

Studying the behavior of psiX under different degrees of network churn^{*}

Dipali Pal, Priyanka Trehan, Vamshidhar Reddy

UMKC, CSEE Department

Overview

We evaluate the behavior of psiX system under different degrees of network churn. A detail study of publication and query is done by measuring data read, data written, number of HOPS, time taken and documents located as a function of percentage of churn introduced into the system.

Experimental setup and data collection logistics

We performed two experiments:

(1) Behavior of psiX system without network churn:

This experiment is conducted on psiX system without network churn to fundamentally understand the psiX performance with increase in number of publishing documents within the network (without churn). Firstly, a network of 40 nodes running on psiX is formed. This network can be formed by starting nodes with same boot host. Subsequent to that, documents are published on to network of these 40 nodes. Consequently, the following information is collected for 203, 394 and 798 number of published documents.

1. Total Time to publish each document (in Sec.)
2. Average Amount of Data Read for each document (in bytes)
3. Amount of Data Written for each document (in bytes)
4. Number of HOPS for all the set of documents.

(2) Behavior of the psiX system with churn:

This experiment is conducted on psiX system under network churn to understand the psiX performance with increase in percentage of churn in the network. Similar to the previous experiment, a network of 40 nodes running on psiX is formed. Subsequent to that, 750^{*} documents are published on to network of these 40 nodes and queries (related to the published documents) are submitted to locate the documents under 0, 5 and 10 percentage of churn in the network with a churn rate of 60 seconds (where a set of node leave and leave at a regular interval).

^{*} Project completed as a part of CS5590 LD course during Spring 2009.

In Documents Publishing part, the following information is collected for 0, 5 & 10 % of churn in the network (of 40 nodes) for 750 published documents.

1. Average Total Time for publishing (in Sec.) .
2. Average amount of Data Read (in bytes) .
3. Average amount of Data Written (in bytes) .
4. Average number of HOPS for each document .

In Queries Submitting part, the following information is collected for 0, 5 & 10 % of churn in the network (of 40 nodes) for 750 published documents.

1. Average Query Time for each query(Sec.)
2. Average amount of Data Read (bytes)
3. Average number of Documents found
4. Average. number of HOPS

Results

Fig. 1 shows the results of publication process without any churn present in the system. In general, the figure suggests an increase in data read, data written, number of HOPS and total time with number of document published. The rate of increase in each instance is higher as we increase the number of documents from 394 to 798. Given the fact that we use a fixed number of nodes in each step, namely, number of documents published = 203, 394 and 798, this increase has a significant contribution from load distribution within the peers. A systematic of load management within individual peers, which has not been studied quantitatively within this project, may have played a role in this case but that should not alter the general trend shown in these plots.

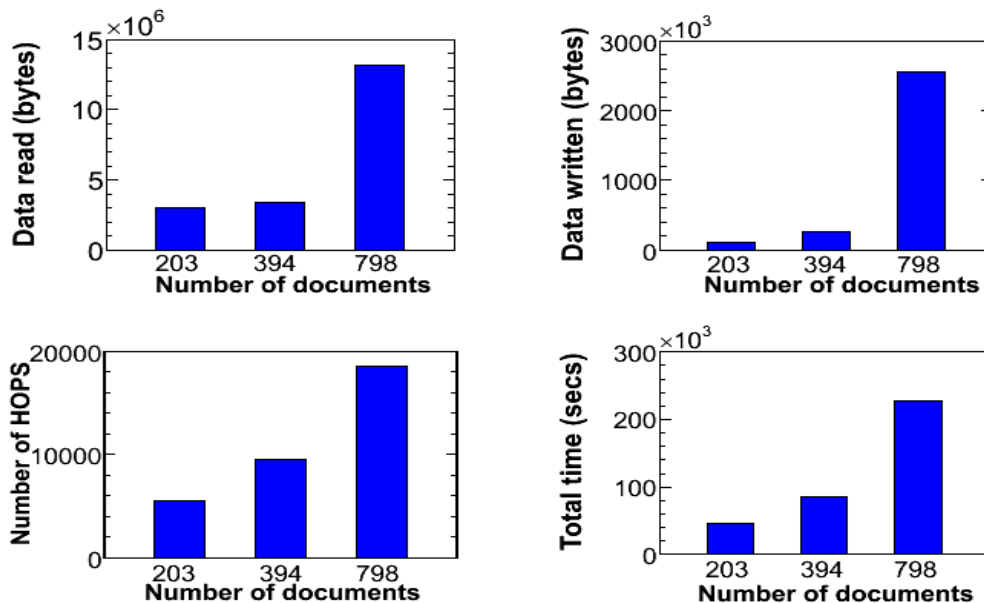


Fig 1: Results of publication performance without churn

Injection of churn within this process, leads us to a rich set of observations on the system performance. Fig. 2 shows evolution of average data read, data written, average number of HOPS and average time taken as a function of percentage of churn injected to the system. The averaging is done over the number of publication processes. First, we observe a systematic increase in average data read with percentage of churn present in the system. The average data written, however, shows some different picture as we see in the upper right hand panel of Fig 2. As we introduce churn to the system, the average amount of data written increases sharply, but after that if we further increase the degree of churn, the data written drops but still remains higher than the system with no churn. The second data bar in that plot might have been contributed by pedestal noise within the system. A systematic study of the process at different time using the same set of nodes is necessary to understand this data point. The increase in number of HOPS with churn percentage is showing a systematic increase. The pattern of time, however, shows a sharp increase as we introduce 5% churn to the system from the system without any churn. The average time is then increased by a small amount if we further increase the churn to 10%. This pattern needs to be checked with larger amount of churn.

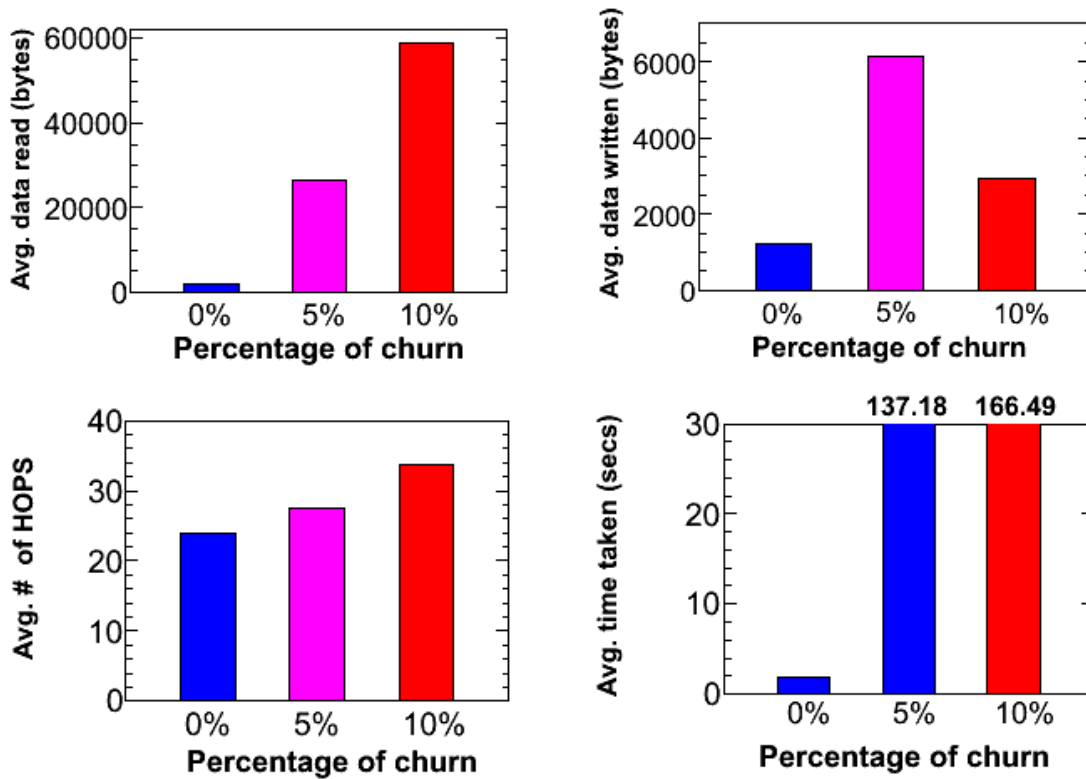


Fig 2: Publication process under different degrees of churn.

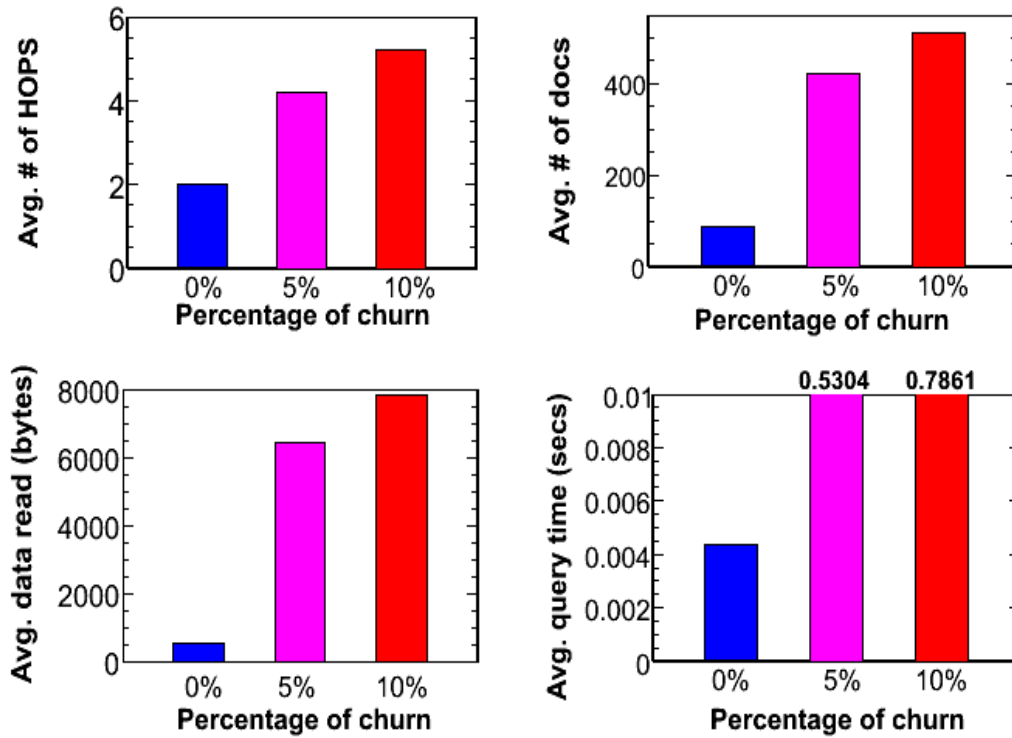


Fig 3: Query process to locate document under different degrees of churn.

Summary and conclusion

In this project, we studied the evolution of the psiX system under different degrees of churn. We found that the average of data read, number of HOPS and time taken for publication process increases with degree of churn within the system. The query process also shows a systematic increase in average number of HOPS average number of documents located, average data read and average query time as we introduce more churn to the system.