

Selected Papers in Security Studies: Volume 7
Data Supply Chain Management:
Supply Chain Management for Incentive
and Risk-based Assured Information Sharing

Technical Report UTDCS-36-11
The University of Texas at Dallas
Department of Computer Science
December 2011

Bhavani Thuraisingham

Selected Papers in Security Studies: Volume 7

Data Supply Chain Management: Supply Chain Management for Incentive and Risk-based Assured Information Sharing

Bhavani Thuraisingham

The University of Texas at Dallas

ABSTRACT

In this paper we introduce the notion of data supply chain management and draw parallels between supply chain management and developing a data product. Then we discuss information sharing in supply chain management and discuss risks and incentives for information sharing. Our objective is to implement the Department of Defense Information Sharing Strategy whose goal is to “*Recognize and leverage the Information Sharing Value Chain.*”

This is the seventh in a series of reports we are writing on Security Studies and the application of information technology for providing security and combating terrorism. We will include papers on both cyber security and national security. The purpose of these series of reports is to guide us in the technologies we are developing for both cyber security and national security. The technologies include systems for assured information sharing and assured cloud computing and tools for secure social network analysis and data mining for security applications such as malware detection. Our research to develop these technologies is supported by the Air Force Office of Scientific Research.

DISCLAIMER: The Views and Conclusions contained in this report are those of the author and do not reflect the policies and procedures of the University of Texas at Dallas, the United States Government or the Air Force Office of Scientific Research.

INTRODUCTION

Supply chain management (SCM) is the management of a network of interconnected businesses involved in the ultimate provision of product and service packages required by end customers [1]. Supply chain management spans all movement and storage of raw materials, work-in-process inventory, and finished goods from point of origin to point of consumption. Now, this definition of SCM can be adjusted reasonably well to a Data Supply Chain environment. We can define Data Supply Chain management (DSCM) to be the management of a network of interconnected data centers involved in the ultimate provision of a Data Product (or package) required by end customers. These end customers could be the CEO of a corporation or a General in the Air Force. For each situation, the end customer needs the data product either for sale or to help him make decisions.

Another definition for SCM provided by the APICS Dictionary is the "design, planning, execution, control, and monitoring of supply chain activities with the objective of creating net value, building a competitive infrastructure, leveraging worldwide logistics, synchronizing supply with demand, and measuring performance globally." Similarly, we can define DSCM to be the design, planning, execution, control and monitoring of data supply chain activities with the objective of creating net value. Essentially, to develop the data product for the end customer, several data supplies have to arrive at the data assembly plant. The assembly plant manager then puts together the data product for the end customer.

In this paper we will examine various aspects of supply chain management and discuss the analogy to data supply chain management. In both cases, the end customer needs a product. The product has to be assembled from multiple components. Each component has to arrive at the right time at the right place. The component has to be of high quality. In the case of attacks, the component should be able to recover. Each component may have policies enforced on it. These components are then assembled and merged and the end product may also be subject to policies. Essentially, the end product has to be secure and be of high quality.

The organization of this paper is as follows: We will discuss the various aspects of supply chain management in Section 2. The parallels to Data Supply Chain Management will be discussed in Section 3. The technologies that we need for DCSM will be discussed in Section 4. Relating data supply chain to DOD's (Department of Defense) vision for information sharing is discussed in Section 5. Security solutions for supply chain and their impact on data supply chain management are discussed in Section 6. Directions for research are given in Section 7. The paper is concluded in Section 8.

1. SUPPLY CHAIN MANAGEMENT

Numerous papers have been written on supply chain management. An excellent introduction is given in the paper by Ganeshan and Harrison [2]. In this section, we will summarize the key points. The next section will discuss the application of the techniques to data products. Ganeshan states that a supply chain is a network of facilities and distribution options that performs the functions of procurement of materials, transformation of these materials into intermediate and

finished products, and distribution of these products to customers. Furthermore, supply chain exists both in services and manufacturing organizations. The authors also state that it is critical that the multiple players in a chain (e.g., suppliers, consumers) be well-coordinated.

Important concepts in supply chain management include supply chain decisions. The types of decisions include location decisions, production decisions, inventory decisions and transportation decisions. Each type of decisions will include both strategic and operational decisions. Location decisions are the geographic placement of the production facilities, stocking points, and sourcing points. The decisions are determined through optimization. That considers factors such as production costs, taxes and distribution costs. Production decisions include the types of products to produce and the plants used for production. Production scheduling is an important aspect here. Inventory decisions include the management of inventories. For example, strategies such as push vs. pull models are examined here. Transportation decisions include the routes to follow, shipment and also distribution issues.

Note that in his book on supply chain management [3], Hugos also adds a fifth decision and that is “Information”. The authors state the following: *“How much data should be collected and how much information should be shared? Timely and accurate information holds the promise of better coordination and better decision making. With good information, people can make effective decisions about what to produce and how much, about where to locate inventory and how best to transport it.”*

Researchers have developed various modeling approaches for each of the decision types. Due to the large number of data requirements, models usually produce approximate solutions. Modeling approaches are divided into three groups: network design methods, rough cut and simulation. Network design methods determine the location of production, stocking and the paths taken by the products. Rough cut methods give guiding policies for operational decisions and assume a single site approach. Simulation methods are used to analyze the supply chain model.

Supply chain management experts have stated that a supply chain works well when all the players share the risks and costs. Clearly, incentives play a major role and have to be realigned if there is a need. Manatsa and McLaren [4] state that sharing accurate and timely supply and demand information throughout a supply chain can yield significant performance improvements to all members of the supply chain. They argue that there is reluctance in sharing information due to the fact that while the recipient of the information benefits, the provider usually incurs a lot of cost. Furthermore, corporations are nervous in case confidential information is given to the competitors. The authors then propose a principal-agent model to handle risks, incentives and sharing.

2. DATA SUPPLY CHAIN MANAGEMENT

In this section, we will apply supply chain modeling for data production in assured information sharing. Data supply chain is the network of facilities and distribution options that perform the functions of the production of data, transforming the data into intermediary and finished data products, distributing the data products to customers, and sharing the data products among customers. Essentially, we have taken the definition of supply chain management and added one

more step which is to share the data among the customers who request the data from its partners in the chain.

Next let us see how the various types of decisions translate to data supply chain management. Location decision in regular supply chain is the geographic placement of the production facilities, stocking points, etc. Note that in the electronic world we have gone beyond borders. Therefore, while geographic location decisions do not play a role here, location in general is important. From which data source should we get the data? Is it A or B?

Production decisions are important for data supply chain. When a customer needs a data product, it is important to determine what raw data to produce. Therefore, locations and production decisions together determine where to produce and what data to produce. When there are multiple options, optimization theories may be used to determine the best options for the location and the substance.

Inventories for data supply chain are the data sub-products that we need in the production. These inventories need storage facilities. Push vs. pull approaches may also apply here as in the case of regular supply chain management. Should we pull the data (i.e. the inventory) when we need it or should the system push the data periodically?

Transportation decisions will determine how the data is moved from location to location (e.g., optimized routing). Furthermore, at each location, the data may go through transformation processes. For example, different pieces of data may be merged or data may be sanitized. Optimized routing algorithms as well as transformation algorithms may be applied here.

With respect to the modeling approaches, network design methods may be used to determine the location of the data and the stocking and transfer of the data. Rough cut methods may take into consideration the policies on the data and the processes at every stage without paying attention to the location or the distribution. Simulation methods will help us analyze the data supply chain model.

With respect to information sharing, data supply chain is created mainly for the purpose of information sharing among the customers (that is, the partner agencies). However, to carry out effective data supply chain management, information sharing is very important. In this context, the information sharing process has to generate the metadata that is needed to produce the data product. The data product itself is shared among the agencies. Sharing metadata for data supply chain may have risks if confidential data gets into the hands of the adversary. Therefore, theories such as the Principal-Agent theory have to be examined for data supply chains.

3. TECHNOLOGIES FOR DATA SUPPLY CHAIN MANAGEMENT

Our goal is to develop a data supply chain model to develop data products that can be shared among the agencies/coalition partners. That is, each data product is developed according to the rules of a data supply chain model. In order for a successful data chain-based approach, the partners of the supply chain also have to share the information, risks and costs. Furthermore, the incentives have to be aligned every step of the way. This means that the approaches used to

supply chain management have to examine for data supply chain management. In addition, several information management technologies play an important role.

Suppose a customer needs a data product. The first step is to determine who to go to to get the data. This means we need metadata that will guide us in getting the locations of the individuals who possess the raw materials (i.e. the raw data). The raw data will be in data sources. The next step is to determine how to get the data from A to B in the form we need. What are the transformations to the data? What path should the data take? How is the data stored at the intermediate locations? Technologies that we need for this process are (i) integration of heterogeneous data sources (ii) cleaning the data every step of the way (iii) understanding the provenance of the data (iv) enforcing appropriate policies – e.g. is the combined data at a higher classification level than the individual pieces of the data?, and (vi) extracting the data that is needed for the processing of the data at every stage.

Conducting this entire process in real-time is a challenge. Therefore, concepts from the raw data such as email, chats, blogs, web pages and social media pages have to be extracted and linked to form networks. This process has to be carried out continuously so that if and when a customer needs a data product, much of the components are already there. This is similar to using existing raw material for a product than trying to develop new raw material. The linked data also has to be analyzed so that the nuggets are produced for effective knowledge management of a corporation or an agency. Therefore, some of the key technologies include semantic web for representing the vast amount of heterogeneous data, data mining for extracting concepts from the data, network analysis, and knowledge management.

Another challenge is to get the right amount of the right parts at the right time to the consumer. That is, if the parts do not arrive on time, then the supply chain process will be disrupted. Also, if there are too many parts supplied (i.e. too much inventory) then it will be very costly. We have heard about the CISCO situation when the company lost several millions of dollars due to too much inventory. Therefore, we need appropriate inventory management techniques. This is also the situation for data. We need the right data at the right time to go to the right place. If the data does not arrive on time, then there will be a delay introducing the final product and this delay could be not just costly but also deadly. Similarly, if there is too much data, then the consumer has to sort the data and extract only the relevant data to complete the data product.

In summary, here are the parallels between data supply chain and regular supply chain. At the lowest level in the data side, you have raw data such as emails and blogs, in the regular supply chain side, you have the nuts, bolts, cement (in the case of building say a house). At the intermediate level you have the network in the data side which will include the nodes and links extracted from the raw data. At the supply chain side you have the doors, windows, and the foundation among other things. At the finish line at the data side, you have the complete data product which could be the negatives (i.e., knowledge) extracted from the networks. At the regular supply chain side, you have the complete house. At every step there are policies. For the data side, you have confidentiality policies, integrity policies and administrative policies. At the supply chain side you have the regulations and guidelines to building a house.

4. SECURITY FOR DATA SUPPLY CHAIN MANAGEMENT

In a panel at the IEEE Logistics Conference in August 2009, security issues for supply chain management was discussed [5]. The panel was moderated by Dr. Charles Nemfakos, a senior fellow of the RAND Corp., and the panelists were Richard Skinner, inspector general of the Department of Homeland Security; Vice Admiral Michael Loose, deputy chief of naval operations for Fleet Readiness and Logistics; Vice Admiral Mark Harnitchek, deputy commander of the Defense Department's U.S. Transportation Command and Dr. Bhavani Thuraisingham, Louis A. Beecherl, Jr. I Distinguished Professor Computer Science at the University of Texas at Dallas. Some of the challenges discussed were malware detection in supply chain management, access control, ensuring that the mission continues even if the components are corrupted and ensuring that the supplies are moved securely from A to B. We will examine each of the challenges.

Malware detection/prevention is about detecting/preventing malware in the software that operates the components of the chain. The component system may be corrupted and this has to be detected as early as possible. Access control would include controlling access to the databases that maintain information about the various supplies and operations within the chain. Ensuring that the mission continues even if the components may fail involves ensuring that the supply chain is not interrupted even a particular member of the chain may fail. This means there are sufficient backup procedures. Finally, ensuring that the goods moved from A to B are secure would mean that surveillance cameras or another means are installed to deter and or detect terrorists.

Next we will apply the solutions to the data supply chain. Malware may infect not only the systems that produce the data, but also the data products along the chain; this means the malware prevention/detection tools have to work at both the system and data level. Next the database that contains the information about the supplies and inventors has to be protected. Here the solution developed for the regular supply chain may be applied. In the case of mission success, even if the components fail, we have to apply the techniques both for the system and for the data. In the case of the data product, the sub-products may contain bad data. The goal is not only to correct the problem at the sub-product level, but also ensure that even if the data is corrupted, the complete product is not corrupted. This is a challenging problem as we are expecting garbage in but high quality product out. One way to handle this problem is to annotate the data products in the supply chain so that during the assembly level one could determine when, where and how the product was corrupted and apply fixes to the final product. Finally, to handle the corruption of the goods or stealing of the goods during the logistics process, the comparison here is handling corrupted or missing data during the supply process. In this case, appropriate techniques have to be developed to manage incomplete and uncertain data. In addition, replicas are needed for backup.

5. RELATING TO DOD'S VISION FOR INFORMATION SHARING

The DoD's information sharing strategy's goal is to "*Recognize and leverage the Information Sharing Value Chain*". In particular, it is stated that *The Information Sharing Value Chain articulates the "opportunity" of information sharing to support informed decision making, shared situational awareness and improve knowledge at every level of the DoD. The risks*

encountered at each step of the information sharing value chain must be managed to mitigate negative consequences [6]. The strategy goes on to say that “throughout history, the supply of and demand for information triggers the inter-related processes of information collection, processing, analysis, and integration to make informed decisions, increase situational awareness, or improve and manage knowledge. Regardless of the mission domain, community or organization’s unique processes for managing information, the universal Information Sharing Value Chain remains the same – to discover and collect information and continuously add value at each stage to best inform a decision maker. “

The information sharing value chain essentially talks about the Data Supply Chain. An agency requests a finished product (it could be either in the push or pull mode). There are several processes that must work together in finding the raw data, manipulating the raw data and assembling the finished products. All of the partners in the chain must be given sufficient incentive to work together, share the costs and risks and celebrate in the success of the finished product. Policies will guide the processing of the data. The finished product may be shared among a second set of partners (e.g, coalition partners) and a second chain may be formed to construct a second finished product from the first one. A question is, what is the relationship between the two chains? This process continues forever with old or unwanted data being discarded when traveling through the chain.

Without the content of the DoD, one data supply chain could be within the DoD. For example, the Secretary of the Defense needs a complete data product daily to brief partners in other agencies and the President. To assemble this product, the Secretary will have several partners within the DoD who will be members of the supply chain. Once the product is assembled, the Secretary will share his product with his porters at DNI, FBI, and DHS. Possibly a second product will be assembled with data from the multiple partners and the final product will be presented to the President. It may not end here. The President may start a third chain by sharing the product with our allies including the Prime Ministers of UK, Canada and Australia. The results will be fed into the product to be assembled the next day.

With respect to giving sufficient incentives, the DoD states that it must “*Create guidance and incentives within the budgeting and resource allocation process to encourage organizations to share information that promotes informed decision making, improves situational awareness, establishes economies of knowledge, and creates unity of effort.*” We need experiments to determine the quality of the finished product with and without incentives. This way the correlation between incentives and the sharing can be understood.

Information sharing does come with risks. While sharing the right information at the right time at the right place can result in significant benefits. Sharing the wrong information at the wrong time at the wrong place would be disastrous. Therefore, supply chain managers must determine the risks during every step of the chain and mitigate the risks and reduce them to an acceptable level.

6. DIRECTIONS FOR RESEARCH

There are many papers and books on supply chain management. An excellent introduction is provided in the book by Michael Hugos [3]. This book addressed many topics including the following:

- Key concepts in supply chain management
- Supply chain operations: Planning and management
- Supply chain operations: Marketing and Delivery
- Using information technology
- Metrics and measuring supply chain performance
- Supply chain coordination
- Defining supply chain opportunities
- Creating supply chain for competitive advantage
- Promise of real-time supply chain.

Data supply chain management can benefit a great deal by applying the concepts and principles developed for supply chain management. For example, how can data be produced and shared in real-time? How can data supply chain management and information sharing result in a competitive advantage? What are the metrics for measuring data supply chain and information sharing performance? For example, by sharing information can we find tangible benefits? These are questions worth exploring that would benefit the DoD information sharing strategy and implementation a great deal. As part of our AFOSR MURI project [7], we are examining risks and incentives for assured information sharing. Incorporating the investigation of data supply chain management is within the scope of this project.

7. SUMMARY AND CONCLUSION

This paper has introduced the notion of data supply chain management and has drawn parallels between this concept and regular supply chain. It then discussed technologies for data supply chain management and discussed its relationship to DoD's information value chain.

Data supply chain is a never-ending process. The raw data is created by individuals from previous data and new concepts in the form of email, blogs and charts. This data is then manipulated to produce an end product which will then become raw material for a new supply chain. This process will continue forever and will involve multiple chains. It is critical to enforce policies at every step of the way, evaluate the risks and give incentives for the members of a chain to share data. Every partner must share in the risks and costs and celebrate in the success of the end product.

REFERENCES

[1] Harland, C.M. (1996) Supply Chain Management, Purchasing and Supply Management, Logistics, Vertical Integration, Materials Management and Supply Chain Dynamics. In: Slack, N (ed.) Blackwell Encyclopedic Dictionary of Operations Management. UK: Blackwell.

[2] Ram Ganeshan and Terry Harrison, An Introduction to Supply Chain Management, Penn State University, 2002.

[3] M. Hugos, Essentials of Supply Chain Management, Wiley, 2006.

[4] Priscilla R. Manatsa, Tim S. McLaren, and Information Sharing In a Supply Chain: Using Agency Theory to Guide the Design of Incentives, Supply Chain Forum, Volume 9, #1, 2008.

[5] Panel on Security for Logistics, 44th IEEE/SOLE Logistics Conference, August 2009.

[6] Department of Defense Information Sharing Strategy,
http://www.au.af.mil/au/awc/awcgate/dod/info_sharing_strategy.pdf