APPLYING THEORY OF CONSTRAINTS AS A CONTINUOUS IMPROVEMENT TOOL IN A LEAN ENVIRONMENT

APPROVED BY SUPERVISING COMMITTEE:

____
Hung-Da Wan, Ph.D., Chair

____
Krystel K. Castillo, Ph.D.

____
Fengshan Frank Chen, Ph.D.

Accepted: ______________________________________
Dean, Graduate School

DEDICATION

To my family

APPLYING THEORY OF CONSTRAINTS AS A CONTINUOUS IMPROVEMENT TOOL IN A LEAN ENVIRONMENT

by

TRUMONE R. SIMS, B.S.

THESIS
Presented to the Graduate Faculty of
The University of Texas at San Antonio
in Partial Fulfillment
of the Requirements
for the Degree of

MASTER OF SCIENCE IN ADVANCED MANUFACTURING
AND ENTERPRISE ENGINEERING

THE UNIVERSITY OF TEXAS AT SAN ANTONIO
College of Engineering
Department of Mechanical Engineering
May 2014

ACKNOWLEDGEMENTS

I will bless the Lord at all times: his praise shall continually be in my mouth.

Psalm 34:1

I would like to thank my advisor, Dr. Hung-Da Wan for all of the help he has given me throughout my graduate studies. He encouraged me to write my first two academic papers which will always link us.

I would like to thank the anonymous referees for their valuable comments during the writing of my first two academic papers. In addition to offering suggestions that would improve the structure and clarity of my work, they also recommended additional articles related to my subject.

I would also like to thank my committee members, Professor Krystal Castillo and Professor F. Frank Chen.

A special thanks goes to the employees who work at the manufacturing facility that my simulation mode is based upon. I appreciate all the support they gave me.

I would like to thank Texas Power Systems and The University of Texas at San Antonio for their financial support.

May 2014

APPLYING THEORY OF CONSTRAINTS AS A CONTINUOUS IMPROVEMENT TOOL IN A LEAN ENVIRONMENT

Trumone R. Sims, M.S.
The University of Texas at San Antonio, 2014

Supervising Professor: Hung-Da Wan, Ph.D.

Theory of Constraints (TOC), with its five steps for constraint elimination, is viewed as a continuous improvement process. The first step in the TOC methodology is to identify the constraint. In a lean environment, this step is not always trivial. This thesis proposes three new methods for this purpose. The first method, Flow Constraint Analysis, takes a holistic view and evaluates whether the customer's demand is being satisfied. This evaluation is made by comparing the takt times and the cycle times of resources in the manufacturing system in order to identify the constraint(s). The second method, Effective Utilization Analysis, can be employed to pinpoint the location of the system constraint to a specific process or station. The actual production throughput is compared against the ideal capacity of the system to locate the bottleneck. This method is based on the relationship between, work in process (WIP), bottleneck rate and lead time for a constant work in process (CONWIP) system. The third method, Quick Effective Utilization Analysis, can be used when there is little or no historical line performance data available.

The second step in TOC is to decide how to exploit the constraint. A non-traditional option for exploiting the system constraint will also be explored in this thesis. This research attempts to perform the exploitation by getting the most from the constraining resource without additional investment.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE: INTRODUCTION

The manufacturing system output is a function of the whole system, not just individual processes. When we view our system as a whole, we realize that the system output is a function of the weakest link. The weakest link of the manufacturing system is the constraint. Consequently, there needs to be focus on the coordination of efforts to optimize the overall system, not just individual processes (Breyfogle, 2003). When a system matures in lean implementation, the production flows smoother and the main constraint becomes less obvious. However, the impact of performance of constraining resources in a lean system, especially one with moving assembly lines, is still evident. Because "every value stream has a primary bottleneck (constraint) that limits its ability to reach its goal" (Bell, 2006, p. 175), it is even more critical to be able to identify system constraints in a lean environment.

Theory of Constraints (TOC) is a well-known methodology for systems improvement that includes principles and practice guidelines that can be adopted by practitioners (Watson, Blackstone, & Gardiner, 2007). The famous novel for operations management, *The Goal* (Goldratt & Cox, 1984), written by Eli Goldratt caught the attention of process improvement professionals and began the use of this methodology. From this book, the five focusing steps (5FS) were brought out: 1) Identify the System Constraint, 2) Decide How to Exploit the Constraint, 3) Subordinate Everything Else, 4) Elevate the Constraint, and 5) Go Back to Step 1, but beware of "Inertia".

On a different path, lean manufacturing is based on Toyota Production System (TPS) and Just-in-Time (JIT) concepts. The lean principles aim at eliminating waste to the maximum extent in order to improve the flow of the value stream (Wan & Chen, 2008).

The concept of integrating Lean, Six Sigma and the Theory of Constraints is being explored more and more while simultaneously being applied to various industries (Pirasteh & Kannappan, 2013). The integration of Lean and TOC will be the focus of this thesis. Constraint identification at a lean manufacturing plant using TOC will be the method of integration.

Motivation

The manufacturing plant where the research for this thesis is being performed consists of Constant Work-In Process (CONWIP) moving assembly lines and several individual automatic stations. The assembly operations are a one-piece flow production environment.When it comes to constraint identification, a lot of focus has been given to the equipment's actual vs. target cycle time. Therefore, a constraint is being defined as a piece of equipment in which the actual cycle time is lower than the target cycle time. This type of information is more applicable from a capacity planning standpoint instead of being used as a continuous improvement tool. Spreadsheets and tables have been created to capture this data. This method of defining and identifying the system constraint would be acceptable if the system was totally automated. But that is not the case; in fact for assembly processes 70% of the operations require manpower. Another weakness in this method of constraint analysis is the exclusion of downtime data. The company's method assumes that every piece of equipment operates with 100% uptime.

In mass production environments, constraints are usually easy to find; just look for large stockpiles of Work-In Process (WIP), backlogs, and frequent expediting (Bell, 2006). But in a lean manufacturing environment, none of these conditions should exist; therefore a different approach has to be taken in order to identify the system constraints(s). Chapter Two will cover some of the current literature gaps that exist when trying to apply theory of constraints to a lean environment.

Currently there are four major constraint identification methods (White, Sengupta, & Vantil, 2012):

1. The machine with the longest active state without interruption
2. The machine with the greatest percentage of cycle time and fail state
3. The machine with the longest average upstream queue length
4. The machine with the largest percentage of utilization.

As can be seen from the above list, all of the major bottleneck detection methods are useful for individual machines.

There are analogous methods for exploiting the constraint once it has been identified (White et al., 2012):

1. Increase the machine reliability (mean time between failures, or MTBF)
2. Decrease the time to repair a down machine (mean time to repair, or MTTR)
3. Increase the size of the buffers around the bottleneck machine.

For moving assembly lines, equipment failures and repairs are not the main reasons for line stoppages. Operators using the equipment, people maintaining the equipment, the people supplying parts to the assembly line and poke-yoke devices are the main causes. In most instances the assembly line stops only for seconds and in some cases it does not come to a complete stop but only slows down. As for increasing the buffer sizes, this is a direct contradiction of lean concepts. Increased buffer sizes leads to increase of Work-In Process and waste. Again, the main ways to exploit the constraint are based on variability reduction methods for individual machines.

There are no major methods for identifying the constraint in systems with paced moving assembly lines. And if the constraint can be identified there is not a method to exploit the constraint. The question, “How do you identify and exploit the system constraint when the typical methods do not apply?” sparked my research into the topic.

This is important because continuous improvement is necessary for a company’s survival and the gains of the blended methodologies have delivered results that were at least four times higher than any one approach alone.

Serial Production Line Defined

Different methods have been proposed to detect the constraints in serial production lines. Before detection methods are discussed a brief definition of a serial production line will be presented. Figure 1 represents a multi-product system with R products (labeled r = 1,…, R), and M manufacturing resources (labeled m = 1,…, M) separated by buffers (labeled b = 1,…, M-1). Note that a manufacturing resource can be of two types, an individual automatic station or a continuous moving assembly line.

Each manufacturing resource is modeled as a single server queue operating under first-in first-out rules. Each manufacturing resource requires an interval of time to process a part. This interval of time is referred to as the *cycle time*. Even though the term *cycle time* will be used, the calculated a*verage cycle time* will represent the interval of time for manufacturing systems with a product mix. The resources in the manufacturing system experience different downtime characteristics. The buffers can have different capacities for WIP from one another.

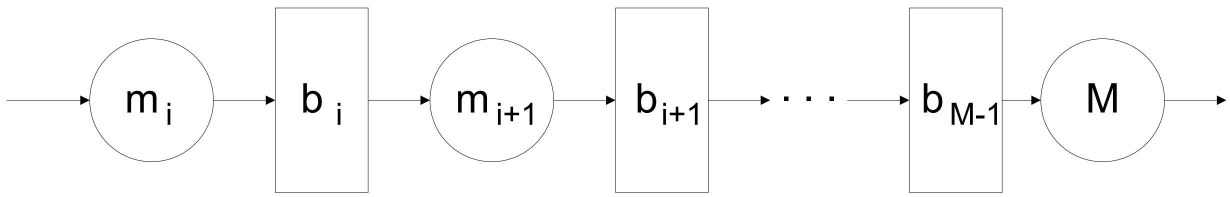

Figure 1: Serial production line

The manufacturing system being studied in this thesis is a serial production line. The facility produces vehicles and there are two product models. The production line consists of tandem assembly lines and several individual automatic stations separated by work in process (WIP) buffers. The buffer holds enough WIP to allow the assembly lines to temporarily run at different speeds without affecting (blocking or starving) one another.

There is a main assembly line is fed by a sub-assembly line. The main line is made up of five moving assembly lines. The names of the lines are Frame 1, Frame2/Final 1, Chassis 2/Final 2, Final 3/Chassis 3 and Inspection. The name of the sub-assembly lines are Trim 1, Trim 2, and Chassis 1. A layout of the main assembly line is presented in chapter three.

Key Lesson Learned

Manufacturing systems that implement theory of constraints techniques exceed the performance of those using only Lean Manufacturing and Just-in-Time (JIT). Theory of constraints systems produce more units and have an increased rate of throughput while at the same time show reductions in inventory, manufacturing lead time, and the standard deviation of cycle time (Watson et al., 2007).

Even though this thesis will present three methods to identify a system constraint, there are other methods in literature that could have possibly accomplished the same task. In fact, there are some authors who have used modified Lean tools to aid in constraint identification (Sproull, 2009) and (Pirasteh & Kannappan, 2013).

It all boils down to knowledge and data. One of the keys to being able to apply and combine different methodologies in different environments is knowledge of the process and the methodology and what each is trying to accomplish. If the right process data is collected and analyzed correctly, constraint identification can be accomplished in multiple ways. Just as each continuous improvement methodology has its own set of strengths and weakness, so will the various constraint identification methods.

CHAPTER TWO: OVERVIEW OF THE LITERATURE

The effectiveness of TOC has been reviewed extensively over time. For example, the extended literature survey by Naor, Bernardes, and Coman (2012), provides a great insight into the theoretical foundation of TOC. Rahman (1998) authored an article reviewing the theory of constraints philosophy and its applications. The article listed journals from which referred papers were selected for further analysis. This list of journals was used as a field of candidates for further consideration for this thesis. The number of journals was reduced to the list below based on the number of TOC articles published:

- *Journal of Operations Management;*
- *Production and Inventory Management;*
- *International Journal of Production Research;*
- *Industrial Engineering;*
- *International Journal of Operations & Production Management;*
- *European Journal of Operational Research.*

The literature search for this thesis includes articles with the publication period from 2000 to present. TOC based dissertations were not considered as a part of the literature search. These selected journals did not provide any articles directly related to the industry and manufacturing system being studied; therefore a second literature search was performed with the primary focus being constraint detection and production assembly lines. The time period was widened and several applicable articles surfaced from the following journals; *Journal of Intelligent Manufacturing, Mathematics Problems in Engineering, Robotics and Computer-Integrated Manufacturing* and *IEEE Transactions on Robotics and Automation.* A summary of the count of relevant articles per journal is shown in table 1.

Table 1: Selected journals for articles

Journal	Number of Articles
Journal of Operations Management	1
Production and Inventory Management	5
International Journal of Production Research	5
Industrial Engineering	8
International Journal of Operations & Production Management	4
European Journal of Operational Research	6
Journal of Intelligent Manufacturing	2
Mathematical Problems in Engineering	1
Robotics and Computer-Integrated Manufacturing	1
IEEE Transactions on Robotics and Automation	1
International Journal of Systems Science	1
Total	35

Several books and websites on TOC were also reviewed. This chapter provides a brief overview of selected literature on applying the theory of constraints. It does not cover all three branches of TOC's management tools: TOC's process of ongoing improvement (five focusing steps), logistics (drum-buffer-rope and buffer management techniques), and problem solving/thinking tools. Since the goal is to identify literature gaps for the process of ongoing improvement branch, the literature from the other two branches will not be reviewed.

Theory of Constraints as a Scheduling Program

Theory of constraints began with a request for help. A neighbor of Eliyahu Goldratt wanted to increase plant output, so the neighbor asked Goldratt for assistance. Goldratt developed a scheduling program that accomplished the goal of increased plant output.

The program was made available to the public and later became known as Optimized Production Technology or OPT. The popularity and use of OPT grew quickly. But the implementation of OPT did not guarantee success. The use of OPT contradicted most U.S. plant performance measurement systems. The conflict was aggravated by management allowing the plant workers to ignore the schedule produced by OPT (Watson et al., 2007). Goldratt decided to try and educate management and the plant workers by publishing the nine OPT rules.

Table 2: The nine OPT rules

1. Balance flow, not capacity.
2. Level of utilization of a non-constraint is determined not by its own potential but by some other constraint in the system.
3. Utilization and activation of a resource are not synonymous.
4. An hour lost at a constraint is an hour lost for the total system.
5. An hour saved at a non-constraint is just a mirage.
6. Constraints govern both throughput and inventory in the system.
7. A transfer batch may not, and many times should not, be equal to the process batch.
8. The process batch should be variable, not fixed.
9. Schedule should be established by looking at all of the constraints simultaneously. Lead times are a result of a schedule and cannot be predetermined.

Adapted from: (Watson et al., 2007)

Publishing these nine OPT rules did not seem to have the effect that Goldratt wanted. So he used an innovative method to get his ideas across to the general public. In 1984 with the aid of Jeff Cox he published a novel about production management entitled *The Goal* (Goldratt & Cox, 1984). The business novel accomplished more than merely educating the various companies who had bought the scheduling software, it became a bestseller.

Theory of Constraints as a Continuous Improvement Process

The Goal provides an outline of the steps required to improve a system. The Five Focusing Steps (5FS) allow the implementation of the theory of constraints concepts. The 5FS are: 1. Identify the System Constraint, 2. Decide How to Exploit the Constraint, 3. Subordinate Everything Else, 4. Elevate the Constraint and 5. Go Back to Step 1, but beware of "Inertia".

The 5FS are constantly evolving. They have evolved into the Process Of OnGoing Improvement (POOGI), which includes the original five steps united with two prerequisites. The first prerequisite is to define the system under investigation and identify its purpose, while the second is to define measurements that align the system to its purpose (Watson et al., 2007).

The second evolution is taking place now. Pretorius (2014) has identified several shortcomings with the five focusing steps. To address these shortcomings, the 5FS are transformed into a decision map that includes all five steps and the two prerequisites, but allows decision points to guide the user through the process. The answer to the first decision point, "Is the constraint physical?" is yes. Therefore to analyze the manufacturing system being studied, the first two steps of the five focusing steps do not change. Figure 2 shows Pretorius' decision map.

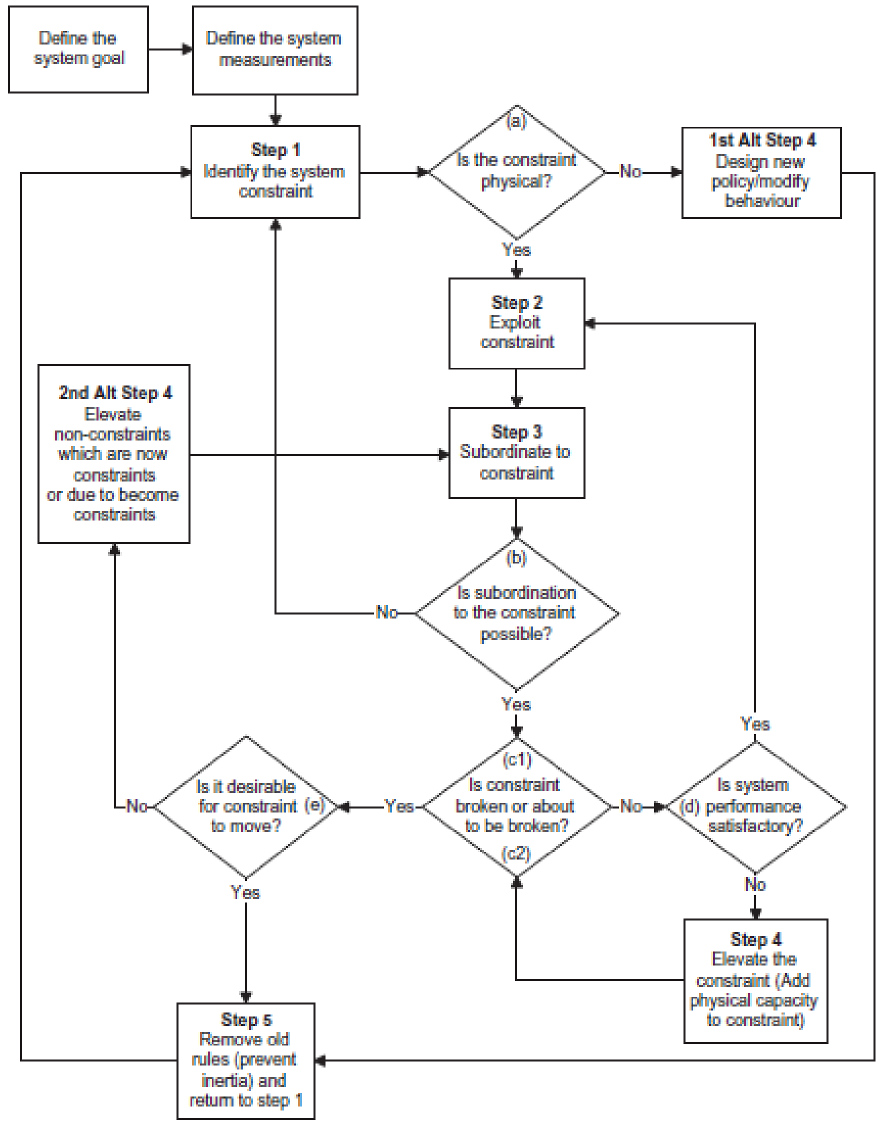

Figure 2: The five focusing steps with all in-between decision points included (Pretorius, 2014).

There are numerous applications of theory of constraints in literature in both non-manufacturing and manufacturing environments. Most of the reported applications have occurred in the United States. The TOC-Goldratt.com web site identifies 62 implementations of 81 referenced involve U.S. firms. The other 19 firms are from Canada, Germany, India, Ireland, Israel, Mexico, South Africa, UK, Uruguay and Venezulea (M. Umble, Umble, & Murakam, 2006).

Constraint Identification Approaches In Bernoulli Production Lines

(Kuo, Lim, & Meerkov, 1996) began defining their approach to finding constraints in production lines by formally identifying the manufacturing system assumptions. The assumptions are used to describe a Bernoulli production line, in which some are listed below;

(i) The system consists of M machines arranged in series, and M-1 buffers.

(ii) Machine 1 is never starved for parts and machine M is never blocked.

(iii) The machines have identical cycle time.

(iv) The buffers separate each pair of machines and are characterized by its capacity.

(v) Machine i, being neither blocked nor starved produces a part with probability p_i and fail to do so with probability 1- p_i.

A constraint in a Bernoulli production line is defined as the resource which has the greatest impact on the sensitivity of the system's performance index as compared to all other resources. This definition is very close to the conceptual definition proposed by theory of constraints. For the manufacturing system, the performance index is defined as the Production Rate of the last machine, which is the average number of parts produced per cycle of time.

Zhuang, Wong, Fuh, and Yee (1998) identify constraint resources by analyzing the distribution of WIP obtained from the mean queue lengths. Zhuang et al. (1998) model a manufacturing system which has the following characteristics: (a) all of the processes consist of multiple identical stations, (b) there are individual queues for most of the identical stations, which means the system does not follow a first-in-first-out discipline, and (c) a station can be blocked by stations not only in the succeeding process but also in the same process. Even though this is a very different manufacturing system from the one being studied in this thesis, Zhuang et al. (1998) made an effort through approximations to transform the system into a traditional single queue, single server system. This method of constraint detection is suitable for analyzing production systems with very large to infinite buffer sizes. For systems containing only limited buffers the results are not sufficiently accurate (Yan, An, & Shi, 2010).

Li and Meerkov (2000) discuss Bernoulli serial production lines with finished goods buffers. They address the problem of satisfying customers demand. They define due-time performance (DTP) constraints and introduce a method for their identification. The tool used for identification is based on the data available on the factory floor through real time measurements. The shortcoming of this constraint identification method is that it is not proved analytically and only numerical justification is achieved (Yan et al., 2010).

Constraint Identification Approaches in Markovian Production Lines

Chiang, Kuo, and Meerkov (1998) followed Kuo et al. (1996) methodology and began their system-theoretic approach to constraint identification by formally identifying the manufacturing system assumptions. The assumptions are used to describe a Markovian production line. The first two assumptions for the Markovian production line are the same as those listed for the Bernoulli production line. The key assumption that differentiates the two production lines is that Markovian production lines have machines where the up-time and the down-time are random variables distributed exponentially with parameters p_i and r_i.

Chiang et al. (1998) introduce three types of constraints:

1. Up-time and Down-time Constraint
2. Up-time and Down-time Preventative Maintenance Constraint
3. Constraint (if one resource is both up-time and down-time constraint).

This constraint detection method is not applicable to systems with an arbitrary number of resources or those in the automotive industry (Yan et al., 2010).

White et al. (2012) propose a method of constraint identification in serial production lines that uses inter-departure time data. The method is based on the assumption that the constraint machine is least affected by other machines in the system. The primary constraint is described as the machine with the minimum combined total time spent in blocked-up (idle) and blocked-down (full downstream buffer or a failed or blocked downstream machine) states.

A simulation model consisting of four-machines and three buffers is used to illustrate the application of the method. The main machine throughput rates are modeled such that a higher speed is expected from each machine as the product moves through the manufacturing system. The key descriptive statistical measure used to rank the system constraints is the mean absolute deviation. Additional research should be performed to extend the proposed method to variable cycle time machines.

The variability of the machine cycle times will not allow the use of the inter-departure time method for true Markovian processes. The method was included in this section of the thesis because of the way machine failures were modeled.

In the papers written by Kuo et al. (1996), Zhuang et al. (1998), Li and Meerkov (2000), Chiang et al. (1998), and White et al. (2012), manufacturing systems with linear topology (i.e., serial production lines) are analyzed, all of which process a single-product. All of the papers except White et al. (2012), assume that the cycle time for the manufacturing resources are the same.

The manufacturing system being studied in this thesis has a product mix of two models and all the manufacturing resources do not have identical cycle times. Also, all of the manufacturing resources are not of the same type. It is the aim of the next three chapters, to present methods to identify constraints for a more general manufacturing system.

CHAPTER THREE: CONSTRAINT IDENTIFICATION USING THE FLOW CONSTRAINT ANALYSIS APPROACH

This chapter presents the first of three constraint identification methods proposed in this thesis. Flow Constraint Analysis, is a holistic approach that evaluates whether the customer's demand is being satisfied. This evaluation is made by comparing the takt times and the cycle times of resources in the manufacturing system. Cycle times will come from one of two sources; a moving assembly line or an automatic station. Resources with cycle times higher than a calculated target value are likely to be the constraints. Figure 3 shows the manufacturing resources and various buffer capacities.

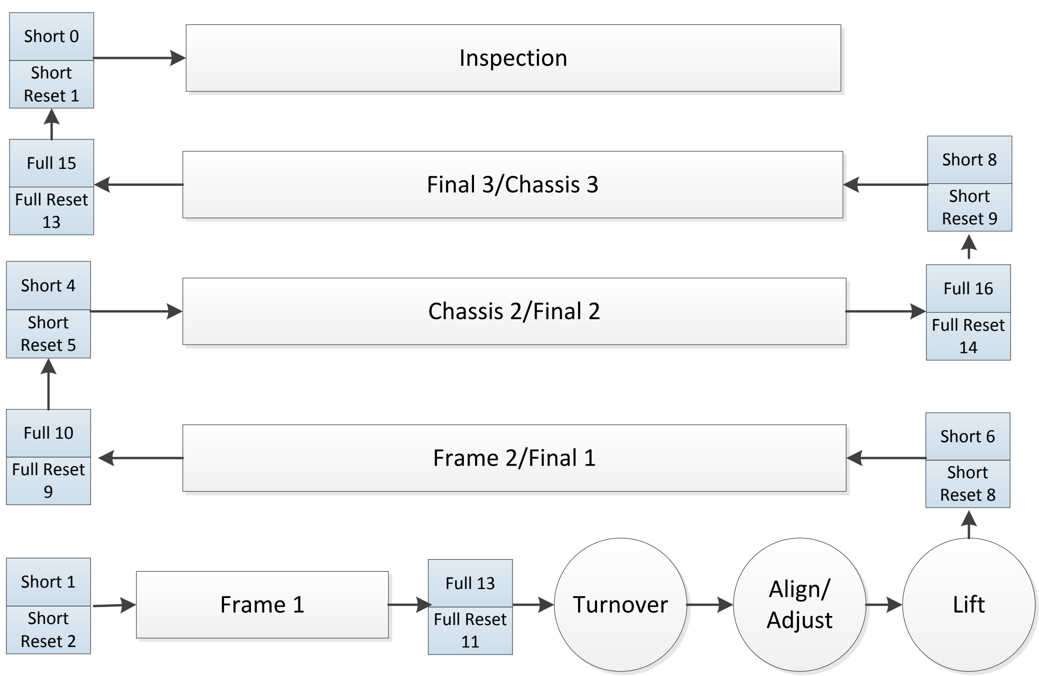

Figure 3: Manufacturing system layout

The Flow Constraint Analysis method involves a two-step process. The first step of the analysis is to determine if a true constraint is located in the manufacturing system. The second step is to identify secondary/tertiary constraints. In the first step, the existence of a constraint is determined by calculating and comparing the takt times and the cycle times. If the takt time is greater than the associated cycle time for each resource in the manufacturing system, the system is capable of meeting customer demand. The system constraint is then defined as being external. If demand exceeds the capacity of any of the resources of the manufacturing system a true bottleneck is said to exist (Fawcett & Pearson, 1991), which means the constraint is internal. Another method of defining an internal constraint resource is through spare capacity. Spare capacity is the difference between cycle time and the takt time (Bell, 2006). The resource with the least amount of spare capacity is the primary bottleneck for manufacturing systems with resources that have varying cycle times.

The manufacturing facility under analysis has an Andon system which collects the time and duration of events that occur. The Andon system also states a takt time for each assembly line in the manufacturing facility. This given takt time will be used during the analysis.

Another type of constraint resource also exists. These constraint resources have sufficient capacity when managed and scheduled carefully, but they could adversely affect the system's performance when managed inappropriately (Fawcett & Pearson, 1991). That is the purpose of the second step in the flow constraint method of analysis. The second step is to identify secondary/tertiary constraints by comparing individual times against each other.

When applying the flow constraint method, the process time for resources falls into one of four categories based on two characteristics. The first of the two characteristics is if the process time is dependent on the model mix, while the second is if the process time remains constant or varies from job to job. Figure 4 shows the steps to follow when applying the method.

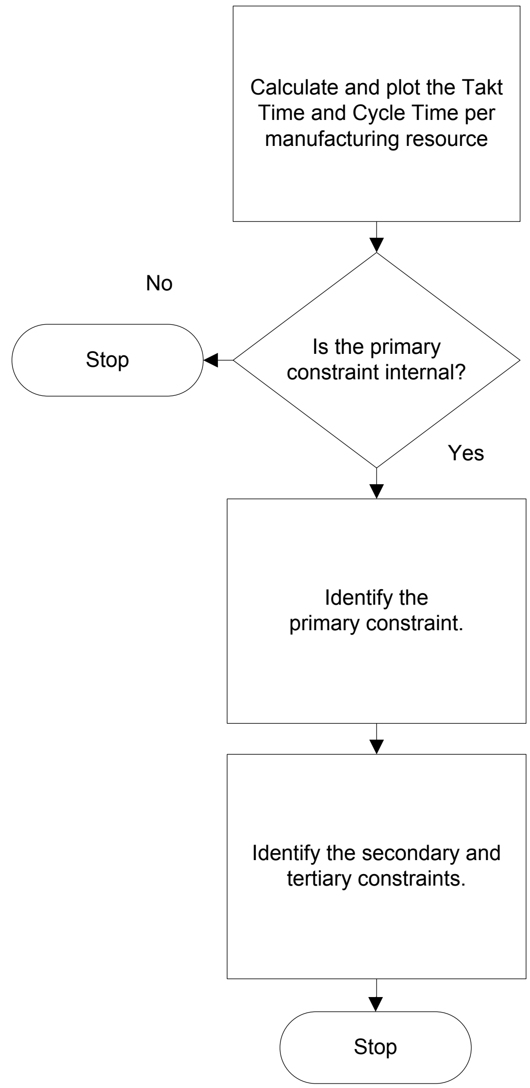

Figure 4: Flow Constraint Analysis Flow Chart

Moving Assembly Line Analysis

Statistical fluctuations apply to the performance of all resources (Stein & Dekker, 2003). One of the most prevalent sources of fluctuations is “natural” variability. Natural variability includes minor fluctuation in process time due to differences in operators, machines, and material (Hopp & Spearman, 2011) . Natural variability is ignored in the flow constraint method. Also, there are no effects to the assembly lines due to model complexity. The assembly lines are assumed to operate at a deterministic cycle. Usually, the cycle time can be obtained from the control panel. For this analysis the average cycle time was calculated for the resources and assumed constant. Continuous moving assembly lines process times can be categorized as non-model dependent and constant.

Results

The point chart below, Figure 5, shows the takt time and cycle time for each assembly line. The first three lines are the sub-assembly lines, while the remaining lines are part of the main line. The cycle time for the Frame 2/Final 1 assembly line is greater than the takt time for the line. Therefore the system is not able to meet customer demand and a true bottleneck exists, which is the Frame 2/Final 1 assembly line.

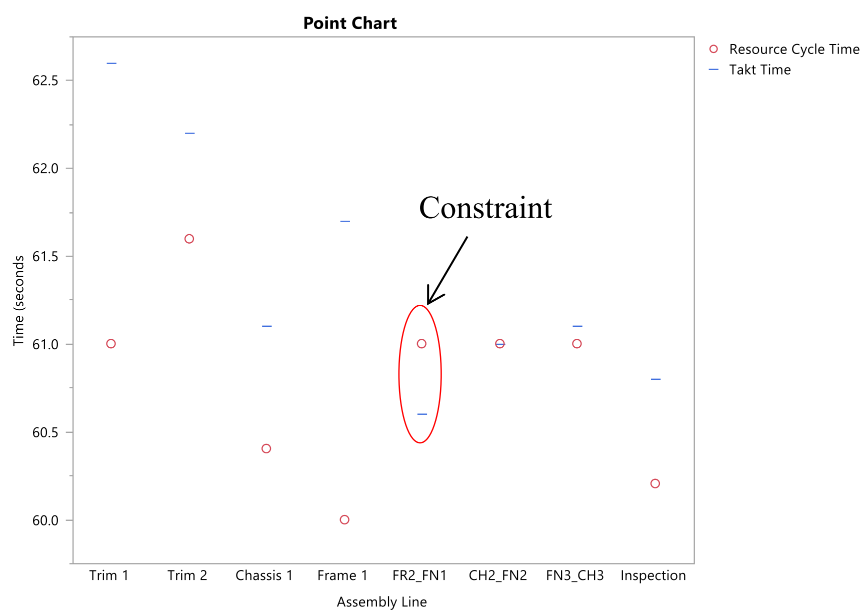

Figure 5: Takt time vs. Cycle time

Protective capacity is the capacity required to correct for late starts in the assembly of products (Stein & Dekker, 2003). Chassis 2/Final 2 and Final 3/Chassis 3 assembly lines cycle times are the same which means the system has potential dual constraints. Since both assembly lines have the same cycle time there is not a sufficient amount of protective capacity.

The resources upstream of the inspection assembly line have higher average cycle times. The inspection assembly line therefore is not the designed constraint. This fact is being mentioned because this is a plant management philosophy.

The sub-assembly products are assembled to the main line products on the Frame 2/Final 1 assembly line. Comparing the cycle time for the sub-assembly lines with the Frame 2/Final 1 cycle time shows that two thirds of the sub-assembly lines have a higher cycle time. The sub-assembly lines could potentially starve the main assembly line of parts.

Individual Stations Analysis

Additional resources are located between Frame 1 and Frame 2 assembly lines. These resources will be added to the discussion to demonstrate the use of the flow constraint method to automatic stations. There is a turnover, two transfers, two alignment/adjust work stations and a belt elevator.

To aid in understanding the function of the resources, the sequence of operations will be described. The product is removed from its skillet (pallet) and rotated 180 degrees. The first transfer removes the product from the turnover equipment and places it into one of two alignment/adjust stations. The alignment/adjust stations are dedicated resources, which means only one type of product is processed by each. Next, transfer #2 removes the product from one of the alignment stations and places it one the belt elevator. The elevator lowers the product onto its skillet where it travels through the rest of the manufacturing system. The definition of station cycle time presented by Hopp and Spearman (2011) will be used to determine the cycle time for the automatic stations.

$$Cycle\ time = move\ time + queue\ time + setup\ time + process\ time + wait - to - batch\ time + wait - in - batch\ time + wait - to - match\ time \tag{1}$$

The system follows single piece flow and first in first out rules. The alignment/adjust station that services Part B has an automatic changeover process which starts prior to the part's arrival, therefore the setup time will be assumed to zero. Queue time will not be considered in the analysis using the Flow Constraint Method. Move time will be called process time for the transfers since this is their only job function; equation (1) reduces to,

$$Cycle\ time = process\ time \tag{2}$$

The turnover and belt lift fall into the independent and constant category. The transfers process time (move time) fall into the dependent and constant category. While the alignment/adjust stations have independent and variable process times. As before with assembly lines, the first step of the analysis is to calculate and compare the takt times and cycle times. For stations with random cycle times the maximum cycle time values are compared against the takt time. If only one station has maximum cycle time values that exceeds the takt time, then that station is the constraint. If multiple stations have maximum cycle time values that exceeds the takt time, a different comparison method is required. A comparison of the probability of the cycle time exceeding the takt time should be performed.

There is not a takt time given for the automatic stations from the Andon system. For demonstration purposes, the cycle time for the upstream assembly line, Frame 1, will be used as the reference time value the equipment should be operating below. This value, 60 seconds, has been selected to reduce the effects of blocking the Frame 1 assembly line.

Results

The turnover and belt lift follow the same type of analysis as the moving conveyors. After the data collection process, the average cycle time for the turnover station was 52 seconds and the maximum cycle time value observed on the belt lift was 39.8 seconds. Therefore these stations were eliminated from consideration as a possible constraint.

The next resources, the transfers are analysed by considering the effects of model mix. The processing time for transfer 1 is the longest for Part A, with a value of 33 seconds. While processing time for transfer 2 is longest for Part B, with a value of 24 seconds. These stations are also eliminated as a possible constraint, see figure 6.

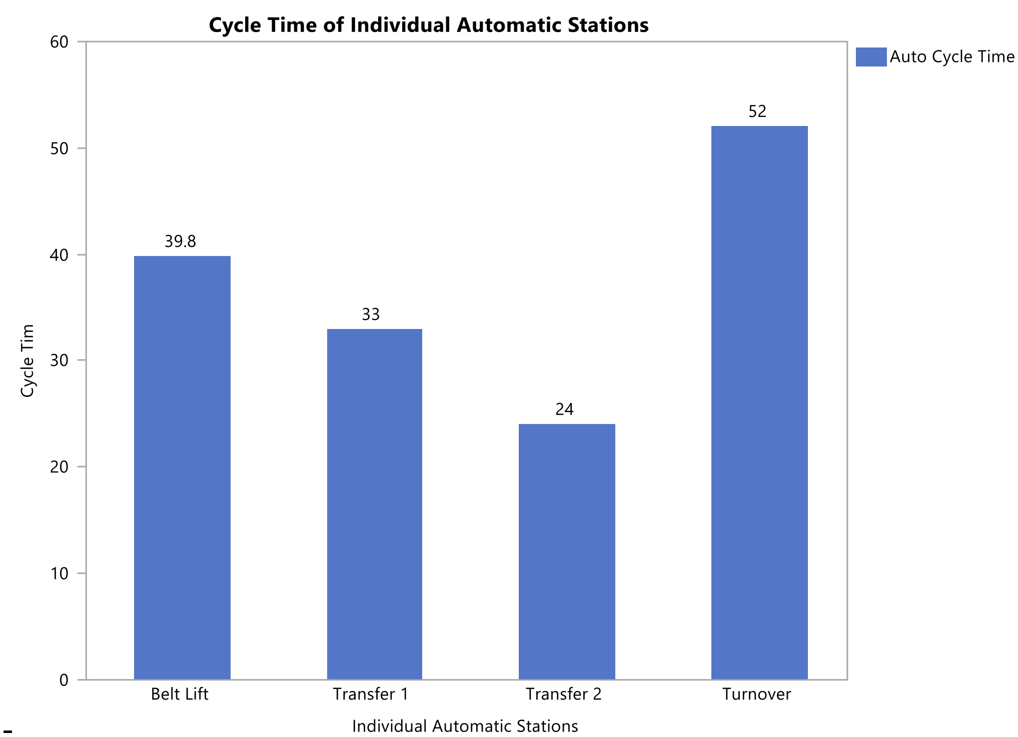

Figure 6: Automatic Stations Cycle Times

The cycle time data for the remaining two pieces of automatic equipment is random and therefore a histogram has been created for the sets of data. The Part 1 alignment/adjust operating range, in figure 7 is well below the target value and is not a system constraint. The Part 2 alignment/adjust histogram shows that the cycle time data is bimodal, in figure 8. Since we are concerned with constraint identification, the first mode is not considered during the analysis. The second mode, is also well below the target value. None of the automatic equipment is a system constraint.

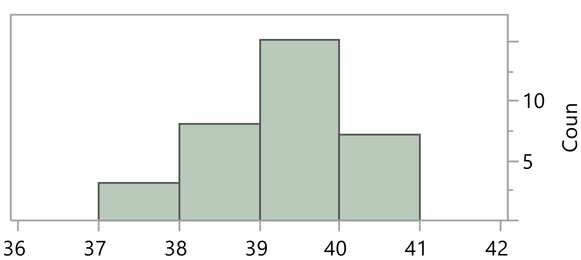

Figure 7: Part 1
Alignment/Adjust cycle time histogram

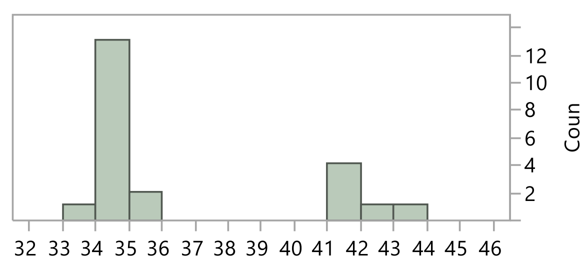

Figure 8: Part 2
Alignment/Adjust cycle time histogram

If both of the histograms had contained the target cycle time value, a probability distribution would have been fitted to the data. Next, the probability of the resource running at a cycle time greater than the target value would have been calculated and the resource with the highest probability would be the constraint resource.

CHAPTER FOUR: CONSTRAINT IDENTIFICATION USING THE EFFECTIVE UTILIZATION ANALYSIS APPROACH

The second proposed method of constraint identification is the Effective Utilization Analysis approach, which can be employed to pinpoint the location of the system constraint to a specific process or station. The actual production throughput is compared against the ideal capacity of the system. This method is based on the relationship between WIP, bottleneck rate and lead time for a constant work in process (CONWIP) system. Resources with the low effective utilization are likely to be the constraints.

Before expounding on the effective utilization method a brief definition of some of the method's terminology is required. There is a design reference line that is theoretically located through the center of the front axes of the vehicle. This line is known as the L10 line. The distance between the L10s is constant and is known as the assembly line pitch.

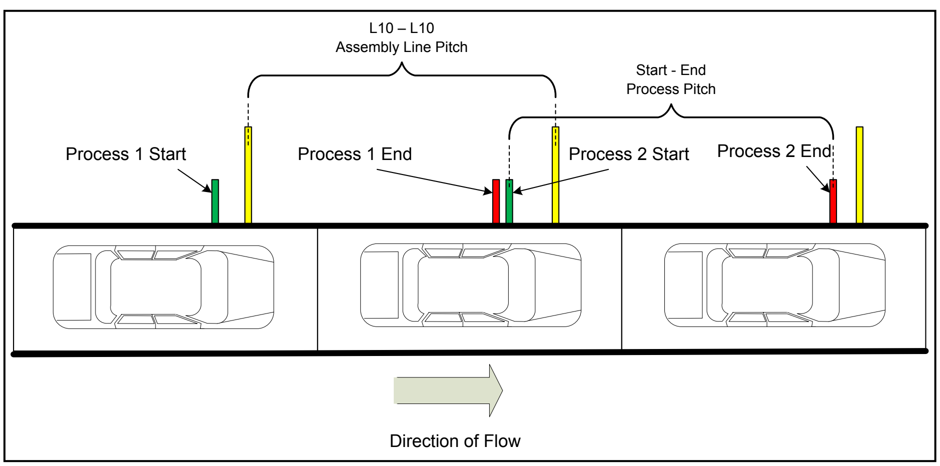

Figure 9: Assembly line pitch vs. Process pitch

The work performed by an individual operator is known as a process. The process should begin and end within a designated area on the plant floor. This area is known as the process pitch. All the processes do not have the same work content. Because of the differences in work content, the process pitch is not the same for all the processes. The start of the process pitches could also be different. In some rare instances the process pitch is greater than the assembly line pitch.

For assembly lines analyzed using the effective utilization method, model complexity and downtime effect the determination of the constraint resource. The effect of model complexity is captured in a variable named average cycle time. The processes performed by an operator are product model dependent. Since there is process time variability, the time weighted average is calculated for each process pitch and recorded as the average cycle time. Downtime durations are recorded in a variable named average downtime. First, the total downtime is calculated by summing up the downtime across all shifts over several days for each pitch. Next, the total downtime is divided by the number of products produced over the same time period. This calculation is performed for each process pitch. The average cycle time and average downtime values are summed. The summation is called process cycle time. This value will typically not be analyzed by itself. The reason is because of manpower allocation. There are typically two operators, one on each side of the product on the moving assembly line.

A comparison has to be made among all the operators working in a process pitch to ensure that the maximum time seen by the product in the process pitch is recorded. Logic is used in the next step of the method to determine the correct value to be recorded. The name of the recorded value is *pitch cycle time*.

Pitch cycle time can be assigned one of three values. The first logical decision is to determine if there are any processes in the pitch. If there are processes in the pitch, then the maximum *process cycle time* is compared against the inverse of the bottleneck rate for the assembly line. If the *process cycle time* is greater than the inverse of the bottleneck rate, then the maximum *process cycle time* is recorded as the *pitch cycle time.*

The next scenario occurs is if there are no processes in the pitch. In this case, the summation of the average downtime(s) for the pitch is compared against the inverse of the bottleneck rate. If the summation of the average downtime(s) is less than 1% of the inverse of the bottleneck rate, then a value of 0 is recorded as the *pitch cycle time.* The 1% value is based off of historical system performance data for the manufacturing system being studied. In theory is there are no processes taking place in a pitch, then that those pitch should not produce any downtime. Even though the product still has to physically travel through the pitch, a value of 0 is recorded because the only other acceptable value would be the inverse of the bottleneck rate. Recording the inverse of the bottleneck for pitches tends to improve the utilization value for the assembly line, therefore 0 is recorded instead.

Let us now discuss the third and final value that the *pitch cycle time* can assume. For the third possible value to be valid, two events can occur and produce the same result. The first event is; yes, there are processes in the pitch but no, the maximum *process cycle time* is not greater than the inverse of the bottleneck rate. The second event is; no, there are no processes in the pitch and no, the summation of the average downtime(s) is not less than 1% of the inverse of the bottleneck rate. If either of these events occurs, the *pitch cycle time* is calculated by summing the inverse of the bottleneck rate and the maximum average downtime value for the pitch. Figure 10 shows the steps involved in applying the method.

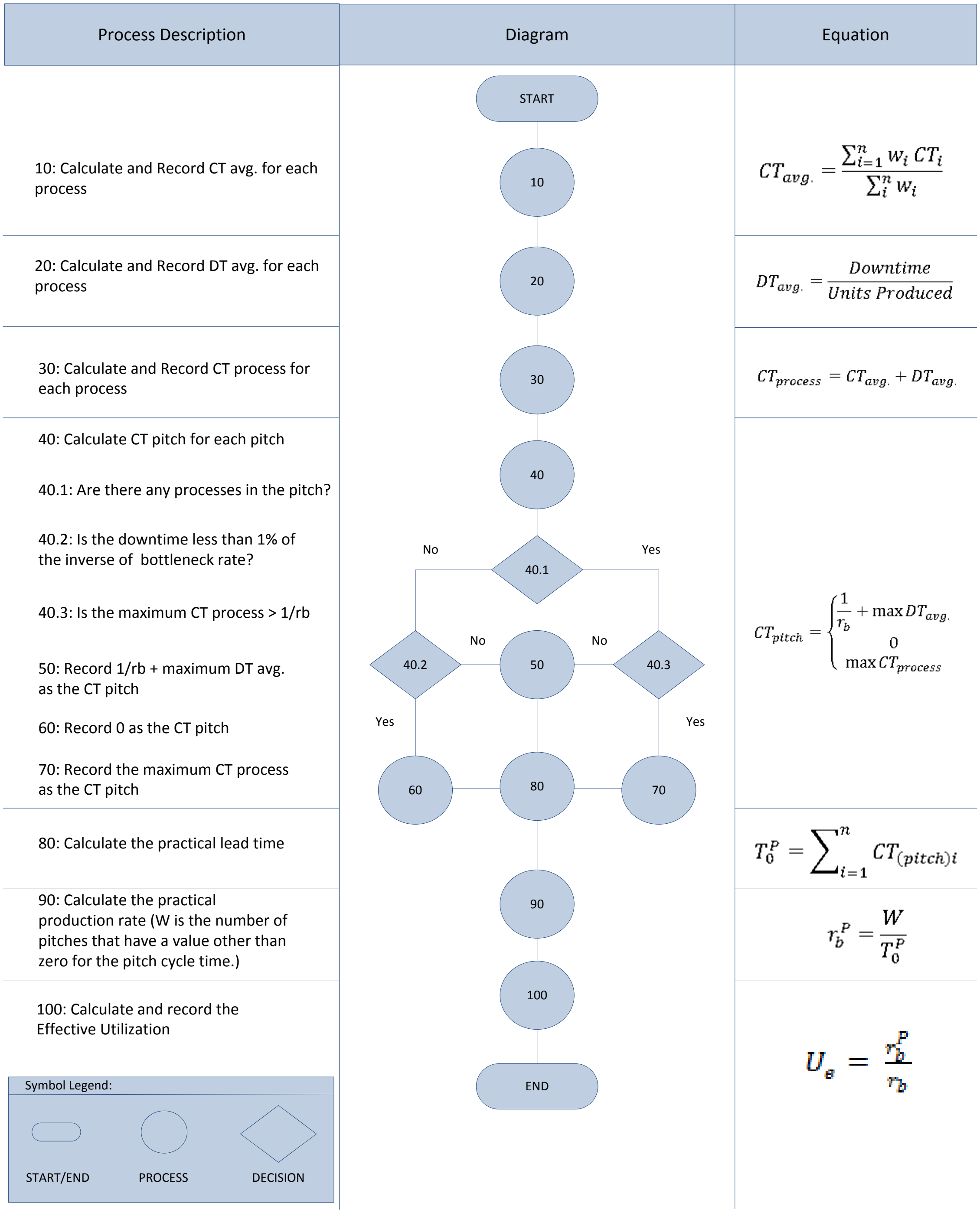

Figure 10: Flow chart of effective utilization method

Moving Assembly Lines Analysis

The analysis begins with the calculation of the practical lead time for the assembly line. This is accomplished by summing the *pitch cycle time* values. The assembly lines maintain a constant work in process (WIP). The actual WIP value has to be reduced to account for pitches with no processes and very little downtime. Therefore, only non-zero *pitch cycle time* values are counted to determine the WIP value for the assembly line. With the practical lead time and WIP calculated for the assembly line, the practical production rate can be determined from equation 3.

$$r_b^P = \frac{W}{T_0^P} \quad (3)$$

The utilization for the assembly line is determined by evaluating the ratio of the practical production rate with respect to the assembly line bottleneck rate, see equation 4. The bottleneck rate will be larger than the production rate because the bottleneck rate doesn't assume any losses due model complexity or downtime. This property can be used to verify calculations while performing the analysis.

$$U_e = \frac{r_b^P}{r_b} \quad (4)$$

Results

The flow chart shown in figure 5 will aid in following the steps required to perform the various calculations. Unlike traditional thinking where high utilization rates are associated with constraint resources, the opposite is true using this methodology. High utilization rates are desired, low utilization rates represent assembly lines that are constantly stopping. Among the sub-assembly lines Trim 1 has the lowest value, while Frame 1 has the lowest value for the main assembly lines, see table 3. Since the focus of the analysis is the main assembly line, the Frame 1 line is the primary constraint from the manufacturing system.

Table 3: Assembly line effective utilization rates

Assembly Line	Trim 1	Trim 2	Chassis 1	Frame 1	Frame 2 / Final 1	Chassis 2 / Final 2	Final 3 / Chassis 3	Inspection
Effective Utilization	93%	97.5%	97.6%	96.2%	96.7%	97.8%	98.1%	99%

Individual Stations Analysis

Now the effective utilization method will be used to determine the constraint for individual automatic stations. The method of calculating a utilization value for independent stations presented here is adapted from Hopp and Spearman (2011). They did an excellent job of explaining the concept and developing the applicable equations, see equation 5. This section of the thesis will apply those concepts and equations to the manufacturing system under study.

Instead of analyzing all the automatic stations, only the alignment/adjust stations will be reviewed. These stations were selected because of their more frequent stoppages.

$$Utilization = \frac{Arrival\ rate}{Effective\ production\ rate} \tag{5}$$

Model complexity and downtime effects will be considered in this analysis method. The effective production rate, in the denominator, will capture downtime effects.

Results

When calculating the utilization for individual stations using this method, a lower utilization is preferred. In the case of completely reliable stations connected as a serial production line, the station with the largest cycle time (the most utilized station) is the constraint station. The station dedicated to product A had the higher utilization.

Table 4: Alignment/Adjust stations utilization

Station	Utilization
Product A	55%
Product B	45%

Based on the utilization rates for the two stations, the station which services product A would be the system constraint because it has the highest rate.

CHAPTER FIVE: CONSTRAINT IDENTIFICATION USING THE QUICK EFFECTIVE UTILIZATION ANALYSIS APPROACH

Calculating the practical production rate, r_b^P, for each of the production lines is the third proposed method, in which downtime will be accounted for on continuous moving conveyors. Imagine being able to work into a manufacturing facility and identify the primary constraining resource in a short period of time with only real time data. That is the greatest strength of this method. The steps involved in calculating the practical production rate were covered in chapter 4 with the description of the Effective Utilization Method, which can be used for determining a manufacturing system's constraint.

$$W = r_b^P \; X \; T_0^P \qquad (6)$$

$$U_e = \frac{r_b^P}{r_b} \qquad (7)$$

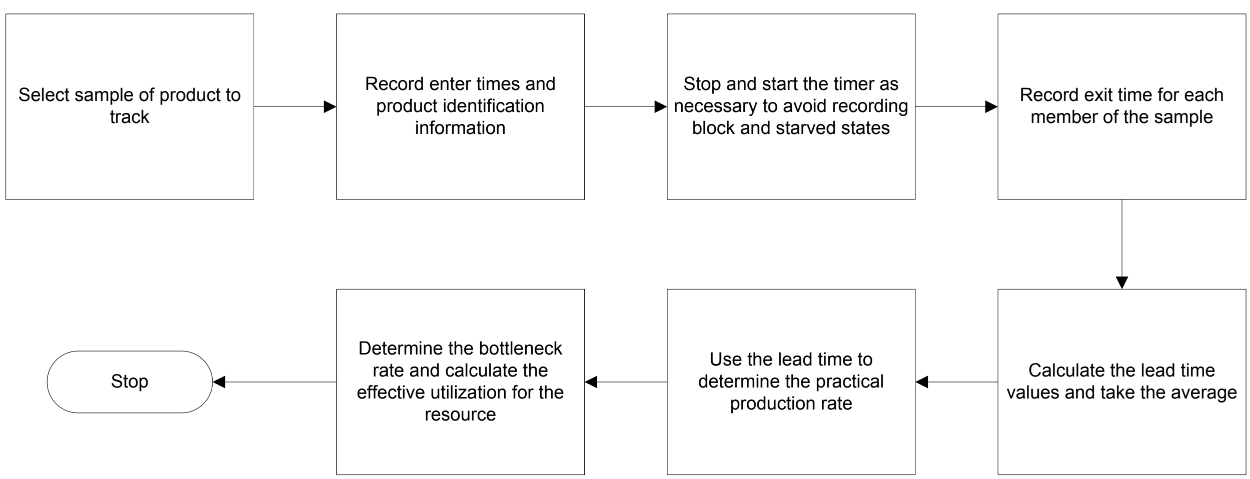

Figure 11 Quick Effective Utilization Flow Chart

Moving Assembly Lines Analysis

A quicker and rougher technique for calculating the practical production rate is to use the Quick Effective Utilization Method which will now be presented. From equation 1, which is Little's law, we are able to calculate the practical production rate r_b^P (Hopp & Spearman, 2011).

The CONWIP level, W, for the production line is obtained by waiting for a planned stop to occur, such as a break or lunch period, and then counting the number of parts on the line. To determine the minimum practical lead time, a small sample size of parts is selected to track through the production line. Start at the beginning of the line and record the time the part enters the line and the sequence/job number associated with the part. Walk the part through the production line and record the time the same part exits the line. Subtract the start time from the stop time to calculate the lead time. The goal is to include the time durations for minor stoppages on the production line in the lead time. The time durations that the production line spends in a blocked or starved state should not be included in the lead time calculation. Then calculate the sample set average lead time; this value is the minimum practical lead time, T_0^P. Table 1 shows a data sheet for the Frame 2 / Final 1 production line. Several columns in the table are left blank. The blank columns would have been used if the production line stopped for long durations of time (i.e. breaks, lunch, and excessive downtime).

Table 5: Data sheet

FR2_FN1	Seq. #	Start T1	Stop T1	Start T2	Stop T2	Time 1	Time 2	T_0 - Lead Time
1	17	8:42	9:10	-	-	0:28	-	0:28
2	9	8:43	9:11	-	-	0:28	-	0:28
3	10	8:44	9:12	-	-	0:28	-	0:28
4	38	8:45	9:14	-	-	0:29	-	0:29
5	8	8:46	9:15	-	-	0:29	-	0:29
6	7	8:47	9:16	-	-	0:29	-	0:29
7	12	8:48	9:17	-	-	0:29	-	0:29
8	5	8:49	9:18	-	-	0:29	-	0:29
CONWIP = 26 Frames							Average	0:28

Results

Using the data from Table 5, the practical production rate can be calculated.

$$26\ units = r_b^P\ X\ 28\ minutes$$

$$r_b^P = 0.929\ units/minute$$

With the practical production rate and bottleneck rate known, equation 7 can be used to calculate the effective utilization of the production line.

$$U_e = \frac{0.929}{0.983} = 94.5\%$$

One of the advantages of using the r_b^P is the ease of accounting for the production line's downtime in the simulation model. By simply multiplying the r_b^P times the assembly line pitch we get the conveyor's velocity which can directly be entered into the simulation data module without the need for additional modules to model production line stoppages.

Individual Stations Analysis

The individual stations can be modeled as a Bernoulli production line because only one station cycles at a time. Kuo et al. (1996) proposed the use of an indicator that allows the constraint resource to be identified using real time data. The frequency of blockages and starvations can be measured on-line. This leads to the following rule for constraint identification;

> Bottleneck Identification Rule: *If the frequency of manufacturing blockage of machine m_i is larger than the frequency of manufacturing starvation of machine m_{i+1} (either measured or calculated), the bottleneck is downstream of machine m_i. If the frequency of the manufacturing starvation of machine m_i is larger than the frequency of the manufacturing blockage of m_{i-1}, the bottleneck is upstream of machine m_i. If, according to this rule, there exist multiple bottlenecks, the primary one is the bottleneck with the largest Severity* (Kuo et al., 1996, p. 251).

Results

The constraint can be identified by just observing blockages and starvations of the machines. If the time spent in the blocked state is larger than the time spent in the starved state, the constraint is downstream of the buffer, otherwise the constraint is upstream.

CHAPTER SIX: DECIDING HOW TO EXPLOIT THE CONSTRAINT

The manufacturing system output is a function of the whole system, not just individual processes. When we view our system as a whole, we realize that the system output is a function of the weakest link. The weakest link of the manufacturing system is the constraint. Consequently, there needs to be focus on the coordination of efforts to optimize the overall system, not just individual processes. When a system matures in lean implementation, the main constraint becomes less obvious. However, the impact of performance of constraining resources in a lean system, especially with moving assembly lines, becomes even more critical. This research attempts to investigate the impact of location of constraints in a system. To facilitate that, the concept of "constraint" should be reviewed first.

The optimization process begins in step 2 of the 5 step continuous improvement process. Refer back to chapter 2 for a review of all 5 steps. Step 2 is deciding how to exploit the manufacturing system's constraint, where exploit means to get the most from the constraining element without additional investment (Breyfogle, 2003).

Constraints are often referred to as bottleneck, which limits the performance of the whole system. By exploiting the constraint, we strive to maximize the utilization of the capability of the constraining component as it currently exists. In other words, TOC urges to rethink what we can do to get the most out of this constraint without committing to potentially expensive changes or upgrades and implement in a short period of time (Dettmer, 1997). Constraints can be both external and internal. External constraints are often beyond the control of management because they are market driven. External or market constraints affect demand, they influence product mix, which in turn affects resource utilization (Fawcett & Pearson, 1991).

Lean manufacturing and TOC were two methodologies developed independently in the past. Based on Toyota Production System (TPS) and Just-in-Time (JIT) concept, the lean principles aims at eliminating waste to the maximum extent in order to improve the flow of value stream (Wan & Chen, 2008). Both Lean and TOC have proven themselves effective in productivity improvement in the past couple of decades. Nevertheless, some of the principles do not seem to agree totally between the two. Some related literature exist, such as the investigation of various configurations of pull production systems, including Kanban, Constant Work-In-Process (CONWIP) and different hybrid configurations that integrate the former two, while varying the location of the dominant constraint (Huang, Wan, Kuriger, & Chen, 2013). Watson et al. (2007) discussed briefly about the differences between JIT and TOC and also pointed out that there is a need for more supporting literature.

This thesis will explore the effects of the dominant constraint location on a manufacturing system that consists of paced moving assembly lines and individual automatic stations. Using simulation modeling, the performance of different scenarios can be compared.

A manufacturing system will be modeled following two different management philosophies. The first philosophy represents the typical manufacturing scenario where the constraint resource is located at the beginning of the assembly line. The downstream resources are allowed to run at a faster cycle time. The manufacturing system that follows this philosophy will be called Model A.

The second philosophy has the constraint resource at the end of the assembly line. This philosophy recommends that your nonconstraints have sprint capacity, that is, the capacity to produce product at faster rates than your constraint operation; thus, if an upstream nonconstraint operation experiences downtime for some reason, when it begins producing again, it should still have the ability to produce product at a fast enough rate to resupply the constraint buffer before the constraint buffer runs dry (Sproull, 2009). Model B will represent a manufacturing system following this philosophy, which is also called over speed. The key metric used for evaluating the models is throughput. Rockwell Arena simulation software was used to model the assembly operations.

In short, the two models compared in this thesis are:

- Model A: Constraint located at the beginning of the assembly line.
- Model B: Constraint located at the end of the assembly line.

Automotive Plant as Choice for Study

In this thesis, a real manufacturing company was used in the development of a simulation study. The manufacturing system being studied is the assembly operations in an automotive plant. There are several tandem, balanced, paced production lines. The plant produces two product families, which will be referred to as Part 1 and Part 2. Each production line has a fixed CONWIP level, but the level of the individual lines is not identical. The lines are decoupled from each other with Work-In-Process (WIP) buffers.

This system is considered a highly matured lean manufacturing system, and the buffer levels are typically well limited. As a result, the performance of constraints can bring a significant impact to the whole system.

It has been stated that performance analysis of production lines strive to evaluate their performance measures as a function of a set of system parameters. The most commonly used performance measures follow (Altiok & Melamed, 2007):

- Throughput
- Average inventory levels in buffers

- Downtime probabilities
- Blocking probabilities at bottleneck workstations
- Average system flow times (also called manufacturing lead times).

TOC states that the performance of the weakest link determines the performance of the whole chain. During this analysis, the constraint will be located at the beginning and then at the end of the manufacturing system. Even though the constraint bottleneck rate will be the same for both models, the system's throughput will change because of process variation and downtime. Throughput has been selected as the system performance measure.

By design, most manufacturing production lines provide for a limited amount of work-in-process inventory (Michael Umble, Gray, & Umble, 2000). The manufacturing system being studied is no exception. Each assembly line is located between two buffers. Both buffers have two set points; one stops the assembly line while the other restarts it. The upstream buffer set points are called short and short reset, while the downstream buffer set points are full and full reset. The buffer set points control the average inventory levels in the buffer, therefore this measure will not be tracked during the study.

Downtime will be restricted to only one of the individual stations. The assembly lines will use another technique to account for downtime called the Utilization Method. This method will be discussed in greater detail in the next section.

Model A has the constraint located at the beginning of the manufacturing system with all of the downstream resources running at a faster pace. Since downtime for the assembly lines will be modeled without the actual assembly line stopping, the probability of being blocked will be small since it will be a function of the automatic stations' downtime. System B has the constraint located at the end of the manufacturing system; therefore there are no opportunities for the constraint to be blocked. Because of these reasons blocking probabilities at bottleneck workstations will not considered a performance measure.

The structure of the data collected for the study does not allow the accurate tracking of average system flow time. Therefore, flow time will not be considered a performance measure.

Model Construction

The plant runs a two shift operation with scheduled overtime. For this study only one shift with 10 hours of production will be modeled. The simulation run contains a warm-up period that allows the system to fill with parts. The replication length and hours per day are set to 18 hours. All of the model resources will follow a capacity schedule which matches the current day shift 10 hour production schedule.

Model verification was performed by inspecting the model statistics to verify that a proper flow of entities was maintained and by observing the model's animation evolution to ensure that the manufacturing lean rules were being followed through the individual automatic stations, such as single piece flow and first in first out sequencing of parts. Figure 2 shows the logic for this section of the model.

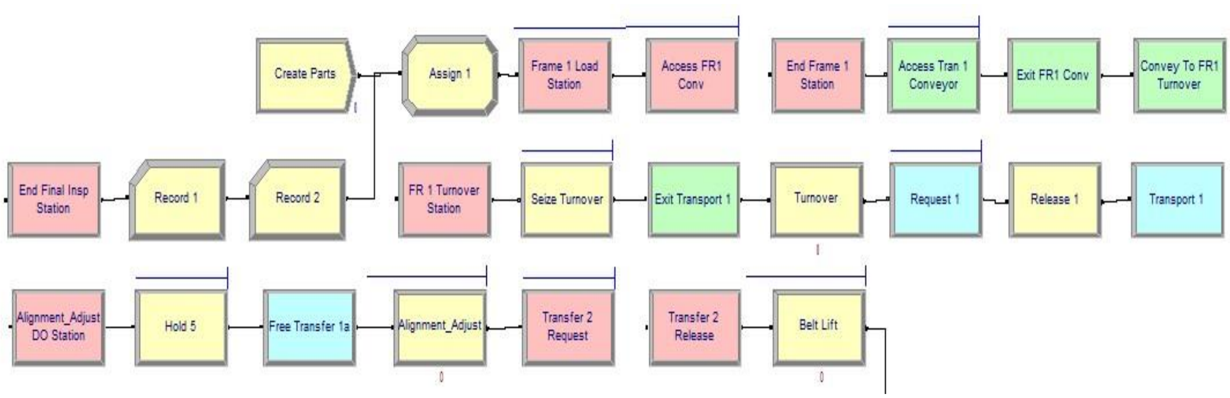

Figure 12: Individual stations simulation logic

The validation was carried out by running 8 replications of the model and comparing the average output from the model with the average output from 8 production days. Figure 3 shows the simulation logic for the production lines after the individual automatic stations.

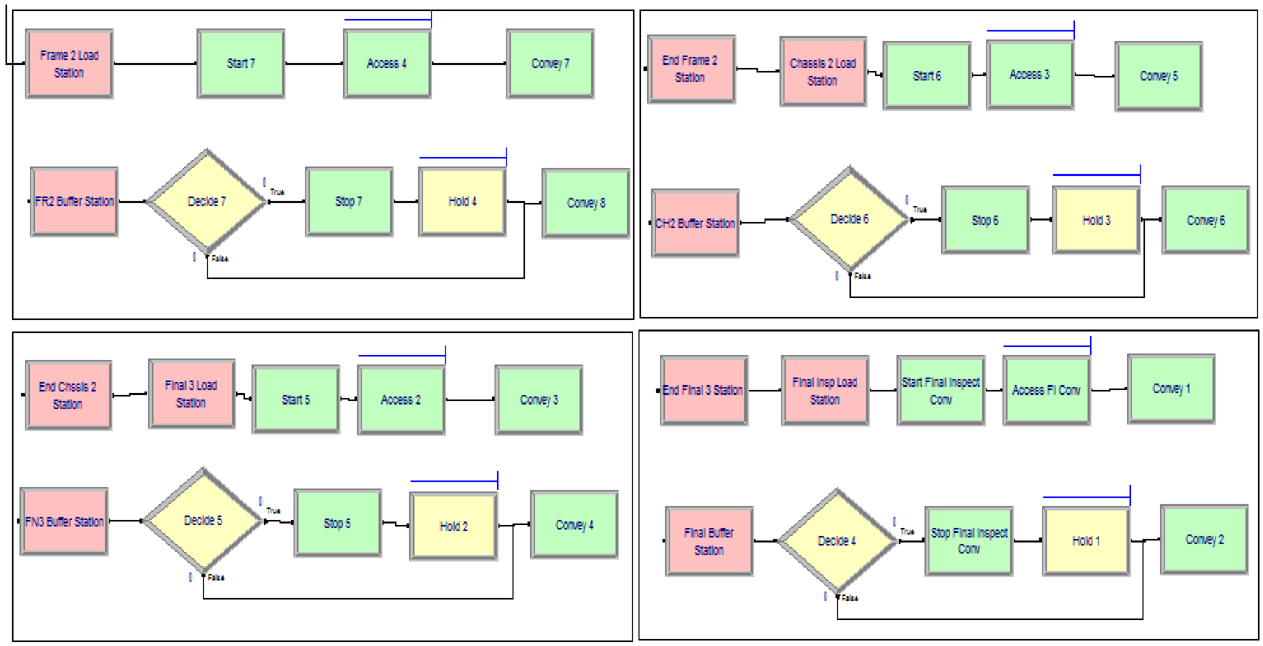

Figure 13: Production line simulation logic

The effective utilization for each of the production lines under study is shown in Table 6. Notice the difference between the utilization values calculated using the Quick Effective Utilization Method from chapter 5 and the Effective Utilization Method for the Frame 2 / Final 1 production line. The data used for the Effective Utilization Method calculations came from a larger sample size that covered multiple days and shifts along with minor calculation differences.

Table 6: Manufacturing system data

Assembly Line	Frame 1	Frame 2 / Final 1	Chassis 2 / Final 2	Final 3 / Chassis 3	Inspection
Effective Utilization	96.2%	96.7%	97.8%	98.1%	99%

Each utilization method identifies the system constraint as the resource with lowest utilization value.

Simulation Study

The purpose of the experiment is to determine one of the following three possibilities that is the output from two discrete event simulation runs using Arena software.

1. Model A produces a higher throughput than Model B
2. Model A and Model B throughputs are the same
3. Model B produces a higher throughput than Model A

There are five moving conveyor assembly lines and three automatic stations. The conveyors will be modeled to run at a constant practical production rate. The system's layout will remain constant.

The resource cycle times will be adjusted to create the two management philosophies. To determine the cycle time for the various conveyors, the model will start with the current physical limitations of the manufacturing system. The lowest cycle time for all the conveyors in the system, which is 60 seconds, is used as the cycle time for the fastest conveyor in the model. From the fastest conveyor, the rest of the conveyor cycle times are increased by increments of 1 second. The individual conveyor cycle times are converted to conveyor velocities by dividing the assembly line pitch by the cycle time. To account for conveyor downtime, the conveyor velocities will be multiplied times the lowest utilization rate of all the assembly lines, which is 96%.

The assembly lines will operate as continuous moving conveyors. This system characteristic means that the full and full reset buffer controls will not be required to be modeled.

The manufacturing system is simulated for 50 one-shift periods. Two different management philosophies are modeled with system throughput being the performance measure. In each of the models there is a model mix of two assemblies (i.e., two product types). The model buffering strategy for short state (starving) matches that of the actual production operations. Model A represents a system where the constraint is located at the beginning of the production line while Model B represents the constraint being located at the end.

In figures 14 and 15, the numbers in circles denote the machine cycle times used for the respective simulation model. The letters "DP" denote a machine that used a discrete probability distribution to model its cycle time. The numbers in the rectangles represent the short set point, which is the minimum number of units allowed in the buffer for the machine after the buffer to continue running.

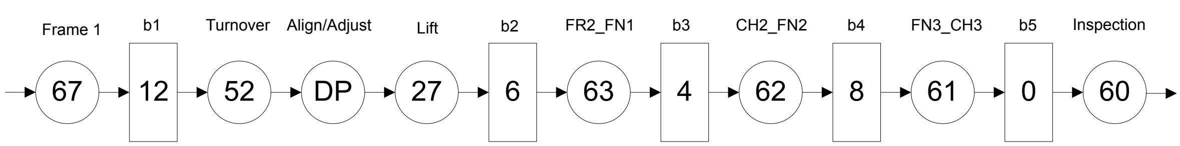

Figure 14: Model A system parameters

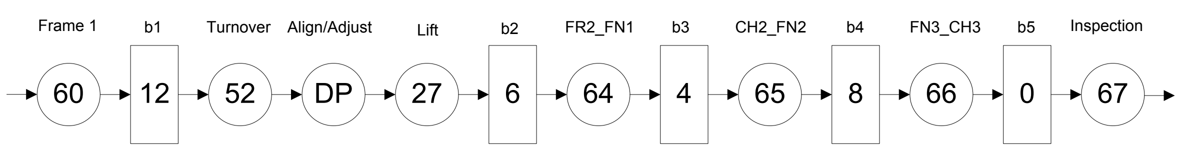

Figure 15: Model B system parameters

Buffer 1 is different from the other buffers in the system because there is not a short set point. Buffer 1 is an accumulating conveyor, which means it can actually be empty without stopping any of the other assembly lines or individual automatic stations. The value in buffer 1 represents the maximum capacity of the buffer.

The turnover station will operate at a constant cycle time of 52 seconds. The align/adjust resources will operate with random process cycle times, see Table 7. The belt elevator will operate at a constant cycle time of 27.3 seconds. Only the align/adjust work center resources will experience random failures.

Table 7: Discrete probabilities

Part A Cumulative Probabilities	0.06	0.28	0.68	0.96	1.00	-	-
Part A Station Cycle Time (seconds)	37	38	39	40	41	-	-
Part B Cumulative Probabilities	0.02	0.50	0.80	0.82	0.96	0.98	1.00
Part B Station Cycle Time (seconds)	33	34	35	40	41	42	43

Results

A t-Test: two-sample assuming unequal variance was performed on fifteen throughput values taken from each model to determine if the simulation results are statistically significant. The p-value was determined to be less than 5%, therefore the null hypothesis was rejected and the means are accepted as not being the same.

Table 8: Overall average of 50 simulated scenarios

	Model A	Model B
Average Throughput (units/minute)	0.6985	0.7560
Average Number Produced	477	516

The overall averages of the simulation results demonstrate that Model B produces a significantly higher throughput than Model A, the results are shown in table 8. In other words, locating the system constraint at the end of the manufacturing system coupled with running an “over speed” setup has benefits. This result contradicts with some literatures which suggest having the bottleneck at the beginning of an assembly system.

Based on observations, having the bottleneck at the beginning may be beneficial for systems with higher buffers and decoupled segments throughout the system, because it controls WIP level and reduces chance of blocking the constraining recourse. For a matured lean manufacturing system, especially with synchronized moving assembly line, having the bottleneck at the end, i.e., the “over speed” setup, actually is more beneficial. One of the benefits is that the constraint cannot be blocked, since there is no other process downstream. Productivity from the constraint in model A is lost because it experiences blockages due to the downtime associated with the align/adjust individual stations. According to Goldratt any productivity lost at a constraint is productivity lost for the total system. In model B, the first few machines have cycle time values that are closer to each other which will allow the system to make up for productivity losses.

Another benefit is that running with over speed forces the work in process (WIP) level to increase and eventually saturate the manufacturing system. It has been well documented that increasing the system’s WIP level will also increase the system’s throughput until the critical WIP level has been achieved (Hopp & Spearman, 2011).

A third benefit is that implementing the over speed philosophy in the planning stages of design of a manufacturing system will require little to no additional resources while simultaneously improving the manufacturing system’s performance.

CHAPTER SEVEN: CONCLUSION

Summary and Discussion

In this thesis, three methods have been proposed to locate constraints in matured lean systems, the uses of the three methods are slightly different as discussed below.

Table 9: Comparison of methods

Analysis Approach	Application
Flow Constraint	• Provides the users with the ability to determine the location of the enterprise constraint. • If the there are multiple system constraints they can be quickly identified. • Use for early planning phases, such as before the manufacturing system has been installed or when there are plans to increase the capacity of an existing facility.
Effective Utilization	• Requires more computations and data. • Use as part of a continuous improvement program.
Quick Effective Utilization	• Requires very little computation and data. • Real-time data is collected and analyzed. • Used in situations where downtime data is not available for the production line, during verification of a simulation model or when the production line being modeled is no significant.

The Effective Utilization Method requires the user of this method to spend more time and effort. Once the analysis is complete, the users of this method will have the ability to narrow the focus of changes to an actual process. The users will also have deeper understandings of the manufacturing system and the effects model complexity and downtime have on the system.

Contribution of this Research

An investigation into the impact of location of constraints in a manufacturing system was also carried out. Having the constraint in different locations (i.e., upstream or downstream) can affect the dynamics of the system and hence results in different performance level. The level of WIP allowed in system also plays an important role.

A simulation study was performed with a model that represented a real manufacturing system with a series of tandem moving assembly lines and single piece flow. Fifty different scenarios were tested, and the results reveal that having the constraint at the end of the system, i.e., the "over speed" setup, can make the lean system more productive. The main reason is that the over speed setup makes the lean system saturated with limited WIP and thus reduces chance of starving the constraining resource.

Future Research Suggestions

This research can be further extended to include other aspects of the TOC methodology. After the constraint is identified, appropriate decisions can be made on exploiting the constraint to further improve the system. Insights into the system constraints can also facilitate redesigning a segment of or the whole system to be compliant with the concepts of TOC. Research into subordinating the non-constraints in a lean manufacturing system would be the next logical step in application.

Further investigating of the impact of different WIP levels allowed in the system, as well as considering other manufacturing system configurations can be considered. The uniting of Lean and TOC simultaneously can contribute to a better paradigm of manufacturing systems design.

REFERENCES

Altiok, T., & Melamed, B. (2007). *Simulation Modeling and Analysis with Arena*. New Jersey: Academic Press, Inc.

Bell, S. (2006). *Lean Enterprise Systems: Using IT for Continuous Improvement*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Breyfogle, F. W. (2003). *Implementing Six Sigma Smarter Solutions Using Statistical Methods*. Hoboken, New Jersey: John Wiley & Sons, Inc.

Chiang, S. Y., Kuo, C. T., & Meerkov, S. M. (1998). Bottlenecks in Markovian production lines: a systems approach. *Robotics and Automation, IEEE Transactions on, 14*(2), 352-359. doi: 10.1109/70.681256

Dettmer, H. W. (1997). *Goldratt's Theory of Constraints - A System Approach to Continuous Improvement*. Wisconsin: ASQ Quality Press.

Fawcett, S. E., & Pearson, J. N. (1991). Understanding and Applying Constraint Management in Today's Manufacturing Environments. *Production and Inventory Management Journal, 32*(3), 46-55.

Goldratt, E. M., & Cox, J. (1984). *The Goal*. Croton-on-Hudson, NY: North River Press.

Hopp, W. J., & Spearman, M. L. (2011). *Factory Physics* (3rd ed.). Illinois: Waveland Press.

Huang, Y., Wan, H., Kuriger, G., & Chen, F. F. (2013, June 26-28). *Simulation Studies of Hybrid Pull Systems of Kanban and CONWIP in an Assembly Line*. Proceedings of Flexible Automation and Intelligent Manufacturing Conference. Porto, Portugal.

Kuo, C.-T., Lim, J.-T., & Meerkov, S. M. (1996). Bottlenecks in serial production lines: A system-theoretic approach. *Mathematical Problems in Engineering, 2*(3), 233-276. doi: 10.1155/s1024123x96000348

Li, J., & Meerkov, S. M. (2000). Bottlenecks with respect to due-time performance in pull serial production lines. *Mathematical Problems in Engineering, 5*(6), 479-498. doi: 10.1155/s1024123x99001209

Naor, M., Bernardes, E. S., & Coman, A. (2012). Theory of constraints: is it a theory and a good one? *International Journal of Production Research, 51*(2), 542-554. doi: 10.1080/00207543.2011.654137

Pirasteh, R. M., & Kannappan, S. (2013). The Synergy of Continuous Process Improvement. *Industrial Engineer, 45*(6), 41-45.

Pretorius, P. (2014). Introducing In-between Decision Points to TOC's Five Focusing Steps. *International Journal of Production Research, 52*(2), 496-506.

Rahman, S. (1998). Theory of Constraints A Review of the Philosophy and Its Applications. *International Journal of Operations & Production Management, 18*(4), 336-355.

Sproull, B. (2009). *The Ultimate Improvement Cycle: Maximizing Profits through the Integration of Lean, Six Sigma, and the Theory of Constraints*. New York: Productivity Press.

Stein, R. E., & Dekker, M. (2003). Identifying The Constraint *Re-Engineering the Manufacturing System*: CRC Press.

Umble, M., Gray, V., & Umble, E. (2000). Improving Production Line Performance. *IIE Solutions, 32*(11), 36-41.

Umble, M., Umble, E., & Murakam, S. (2006). Implementing Theory of Constraints in a Traditional Japanese Manufacturing Environment: The Case of Hitachi Tool Engineering. *International Journal of Production Research, 44*(10), 1863-1880.

Wan, H.-d., & Chen, F. F. (2008). A Leanness Measure of Manufacturing Systems for Quantifying Impacts of Lean Initiatives. *International Journal of Production Research, 46*(23), 6567-6584.

Watson, K. J., Blackstone, J. H., & Gardiner, S. C. (2007). The Evolution of a Management Philosophy: The Theory of Constraints. *Journal of Operations Management, 25*(2), 387-402.

White, T., Sengupta, S., & Vantil, R. P. (2012). A New Way to Find Bottlenecks. *Industrial Engineer, 44*(11), 45-49.

Yan, H.-S., An, Y.-W., & Shi, W.-W. (2010). A New Bottleneck Detecting Approach to Productivity Improvement of Knowledgeable Manufacturing System. *Journal of Intelligent Manufacturing, 21*(6), 665-680.

Zhuang, L., Wong, Y. S., Fuh, J. Y. H., & Yee, C. Y. (1998). On the role of a queueing network model in the design of a complex assembly system. *Robotics and Computer-Integrated Manufacturing, 14*(2), 153-161. doi: http://dx.doi.org/10.1016/S0736-5845(97)00023-9

VITA

Trumone Sims is from Tallulah, Louisiana. He graduated with a Bachelor of Science degree in mechanical engineering from Southern University and A & M College in Baton Rouge, Louisiana in May of 1994. He is currently pursuing a Master of Science degree in advanced manufacturing and enterprise engineering from the University of Texas at San Antonio, San Antonio, Texas.