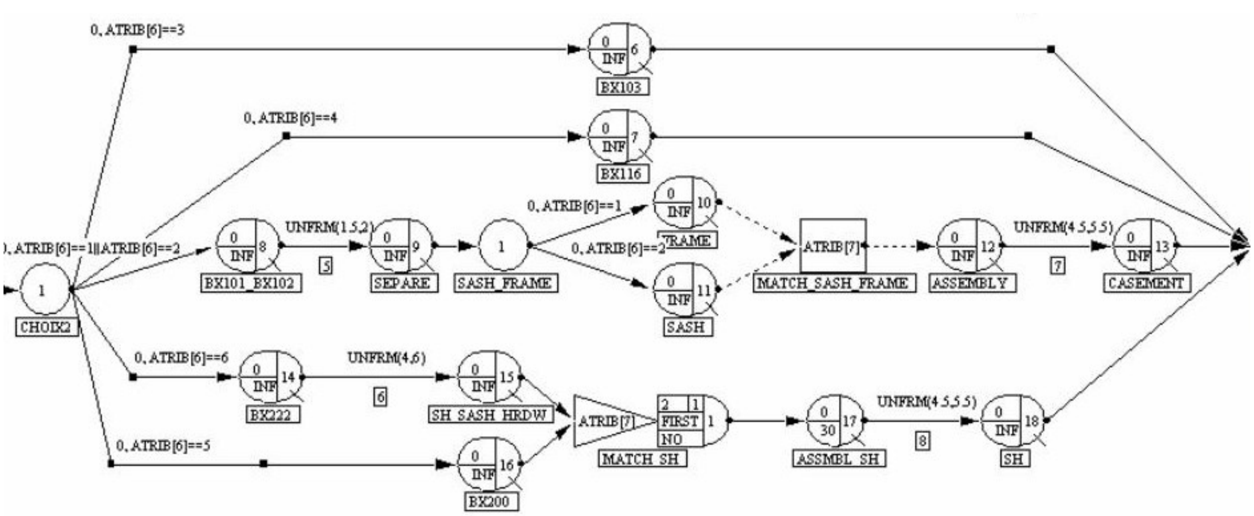


Figure 3 – Partial representation of the SLAM II model

Some important assumptions have to be made. There are always 12 employees available for the period being studied. There are no machine breakdowns during this same period. The authorization for the start of a new lot is sent instantly from the end of the production line to the first station in the line and transportation time between stations is negligible. We assume unlimited raw material and customer demand so that the first station never starves; the finished goods inventory is considered to have an infinite capacity.

The operation times at each station are assumed to be uniformly distributed and don't vary much from one station to the next. Partial time study results, historical production data and routing cards for all products formed the basis for three SLAM based simulation models. Figure 3 gives a partial representation of one of the three simulation models used. The first model represents the push system with no control on WIP. In this scenario, each employee in the production line produces as much as possible without considering what is happening upstream or downstream on the production line. The second is an implementation of CONWIP in the same plant with caps of 3, 4 and 5 lots. For CONWIP, this means that when all products in a given order have left the production line, the authorization is given for a new order to enter the production line. The third model is an implementation of CUWIP in the same plant which was observed with caps (maximum WIP levels) of 10 to 70 UE in increments of 10. In this scenario, authorization for the production of the next product on the backlog list is delayed until sufficient UE are available. UE are released at the end of the production line when products have completed processing.

5. Methodology

All models were run for a period representing three weeks of production during the peak summer production season. Statistics were collected for the average cycle time and the average level of WIP (UE) in the plant as recommended by Law and Kelton (1991). The average cycle time for an order is defined as the average, for all orders, of the total processing time from the start of work on the first product of the lot to the end of the production time on the last product of the lot. It includes delays, waiting times, start up times and transportation times between stations. The average WIP is defined as the average, for a simulation period, of the sum of the material being processed plus the material waiting between workstations for the entire line.

Average throughput is calculated based on Little's law:

$$TH = \frac{WIP}{CT}$$

where: TH = Average throughput
WIP = Work in process
CT = Average cycle time

The first step was to collect data on customer demand. This information came in the form of daily reports of sales. Window orders were manually extracted from these reports and entered into Excel where they were divided into the 31 window types. This Excel file contained the order number, the date the order was received, the type of windows in the order and the number of each of these types. This file containing all the orders was transformed into a comma-delimited text file to be transformed into production lots using a C++ routine.

At the beginning of any simulation run, the Visual SLAM processor calls function INTLC to allow the user to set initial conditions and to schedule initial events. The INTLC function built for these simulation models took all of the customer orders and transformed them into UE by creating entities and assigning attribute values before entering them into the network. Each entity represents one of the vinyl components used in constructing the final products.

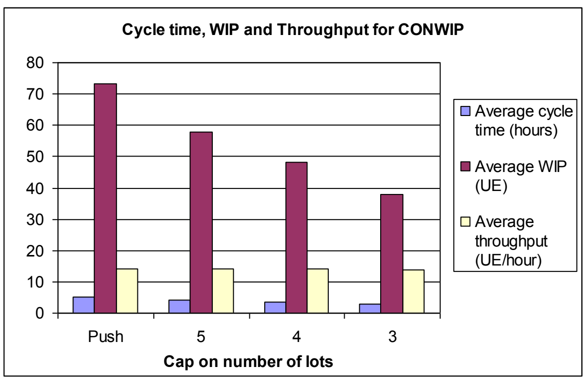
To evaluate the effectiveness of each method (push, CONWIP and CUWIP), we compared model parameters (the average WIP, average CT and average TH) for each method. A very important measure of effectiveness of these methods is the standard deviation of WIP, CT and TH. Recall that we are trying to minimize the variation of WIP to make plant management more efficient. The coefficient of variation can be used to calculate the variation around the mean for all of the above observed model parameters. The standard deviation is divided by its mean to give a coefficient that reflects that amount of variation around the mean. In this case, a small coefficient of variation indicates a parameter that is stable with the changing demand.

6. Results

The simulation study was first used to determine the average WIP level and average cycle time in the plant (with the push method). The production of 2408 UE, with the push method, took 10100 minutes which results in a throughput of 14.30 UE/hour. In comparison, the actual throughput rate for the orders used in the simulation was 14.36 UE/hour thus confirming the validity of the model. The simulation model for this same period had an average WIP total of 73 UE. Since the objective in a pull system is to reduce the amount of WIP, the other two methods (CONWIP and CUWIP) would not be tested with a WIP level above 70 UE. The results are presented in Figure 4. The CT is reduced as the WIP level goes down which results in a constant TH for all methods studied (push, CONWIP and CUWIP). The lead-time quoted to the customer depends on this throughput rate to remain unchanged.

Since the quoted lead-time is approximately three weeks in the window industry (for custom sized windows), the observed reduction of 2-3 hours in the CT between the push system and CONWIP does not have a significant impact on customer service. On the other hand, in an industry where the lead-time is measured in hours, such a reduction could mean considerable savings.

It may be noted that the plant was simulated with 12 employees and, since each vinyl component is assigned one UE, if the WIP level is set at or below 12 (in the CUWIP model), there is a drastic reduction in the TH. This is a result of not having enough products on the production line to keep all employees busy.



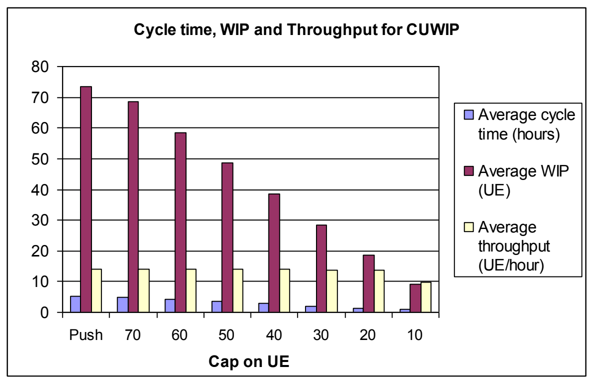


Figure 4 – Cycle time, WIP and throughput for CONWIP and CUWIP

Figure 4 shows, for push, CONWIP and CUWIP methods, the evolution of average cycle time when we put a cap on the number of lots (CONWIP) or the number of UE (CUWIP). The cycle time goes down proportionately to the reduction in average WIP. The caps on the bottom of the figures show the number of orders (CONWIP) or UE (CUWIP) permitted on the production line at any given time. Figure 4 also shows that the throughput remains unchanged as the cap is lowered.

Table 1 shows the average coefficients of variation (CV) or relative standard deviation for cycle time duration and WIP levels for all three methods (push, CONWIP and CUWIP). CV is the ratio between simulation standard deviation and simulation average. The results in table 1 are stated for 5 orders with CONWIP and 50 UEs for CUWIP. We see, in Figure 4, reductions in cycle time for both CONWIP and CUWIP when a cap is applied. Table 1, however, shows that CUWIP gives a much lower CV, nearly equivalent to the one obtained for a push system.

Method	Push	CONWIP	CUWIP
Cycle time	0.01	0.18	0.02
Average WIP	0.01	0.22	0.02

Table 1 : Average coefficients of variation for the three methods

7. Conclusion

Based on a case study, CUWIP appears to provide a better control of WIP as compared to push and CONWIP system in a variable demand environment. From a practical point of view, the approach presented in this study allows production plants faced with variable demand to implement a production control method tailored to their needs. CUWIP gives the WIP control advantages of CONWIP without the relatively high variability typical of CONWIP in a highly variable make to order production environment with FIFO order sequencing rules. This study isn't as much about presenting new innovative methods of controlling WIP as it is to reiterate the importance to think critically about the impact of high WIP in this type of production environment and to implement the easiest methods of controlling this WIP. CUWIP will result in less clutter on the shop floor making it much easier to react to quality problems such as machine failures and defects. When WIP levels are low, operators waste less time searching through WIP for the next job to process.

Further work in this area should concentrate on finding practical, easy to implement, methods of dynamically adjusting the cap as a reaction to rapidly changing demand.

References

Bonvik, A. M., Couch, C. E., Gershwin, S. B. (1997). A Comparison of production-line control mechanisms. International Journal of Production Research, 35(3), 789-804.

Framinan, J. M., González, P. L., Ruiz-Usano, R. (2003). The CONWIP production control system: review and research issues. *Production planning and control*, 14(3), 255-265.

Gstettner, S., Kuhn H. (1996). Analysis of production control systems Kanban and CONWIP. *International Journal of Production Research*, 34(11), 3253-3273.

Hopp, W. J., Spearman, M. L. (2001). Factory Physics: Foundations of manufacturing management. Chicago: Times Mirror Company.

Hopp, W. J., Roof, M. L. (1998). Setting WIP levels with statistical throughput control (STC) in CONWIP production lines. *International Journal of Production Research*, 36(4), 867-882.

Law, A. M., Kelton, W. D. (1991). Simulation modeling and analysis. New York: McGraw Hill.

Leu, B.-Y., Chang, S.-H. (2001). The effect of container size on a CONWIP production line: a simulation study. *International journal of industrial engineering*, 8(2), 80-90.

Roderick, L. M., Toland, J., Rodriguez, F. P. (1994). A simulation study of CONWIP versus MRP at Westinghouse. *Computers and industrial engineering*, 26(2), 237-242.

Ryan, S. M., Choobineh, F. F. (2003). Total WIP and WIP mix for a CONWIP controlled job shop. *IIE Transactions*, 35, 405-418.

Ryan, S. M., Vorasayan, J. (2005). Allocating work in process in a multiple-product CONWIP system with lost sales. *International Journal of Production Research*, 43(2), 223-246

Spearman, M. L., Woodruff, D. L., Hopp, W. J. (1990). CONWIP: a pull alternative to Kanban. *International Journal of Production Research*, 28(5), 879-894.

Spearman, M. L, Zazanis, M. A. (1992). Push and pull production systems: issues and comparisons. *Operations Research*, 40(3), 521-532.

Takahashi, K., Nakamura, N. (2002). Comparing reactive Kanban and reactive CONWIP. *Production planning and control,* 13(8), 702-714.