A Case Study of Computer Service

Parts Inventory Management

Pinaki Kar
New York, NY
pkar@alum.mit.edu

Ming Li
Analog Devices
Norwood, MA
li_ming@alum.mit.edu

Edmund W. Schuster†
23 Valencia Drive
Nashua, NH 03062
edmundwschuster@aol.com
www.ed-w.info

† Corresponding Author

Pinaki Kar is interested in the application of operations research and statistical techniques for planning and decision support across a wide range of business issues. His experience spans multiple industries that including, pharmaceutical, chemical, high-tech and insurance.

Ming Li is interested in the application of optimization methods and heuristics in the fields of inventory control and logistics planning. He has experiences in the fields of telecom and semiconductors industries.

Edmund W. Schuster is interested in the application of models to logistical and planning problems experienced in industry. Ed can be reached through his personal web site, **www.ed-w.info**

1.0 INTRODUCTION

In this article, we present a case study of the service parts logistics operations for a company involved in supporting an installed base of computing machines. The company offers hardware and software support for network servers and workstations housed on both UNIX and Windows platforms. Customers predominantly are Fortune 500 firms with massive information processing infrastructures. To provide quick service when component failures occur, the company maintains a network of parts banks in the United States and Europe. Coordination of this network occurs from a central call center that is the initial point of contact for all customer service requests. The company also employs field engineers, both directly and under contract, in each of the geographical areas that it provides service.

Mid-sized and headquartered in the Boston area, primary competition comes from the service parts operations of large OEM vendors in the computer industry. Privately held, the current owner of the company was also the founder. Since its inception, the company has maintained a competitive strategy based on 1) high service levels, 2) flexible service plans to meet the needs of a wide range of customers, and 3) a single point of service for equipment manufactured by different vendors. The commitment to deliver flexible service plans differentiates the company from OEM vendors that tend to offer a standardized service plan for all customers. This strategy has led to impressive revenue growth during the past decade.

However, with the economic downturn beginning in 2001 the company has experienced declining profit margins primarily from the rising cost of operations. This happened because of an increase in the complexity of the installed base of network servers and workstations. In addition, increased competition from OEM vendors forced downward pricing pressure for new service contracts further squeezing profit margins. Rising operational costs, invigorated competition, and the lack of pricing power in the market led the company to conduct a review of current operations and established practices.

This case study comprises two parts. In the section 2.0, we provide a summary of the business operations and important issues faced by the firm. In section 3.0, we describe two mathematical models that address these issues. These models provide a means of determining the amount and location of inventory necessary to achieve optimal service levels. The balance of section 3.0 focuses on the prospects of gaining more information about the demand rate for service parts. We explore the use of change point methods and the application of Auto-ID Technology.

To maintain confidentiality, we do not disclose the company name in any part of this article.

2.0 BUSINESS OPERATIONS

In the following subsections, we discuss the fundamental aspects of the business including the structure of service contracts, the evaluation of new contracts, and the typical process flow for a service request from a customer. The section concludes with a discussion of the main issues that caused the recent decrease in profitability.

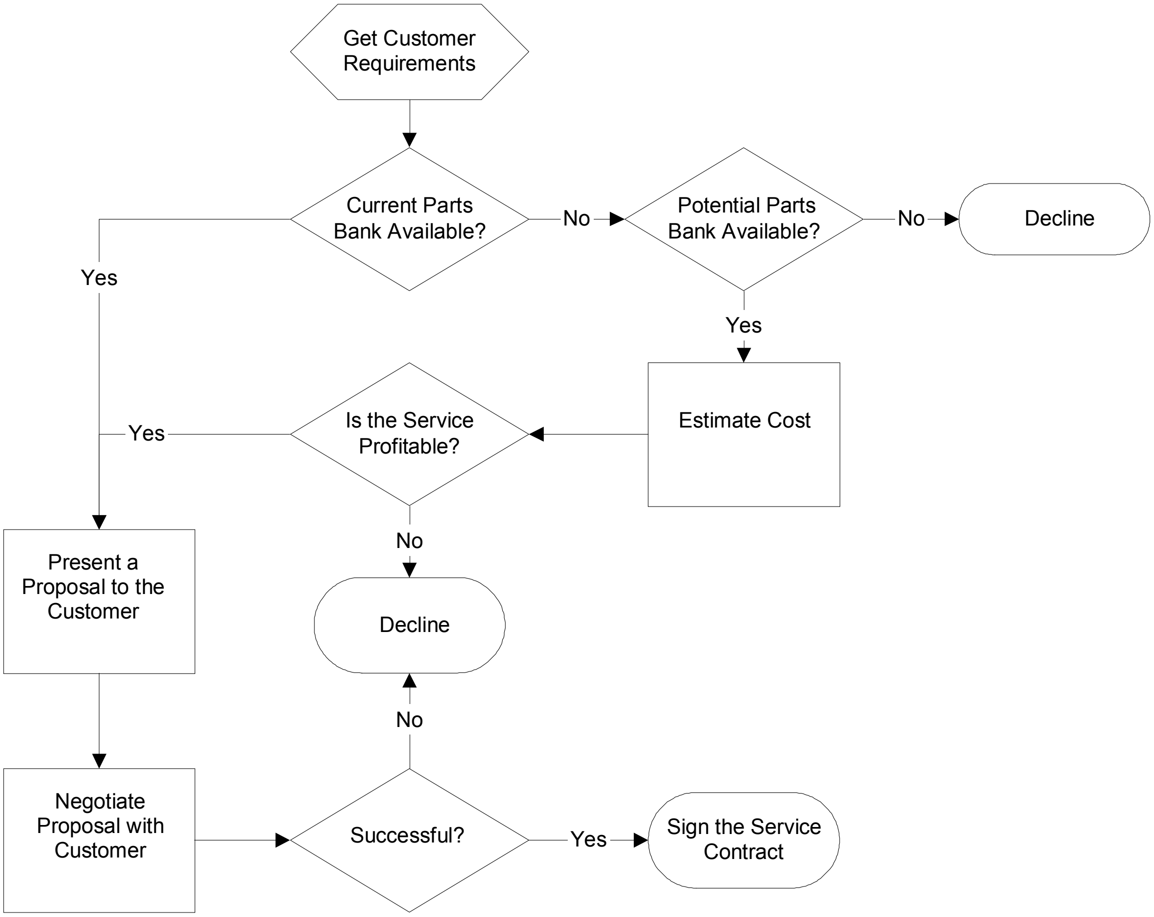
2.1 Contracts

Detailed contracts are negotiated for all business conducted by the company. A typical service contract specifies the following:

- Types of equipment supported – This includes a listing of the make and model of the computing machines supported at each location. There is usually no indenture level information for any of the equipment. Much of the information captured in the contracts is incomplete. Updates are infrequent.
- Term of contract – Length of time the contract is in force, usually one or two years.
- Type of service – hardware or software support, or both.
- Guaranteed response times - The contract specifies the maximum allowable response time for a service request at each location. The most frequent choices for response times are same day (2, 4, 6, and 8 hour response) and next day.
- Penalty clauses – Not meeting guaranteed response times for a pre-specified percentage of the requests results in a penalty. This pre-specified percentage is the contract service level. Usually, the service level is between 80% and 90% and is specified in the contract.

Before entering into each service contract, the company performs a detailed evaluation of the customer's requirements, taking into consideration the existing capabilities of the parts bank network. This evaluation results in an estimate of the incremental profit contribution for each potential new customer. A flow diagram of the evaluation process appears in Figure 2-1.

Figure 2.1 Business Process for Evaluation and Signing of Service and Support Contracts with Customers



Because the company has done a number of service contracts in the past with profitable outcomes, the process described in Diagram 2-1 is an established practice. Though this process provides a structure to ensure profitability during the life of the contract, it depends upon information about the installed base. Often, during the lifetime of the contract, the installed base changes causing additional complexity, lower service and increased costs. Once the contract is in place, it is seldom renegotiated because of a change in the installed base. This presents a significant management challenge for the company to overcome.

2.2 Service Network

The service parts network is a two-echelon distribution system consisting of a central depot and 50 regional parts banks maintained by third part logistics providers (3PLs). These facilities serve about 2,000 customer sites. The 3PL providers make local delivery of parts to customer sites. The company also maintains parts banks onsite at the customer location for contracts that guarantee a two-hour response time. A min/max (s, S) continuous review replenishment policy is the method used by the company to determine reordering for all of the parts banks. All defective parts replaced at customer sites are shipped back to the central depot for repair and testing before being returned to stock.

Since most of the service parts have similar weight, the transportation cost depends on the distance between the parts banks and the customers. Parts banks are usually located within 200 miles of the customer site if the customer requires one-day part delivery service.

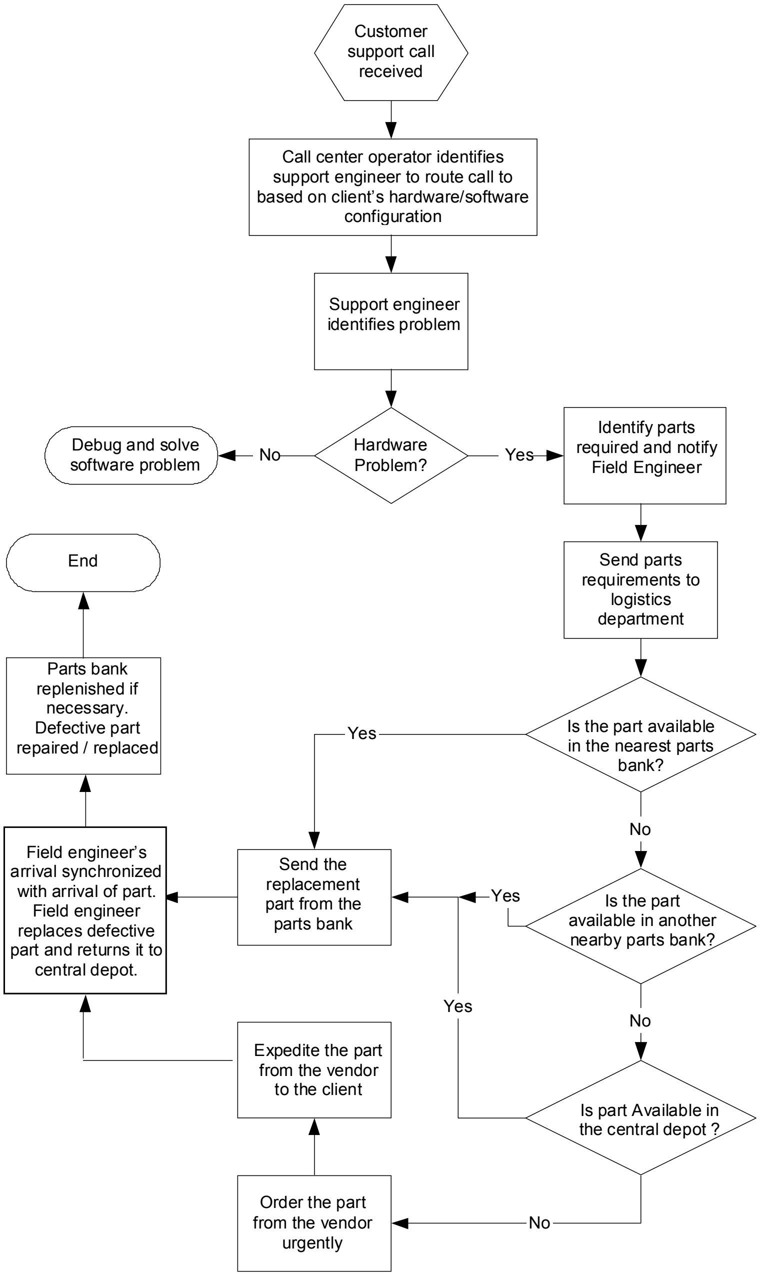
2.3 Inventory Planning

The safety stock levels at the parts banks are determined based on the historical demand rate. For lower demand items, the safety stock level is simply the demand during the previous month. For higher demand, the safety level is two times the demand during the last 15 days. These rules apply to both the parts bank and the central depot. For a new part where no historical demand data exists, the company uses the safety stock level recommended by the manufacturer of the part.

2.4 Fulfillment Process

Once a service contract is in force, the company maintains a disciplined fulfillment process described by Figure 2-2.

Figure 2-2 The service operation



This fulfillment process is the critical aspect of day-to-day operations. To meet the specified service levels, the right stock must be at the right place at the right time. As a result, a significant investment in inventory exists. Managing this critical asset is the most important variable in maintaining profit margins. It is easy to over invest in inventory as a simple means of increasing service levels. With increasing complexity of the installed base, this leads to reduced cash flow.

2.5 Critical Business Issues

Like any business, the company faced a number of important operational issues. Three primary issues were thought to be significantly affecting the profitability of the company.

- **High Inventory Cost, Low Service Level** - Inventory carrying cost for all the parts located in the Parts banks and the Central Depot was 25% of annual service contract revenue. This investment in inventory produced an average parts availability of less than 80%. This compares unfavorably to other companies in the computer service and support industry as surveyed by Cohen and Agarwal (1999). The following table lists some comparative results of the survey.

COMPANY	Inventory Investment/ Service Revenue (%)	Service Level (% On-time Parts Availability)
1	10%	91.6%
2	23%	88%
3	40%	88%
4	17%	90%

- **High Scrap Rate** – Because of obsolescence, the company scrapped approximately 4% of the service parts inventory cost per month.
- **High transportation costs** - The annual transportation cost equaled 15% of the total value of the inventory. In contrast, the survey conducted by Cohen and Agarwal (1999) finds that the average transportation costs for companies ranged between 2.2% to 11.4%. Stock outs on frequently used parts resulted in high expediting costs. In addition, there were significant lateral transshipments between Parts banks to meet backorders and correct stock allocation based on recent demand.

These three issues produced a great amount of downward pressure on operating margin. With decrease profits, the company was willing to consider new methods that would improve productivity.

3.0 SUGGESTED SOLUTIONS

To address the business issues outlined in section 2.5, the firm needed to make better trade-offs between inventory carrying cost and service. Given the complexity of the service parts network, these cost and service trade-offs required implementation of mathematical models for effective solutions. It is an established fact that traditional time-series methods to calculate optimal inventory levels do not work for service parts because the demand is "sparse." This means that demand is often zero for many periods. From our initial review of operations, we recommended the implementation of two models, METRIC developed by the RAND Corporation and used by the Air Force, and a mixed integer programming model to optimize the number and location of parts banks (see Bazovsky 1961; Graves, S.C., 1985; Hillestad and Carrillo 1980; Sherbrooke 1992). Both of these models are established methods that optimize cost and service trade-offs. Through implementation of these models, we project significant cost savings to the company. This alone would be a major improvement.

However, METRIC and the mixed integer LP did not address the issue of a lack of information about the installed base. This information is critical in determining projected demand. There are two ways to deal with this problem, the change point method, or the application of Auto-ID Technology.

3.1 The Change Point Method

At the most basic level, demand for service parts arises from the failure of one or more components in a computing machine. In our study of the company, we considered only failures that occur during the useful life of the part. Warranties usually cover initial malfunctions during the "burn-in" phase. Because of rapid obsolescence for computing machines, wear-out is hardly an issue. Few components reach the end of the operating life prior to the replacement of the computing machine.

Excluding burn-in malfunction and wear-out, the remaining failures that occur during the useful life of a part happen because of "chance" occurrence. In this case, sudden breakdowns take place without advanced signs of pending failure. For this situation, the Poisson distribution correctly approximates the failure rate. The Poisson distribution is discrete and commonly used in reliability modeling.

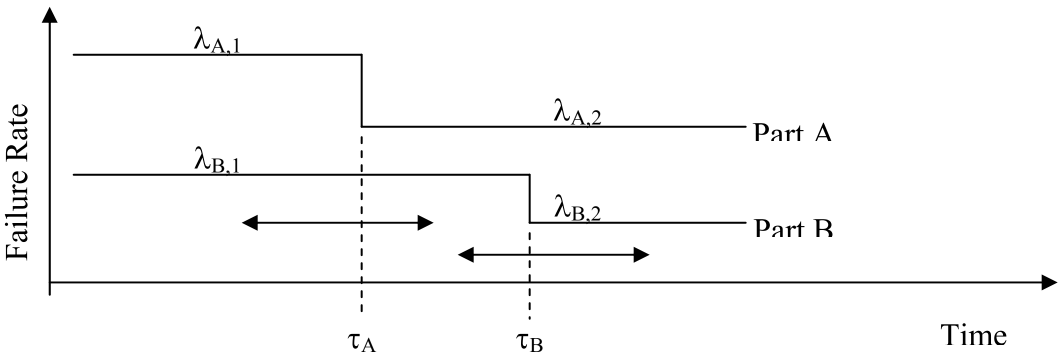
The demand rates for service parts depend on 1) the number of machines being supported (the *installed base*) and 2) how much these machines are used (the *program usage*). However, service providers for computing machines, like the company in this case study, often have only a general idea of the installed base, and program usage. Because of this lack of information, there is no alternative but to rely upon time series data of failures as a means of calculating safety stock levels. The problem with this method is that there is often a significant time lag between a change in the installed base, or program usage, and a change in the failure rate. As well, it is very difficult to correlate the installed base, measured in machines, to failures measured in demand for parts, because some components are interchangeable with other machines at the same location. Adding to complexity, service support organizations typically stock thousands of parts making manual review of changing failure rates difficult to accomplish.

To deal with this problem, we developed a mathematical technique for inferring changes using only historical failure data. The underlying concept of the technique is that a sudden change in the failure data might signal changes in the installed base or program usage (Akman and Raftery 1986; Loader 1992). This technique is robust, easy to implement and can detect with statistical significance the exact time when a change in the demand rates for a given part has occurred. When implemented as part of an automated service parts planning system, the technique highlights possible changes in the installed base through analysis of time-series failure data. By highlighting potential changes, planners can then determine whether to adjust stocking levels at the parts banks.

3.1.1 Conceptual Framework

To provide a conceptual example, we direct attention to Figure 3.1. This graph shows the failure rates over time for two theoretical parts, labeled A and B, that constitute a computing machine. The failure rates, represented by λ_A and $\lambda_{B,}$ are constant though time with change points occurring at τ_A and $\tau_{B.}$

Figure 3.1 Change points for the failure rates for two constituent parts of a given machine.



Using the statistics, we can determine if the failure rate before and after τ_A and τ_B is significantly different at a level of 95%. If this statistical significance exists, then we infer that τ_A and τ_B, indicate an actual change in the installed base or program usage.

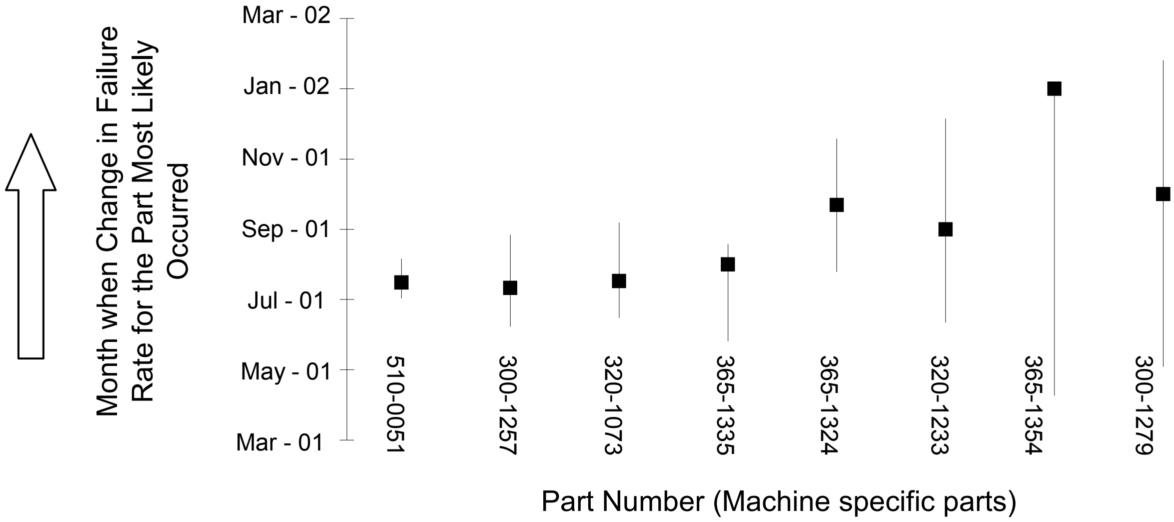
Planners can use the indicator information to determine if a change actually took place through physical verification of the site containing the computing machines. This becomes extremely important for inventory planning. With an actual decrease in the installed base or program usage, planners can immediately reduce the total inventory in the two-echelon distribution system, avoiding an inventory buildup of potentially obsolete parts. This in turn reduces inventory write-offs and improves profit margins.

In the case of a verified increase in the installed base or program usage, planners can boost inventories in the two-echelon distribution system, ensuring that service will remain at targeted levels specified by the contract. This helps the company to avoid financial penalties associated with low service levels.

3.1.2 An Example

We obtained time-series failure data from the company to test our theories about the application of change point methods to diagnose a change in the installed base. The company provided failure data for two computing machines installed at a customer site that both contained common parts. We then calculated change points for each part. This result appears in Figure 3.2.

Figure 3.2 Change point – month where change in installed base for each part in the indenture most likely occurred- for parts specific to the machine



The dots in the graph represent change points and the lines represent the confidence level at 95%. A short line indicates there is a high probability that the change point occurred within a tight interval of time. The plot suggests a change in the demand rate for parts between July and September 2001 (part numbers 510-0051, 300-1257, 320-1073, 365-1335, 365-1234, 320-1233). The other two parts (365-1354 and 300-1279) shown in the graph have very high confidence bands. This is because these parts have low failure rates (see figure 3.3 below) and hence very few data points to estimate the change point. Nevertheless, a change for these parts is also indicated around the same time. Since all are specific to the two computing machines there is strong likelihood that a change took place in the installed base or program usage. As an action item, the planner makes verification and adjusts stocking levels at the parts bank based on the new demand rate.

Figure 3.3 Failure Rates for parts with and without considering the change in installed base

		Failure Rates with and without considering Change Point for Part in Change Month			
		Change in Installed Base not Considered	Installed Base Change in Change Month		% Over-estimation of failure rate if change ignored
PART ID	Change Month	Avg. Failure rate	Failure rate prior to change	Failure Rate after change	
510-0051	Jul-01	6.8	12.8	2.4	278%
300-1257	Jul-01	3.2	6.5	1.5	211%
320-1073	Jul-01	39.2	49.4	31.9	123%
365-1335	Aug-01	7.7	11.8	4.7	163%
365-1324	Sep-01	2.6	3.9	0.8	323%
320-1233	Sep-01	3.1	4.7	1.5	206%
365-1354	Jan-02	2.3	2.7	0.0	
300-1279	Oct-01	3.8	4.7	2.4	156%

The inferred failure rates for use in the planning algorithms for the parts with and without considering the change point are show in Figure 3.3. The failure without considering the likely change in installed base is significantly higher than the failure rates if the change in installed base were considered. Not considering the change would have led to a higher stock level for all the parts regardless of the planning technique used.

While change point methods provide indirect means of determining if an increase or decrease in the installed base has occurred, Auto-ID technology offers the prospect of immediate identification and monitoring of the entire installed base. In the next section, we discuss some of the aspects of Auto-ID technology that are important to increasing productivity in the service parts industry.

3.2 The Application of Auto-ID

Auto-ID technology holds great promise to solve age-old problems in organizing information about the installed base, through the application of low-cost, Radio Frequency Identification (RFID) tags (Dinning and Schuster 2003). By placing RFID tags on individual parts, along with readers located in a machine or operated by a technician, Auto-ID technology will provide instant two-way communication by merging information with physical goods. Auto-ID technology also incorporates the Internet as a means of transmitting data gathered from RFID tags, combined with a new communication protocol called PML (Physical markup language). Using PML, the opportunity exists to gather online telemetry information about the performance of a part located in a computing machine. With time, RFID tags will become part of integrated circuits within computing machines, thus reducing the need to apply separate tags.

Though the economics of tags are not fully understood at this time, the future looks bright for application of this technology to improve service parts inventory management. Since many parts for computing machines are expensive, the cost of a tag, ranging from 30 cents to ten dollars, is a small fraction of the purchase price. It is true that identification using barcodes is currently cheaper. However, the older bar code technology suffers from several important shortcomings.

All bar codes, traditional and two dimensional, have limitations including a) the need for a direct line of sight from the scanner to the bar code, b) the ability to read only one code at time, and c) bar codes often require human intervention to capture data or to properly orient parts in relation to readers. In addition, bar codes provide only one-way communication and inventory systems are typically only updated on a batch basis. Bar codes seldom provide real time information or Internet connectivity to the data. There is always a chance the bar code will be missed or in other cases, read twice. Finally, bar codes can be damaged or compromised in a way that makes them impossible to read. Auto-ID technology is designed to overcome all of these limitations and make it possible to automate the scanning process, providing real-time data about the installed base.

3.2.1 How Auto-ID Works

Auto-ID Technology provides an infrastructure and set open standards that make it possible for wide adoption of RFID technology by industrial firms. Four components comprise Auto-ID Technology:

- EPC (electronic product code)
- ONS (object naming service)
- PML (physical markup language)
- Savant (data handling)

The EPC is a numbering system that contains enough combinations to identify trillions of objects. This is necessary because the ultimate goal is to provide a structure for low cost identification at the item level, meaning every single part will have its unique code. PML is the communication format for the data and it is based on XML (extensible markup language) that is gaining popularity in eCommerce transactions. PML represents a hierarchal data format to store information. By having a standardized means of describing physical objects and processes, PML will facilitate inter- and intra-company commercial transactions and data transfer. ONS acts as a pointer to connect the EPC to the PML file stored on a network, either a local area network or over the

Internet. It performs a similar function to the Domain Naming Service (DNS) of the Internet, which connects a text web address to an underlying IP address. An IP address is comprised of a 32-bit numeric address written as four numbers separated by periods, to find resources over the Internet. However, with the EPC, we start with a number and use ONS to find the product information linked to that number. Savant is a lower level software application that processes the data and performs error checking and de-duplication procedures in the event that more than one reader receives a signal from the same tag. It handles the scalability problem associated with the massive amount of data captured by Auto-ID. To summarize, the EPC identifies the product, PML describes the product, and ONS links them together.

With an extensive information technology infrastructure, Auto-ID is positioned to provide the real time data needed to monitor changing installed base or program usage. Though it might be five years before Auto-ID technology penetrates the service parts market, it has a bright future.

4.0 CONCLUSION

From this case study, we conclude that the application of several different mathematical models, and new identification technologies such as Auto-ID, will greatly improve operational costs in service parts inventory management. Specifically, the inventory control method METRIC reduces inventory carrying cost by providing a specific procedure to calculate safety stock based on expected service levels. This is an improvement over heuristics used by the company to calculate safety stock. Mixed-integer programming, an established method that provides insight about optimal network design, works well when applied to the parts bank location problem encountered by the company. With this method, planners can make sure the correct number of parts banks are established to balance cost with distance to the customer. Finally, we note that the most significant problem experienced by the company involved incomplete information about the installed base and program usage. Change point methods have the potential to be an indirect indicator of changes in the installed base or program usage through the analysis of failure rates. In contrast, Auto-ID technology offers the opportunity to obtain detailed information from two-way communication with computing machines located at a customer site. In the near term, change point methods offer the most cost effective alternative because Auto-ID technology is, in many cases, still in a development stage. Longer term, we feel that Auto-ID technology will become a productive means of controlling service parts inventory networks. We plan more research in this area.

REFERENCES

Akman V.E. and A. E. Raftery, 1986, "Asymptotic Inference for a Change-Point Poisson Process," *The Annals of Statistics* 14, pp. 1583-1590.

Bazovsky, I., 1961, Reliability Theory and Practice, Prentice-Hall Inc. (Englewood Cliffs, NJ).

Cohen, M. and V. Agarwal, 1999, "After-Sales Service Supply Chains: A benchmark Update of the North American Computer Industry," *Fishman-Davidson Center for Service and Operations Management*, The Wharton School, University of Pennsylvania (Philadelphia).

Dinning, M and E.W. Schuster, 2003, "Fighting Friction," APICS – The Performance Advantage 13:2.

Graves, S.C., 1985, "A Multi-Echelon Inventory Model for a Repairable Item with One-for-One Replenishment," *Management Science*, 31, pp. 1247-1256.

Hillestad, R.J. and M. J. Carrillo, 1980, "Models and Techniques for Recoverable Item Stockage When Demand and the Repair Processes are Non-Stationary – Part I : Performance Measurement." *RAND Corporation*, N-1482-AF (Santa Monica, CA).

Loader, C.R., 1992, "A Log-Linear Model for A Poisson Process Change Point," *The Annals of Statistics* 20:3, pp. 1391-1411.

Sherbrooke, C.C., 1992, Optimal Inventory Modeling of Systems: Multi-Echelon Techniques, John Wiley and Sons Inc. (New York).