2011 2nd International Conference on Networking and Information Technology
IPCSIT vol.17(2011) © (2011) IACSIT Press, Singapore

Cost efficient Solutions for Data Flow in a Corporation Modeled with a Complex Network

Versavia Ancusa and Razvan Bogdan

Computer and Software Engineering Department, Faculty of Automation and Computers,
"Politehnica" University of Timisoara, Romania
versavia.ancusa@cs.upt.ro, razvan.bogdan@cs.upt.ro

Abstract. This paper aims at modeling and simulating the data flow in a corporate environment. The goal of this simulation is to study the average values related to that flow. Further on, such an analysis can be employed to track down cost-efficient solutions. A large spectrum of experimental results is discussed. Different examples are offered where such an analysis can be used.

Keywords- complex network, simulation, data flow, cost efficiency

1. Introduction

Estimating cost-efficient solutions is the key to a successful business. As studied by various economists [1] there is a definite link between the corporate social performance, organizational size and financial performance. On the other hand, the computer science has barely recently started to study and implement the modeling of complex social networks [2]. Due to the increase in computation power, more and more intricate models can be simulated and analyzed, thus making it very interesting and economically judicious to implement multifaceted cost analysis scenarios on complex networks models of corporations.

The organization size is a definite factor in studying the economic proficiency of a corporation. However, complex networks theory comes to prove that in an organization, human interaction structure (Figure 1) is relatively scalable. Because this social network can be mathematically described with a fractal, the complexity of this type of network is mathematically precise. Another advantage in mathematically modeling such a network comes from the possibility of approximation with a compound star topology [4].

In characterization of corporate social networks, an important factor is degree correlation. For the most part, social networks seem to be assortative [2], meaning the nodes are grouped together by category. Such grouping makes it easier to simulate and analyze the data. Not only that, but in a corporate environment the grouping mainly corresponds to the hierarchical structure, which is well defined and easy to implement. The variables consist in particular friendships or business relations that none the less, fall in the complex network pattern.

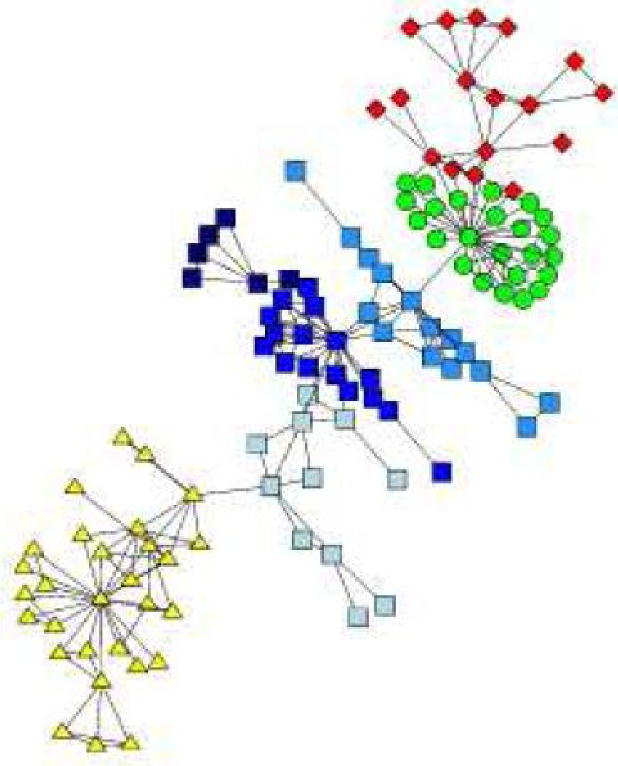

Figure 1. Human interaction structure [3]

In this paper, we shall model, using a star topology, the data flow in a corporate environment and study the average values related to that flow. This analysis can be used in order to track down cost-efficient solutions. For example: if there is heavy dataflow between two nodes, fault-tolerant electronic communication can replace the hardcopy, saving not only money, but also time required by printing that hardcopy. The environmental impact can also be minimized with a similar analysis.

This paper is structured as follows: section II presents the simulation and the cost analysis of the proposed problem. Following next, we will present the review of the results. In section IV some future work directions are presented.

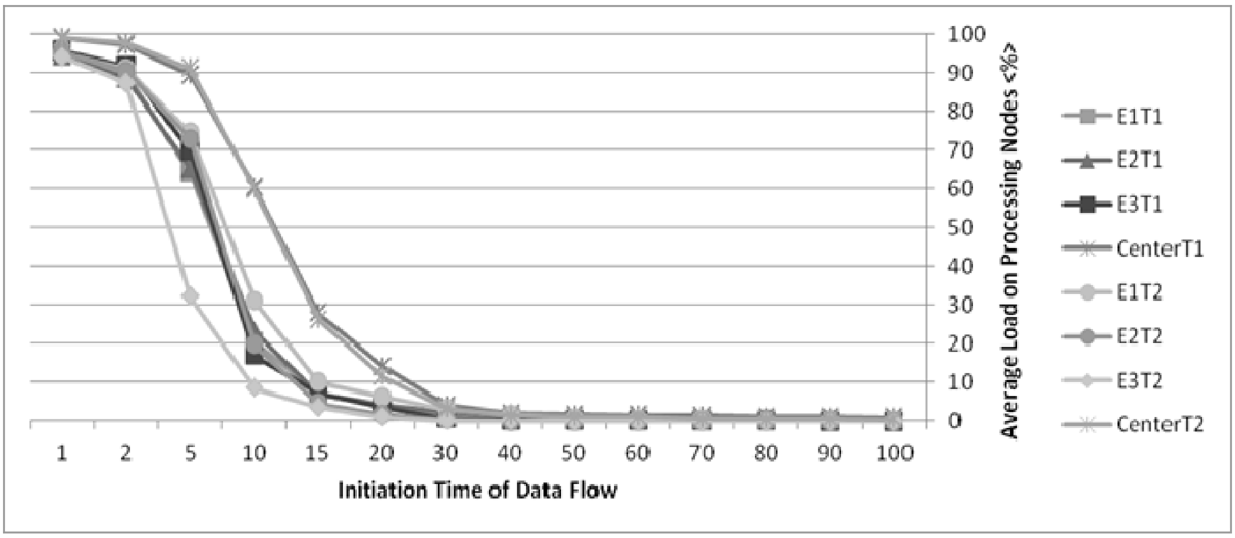

Figure 2. Percent of average load on a processing node

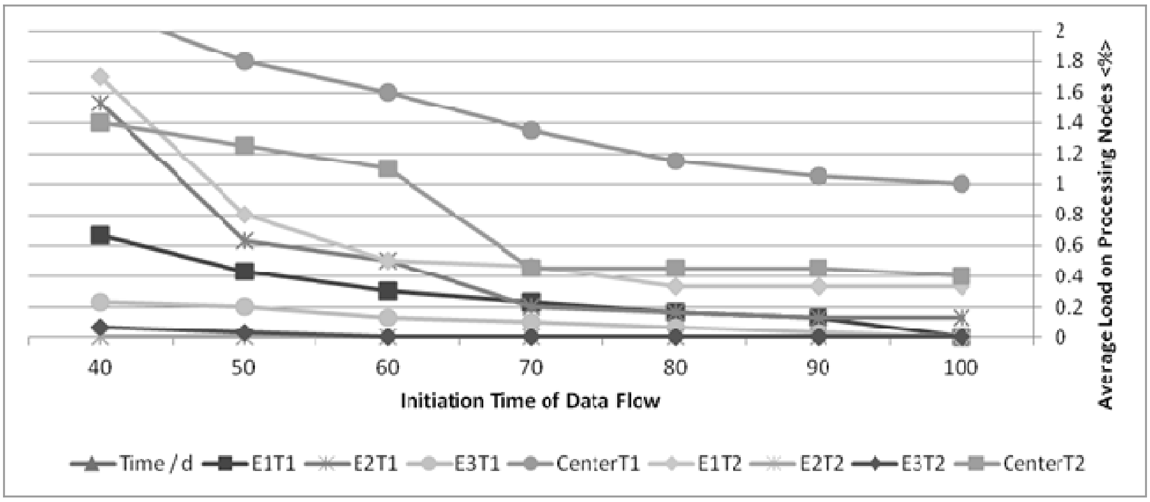

Figure 3. Non-congested data flow

2. Simulation

We chose to implement a simple two-team model by using AweSIM [5]. Each team has three members and a communication center. Each member and the communication center generate data, with a random exponential distribution. The data is gathered in a local node by each team member and each center, and after that are processed. If the processing entity cannot accept any more data, the team members send data to the central communication center. The communication center for team 1 can forward overflowing data to the communication center for team 2. The communication center for team 2 will service as many requests as possible and when overflowed will out-source to an outside center. The out-sourcing we can count as failure, because as easy it is to model that outside center, that node can also be a lost data node. All the initiation nodes are modeled identically; only the processing nodes are somewhat different (the team has the same type of processing, different from the data centers).

As stated in [1], the main measure in order to analyze various aspects of a corporation revolves around different mean values, therefore we shall study the average loads and wait times in the processing nodes and average line utilization. One note should be made regarding the time used in the following figures. In order to maintain the generality of the model, we shall compute a relative time as a multiple of the minimum data entity path in the complex network.

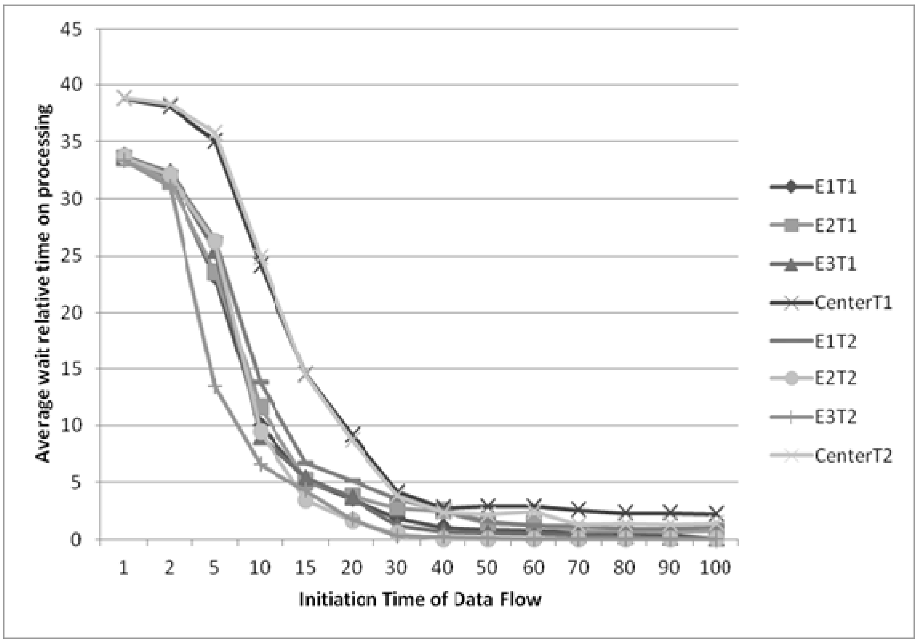

Figure 4. The average wait time in the processing queue

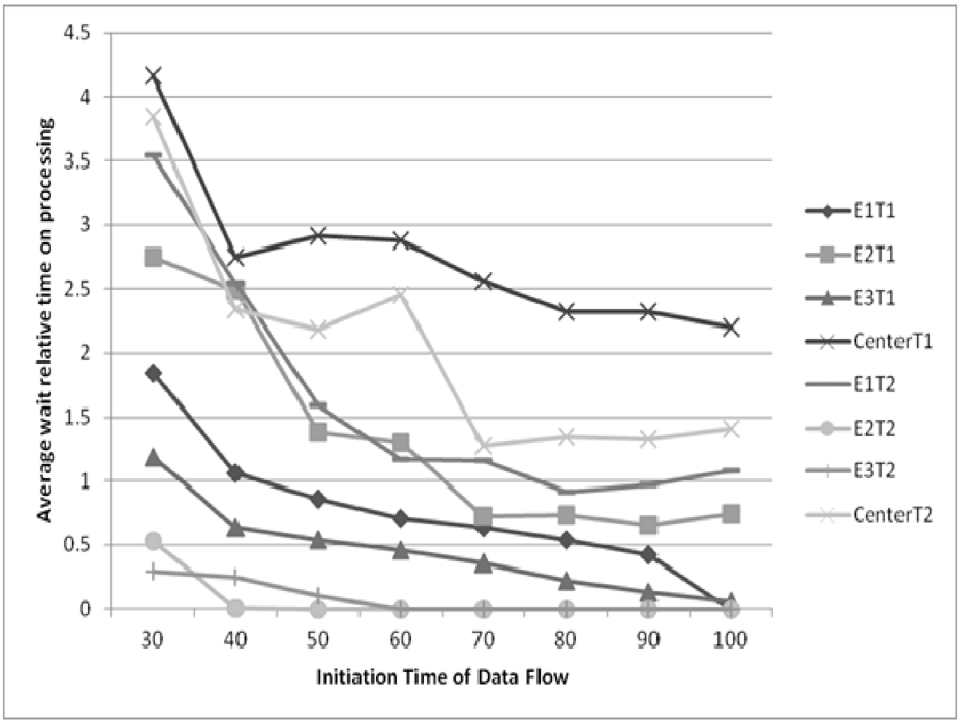

Figure 5. Waiting times

The real time value can thus be easily computed, and applied in any type and size of corporation.

Figure 2 presents the percent of average load on a processing node. If data is initiated very frequently from all the nodes (for example, in a printing house), the load increases rapidly, attaining congestion status if the initiation time is 5 times the time necessary to process a data. As expected, the centers congest before the other nodes, however the difference is less than 15% load when the congestion starts.

Looking at the non-congested data flow (Figure 3), team 1's center has a slightly higher load than team 2's data

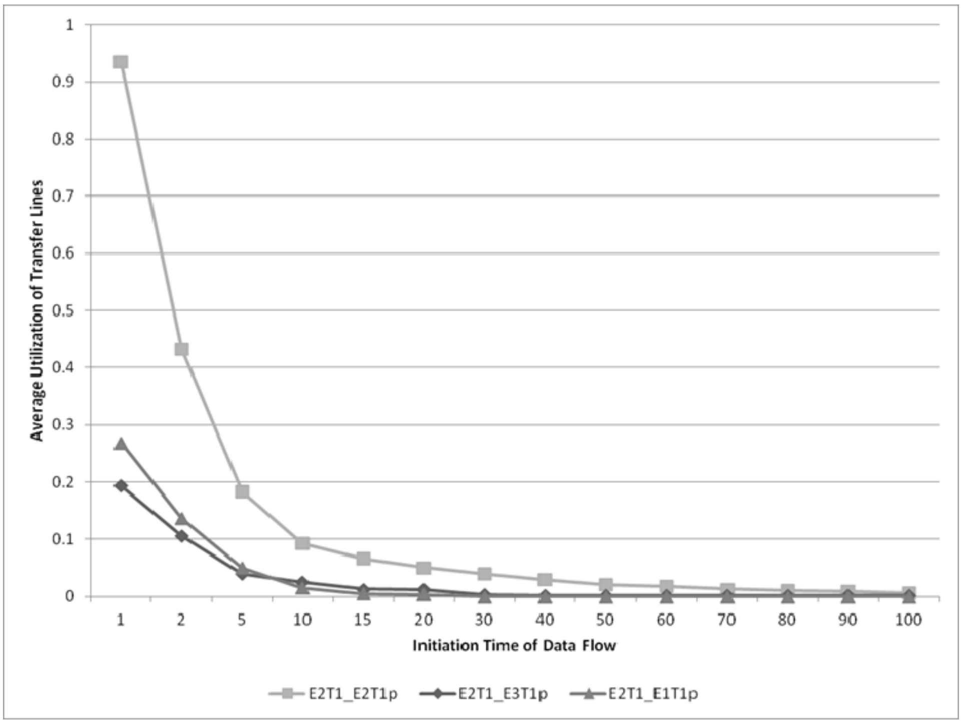

Figure 6. The average utilization of direct transfer lines

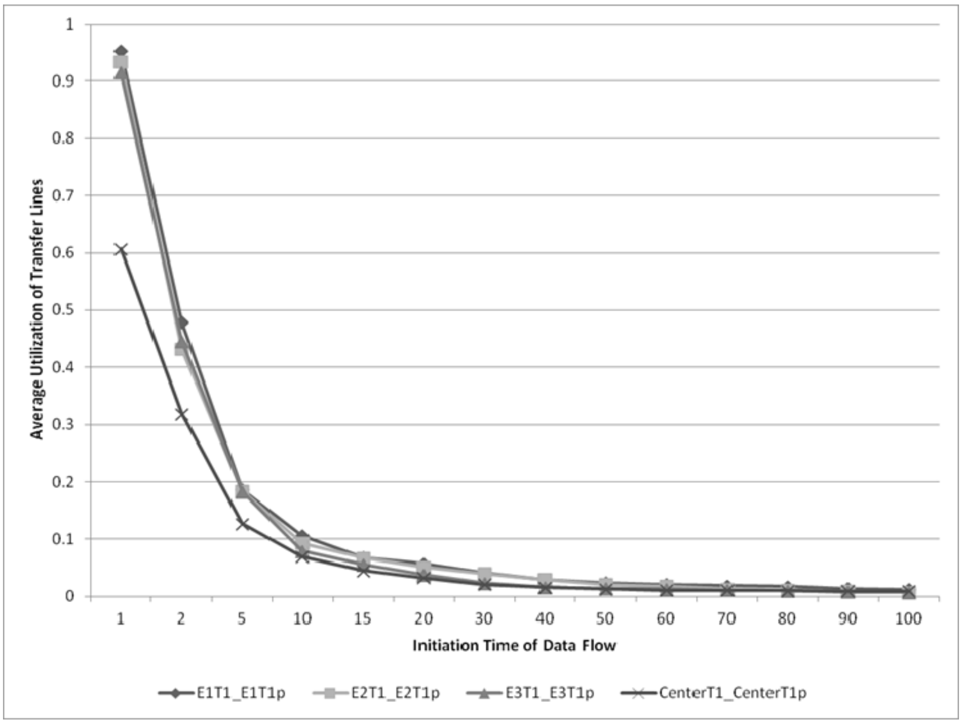

Figure 7. The forwarding routes inside a team

center, due to the possibility of outsourcing for team 2. The team members' behavior are similar, though it should be noted there is, to some extend, a member in each team with a higher load than the others. A cost-efficient setup would identify that node and make sure that it has an increased processing capability, therefore delaying the congestion.

Figure 4 describes the average waiting time in the processing queue. For lower values of initiation time (a higher frequency of data flow generation), the waiting time increases significantly. The main data processing centers for each team clearly have a higher waiting period, because all overflowing jobs are referred to them.

The difference between the waiting time for the data center and normal processing node is not so clear unless there is congestion. As shown in Figure 5, before congestion all the waiting times are not significantly differentiated.

The average utilization of direct transfer lines between an initiator and processing node is presented in Figure 6. The main processing center has a lower utilization, compared with the rest of the team, which reaches full utilization of the lines for initiation time equal to an individual processing task. This means that the processing center still has untapped potential, even when there is congestion, suggesting a cost efficient measure like reducing bandwidth on the data center or maybe maintaining the same bandwidth, while increasing the forwarding of jobs from the low-level nodes.

Looking at the forwarding routes inside a team (Figure 7), the further away a node is from another, the average utilization of that route decreases. Cost-efficient solutions may promote forwarding data to other same-level nodes, if the utilization of one's direct processing route is higher than 40%.

3. Conclusions and Future Work

This paper presents a pertinent way to make a cost-performance analysis, as opposed to the general empirical way to approach this problem. Using computer simulation on a general model, shows the viability in applying this method for a real company.

Future work includes adding different types of nodes, different types of processing, a prioritization schema for processing and so on.

4. References

[1] Peter A. Stanwick, Sarah D. Stanwick, “The Relationship Between Corporate Social Performance, and Organizational Size, Financial Performance, and Environmental Performance: An Empirical Examination,” Journal of Business Ethics, vol. 17, no. 2, pp. 195-204, DOI: 10.1023/A:1005784421547, January 1998.

[2] M. E. J. Newman, “The structure and function of complex networks,” SIAM Review, no. 45, pp. 167-256, 2003.

[3] M. Girvan, M. E. J. Newman, “Community structure in social and biological networks,” Proc. Natl. Acad. Sci., 2002.

[4] Juyong Park, M. E. J. Newman, “Solution of the 2-star model of a network,” Physical Review E, vol. 70, Issue 6, id. 066146, 2004.

[5] Alan B. Pritsker, Jean J. O'Reilly, “AWESIM: the integrated simulation system.” Proc. of the 30th conference on Winter simulation (WSC '98), D. J. Medeiros, Edward F. Watson, John S. Carson, and Mani S. Manivannan (Eds.), IEEE Computer Society Press, Los Alamitos, CA, USA, 249-256, 1998.