ComEX

IEICE Communications Express, Vol.4, No.1, 14–19

SFB: a scalable method for handling range queries on Skip Graphs

Ryohei Banno[a)], **Tomoyuki Fujino,**
Susumu Takeuchi, and Michiharu Takemoto

NTT Network Innovation Laboratories, NTT Corporation,
Midori-cho, Musashino-shi, Tokyo 180–8585, Japan

a) *banno.ryohei@lab.ntt.co.jp*

Abstract: Skip Graph is a promising candidate algorithm for large scale distributed systems. The principal feature is range query functionality, but Skip Graph does not have a definite method of multicasting inside ranges designated by query issuers. Even though several simple ways can be considered, they are inefficient regarding the latency or traffic volume. In this letter, we first introduce Multi-Range Forwarding (MRF) used in Multi-key Skip Graph. MRF can be used even in normal Skip Graph, and is efficient compared to the simple ways. Second, we propose a method named Split-Forward Broadcasting (SFB). We analytically evaluate SFB and explain that SFB can roughly halve the average number of hops of MRF.

Keywords: Skip Graph, range query, structured overlay networks

Classification: Network

References

[1] J. Aspnes and G. Shah, "Skip graphs," *ACM Trans. Algorithms*, vol. 3, no. 4, 37, November 2007. DOI:10.1145/1290672.1290674

[2] http://www.piax.org/en (accessed September 7, 2014).

[3] A. G. Beltran, P. Milligan, and P. Sage, "Range queries over skip tree graphs," *Comput. Commun.*, vol. 31, no. 2, pp. 358–374, February 2008. DOI:10.1016/j.comcom.2007.08.003

[4] Y. Konishi, M. Yoshida, S. Takeuchi, Y. Teranishi, K. Harumoto, and S. Shimojo, "An extension of skip graph to store multiple keys on single node," *J. Inf. Process. Soc. Japan*, vol. 49, no. 9, pp. 3223–3233, September 2008. (in Japanese)

[5] A. G. Beltran, P. Sage, and P. Milligan, "Skip tree graph: a distributed and balanced search tree for peer-to-peer networks," Proc. 12th International Conference on Communications, Glasgow, Scotland, pp. 1881–1886, June 2007. DOI:10.1109/ICC.2007.313

1 Introduction

Skip Graph [1] is an algorithm of structured overlay networks, which can be used to construct distributed systems [2]. The distinct feature of Skip Graph is supporting

DOI: 10.1587/comex.4.14
Received October 22, 2014
Accepted December 16, 2014
Published January 23, 2015

range queries. In Skip Graph, each node has a key and can issue a query by specifying a target range in the key space. Issued queries are delivered to nodes whose keys are included in the range.

Skip Graph composes a multiplex structure of a skip list. This contributes to the suppression of both the size of routing tables and the path length of forwarding queries. An issued range query is forwarded from the start node towards the specified range with expected $O(\log N)$ hops, where N is the number of nodes.

Although Skip Graph enables a query to be delivered to one of the nodes in the specified range efficiently, it does not have definite methods of delivering the query inside the range from that node. Several simple ways, e.g., sequential forwarding, can be considered as mentioned in [3], but they are inefficient from the viewpoint of the latency or traffic volume.

The contributions of this letter are threefold: First, we introduce that Multi-Range Forwarding (MRF) used in Multi-key Skip Graph [4] can be used even in normal Skip Graph, and is efficient compared to the simple ways. Second, we propose a novel algorithm named Split-Forward Broadcasting (SFB), which improves the average number of hops of MRF. Third, we analytically compare the essential characteristic of SFB with MRF.

2 Related work

Beltran et al. [3] have compared some methods of handling range queries with respect to the average number of messages and hops. The compared methods are as follows:

Sequential

Queries are forwarded along the doubly linked list at level 0, until the upper or lower bound of the range is found.

Broadcasting w/o memory

Each node within the range forwards the received queries to all its neighbors within the range.

Broadcasting w/ memory

It is an improvement in the number of messages of the broadcasting w/o memory method. Each message stores the list of nodes that have been visited so that it is avoided that nodes within the range receive the query several times.

Tree-based

Queries are forwarded by using links which are peculiar to Skip Tree Graph [5].

The tree-based method assumes that there are additional links which are defined in the algorithm of Skip Tree Graph. We consider methods of handling range queries on normal Skip Graph for versatility, so this method is outside the scope of this letter.

The sequential method takes expected $O(\log N + N_R)$ messages and hops, where N_R is the number of nodes within the target range R. On the other hand, both broadcasting methods require $O(\log N + N_R \log N_R)$ messages and $O(\log N)$ hops. Though the sequential method outperforms the broadcasting methods with respect to the number of messages, it requires a larger number of hops. That is,

DOI: 10.1587/comex.4.14
Received October 22, 2014
Accepted December 16, 2014
Published January 23, 2015

suppressing both the number of messages and hops by using these simple ways involves difficulties.

3 Applicability of Multi-Range Forwarding

For overcoming the above problem, we first introduce Multi-Range Forwarding (MRF) which is a routing mechanism originally used in Multi-key Skip Graph (MKSG) [4]. MKSG is an expansion of Skip Graph for enabling nodes to possess multiple keys, but we show that MRF can also be used in the normal Skip Graph. By using MRF in normal Skip Graph, a query with its target range R is forwarded as follows:

> When a node whose key is within R receives the query, the node divides R into subranges by its key. The query is duplicated and forwarded to neighbors, which are connected to the node at the highest level among neighbors placed within R, with each subrange attached instead of R.

For example, in the lower half of Fig. 1, when a node A whose key is 10 receives a query of target range $0 \leq key \leq 50$, the range is divided into two subranges: $0 \leq key < 10$ and $10 < key \leq 50$. If A has a right (i.e., larger side) neighbor E which has a key 36 and which is linked to A at the highest level among neighbors within $10 < key \leq 50$, A forwards the query with the subrange $10 < key \leq 50$ to E. E also divides the subrange into $10 < key < 36$ and $36 < key \leq 50$, and forwards them to corresponding neighbors. The query is recursively forwarded, and is finally expired when there are no more nodes to receive it within the range.

With these rules, each node receives the same query only once. Therefore, MRF requires only $O(\log N + N_R)$ messages and $O(\log N)$ hops, and has both strong points of the sequential method and the broadcasting methods.

4 Proposition of Split-Forward Broadcasting

We propose a method named Split-Forward Broadcasting (SFB), which can reduce the average number of hops of MRF. The difference of forwarding processes of queries between SFB and MRF is depicted in Fig. 1.

We explain the algorithm of SFB by using the example shown in the upper half of Fig. 1. When node A receives a query with its target range $R = \{0 \leq key \leq 50\}$, node A searches for neighbors which are connected to node A at the highest level among neighbors placed within R, one on each side. Regarding the right side, node E is the corresponding neighbor at level 2. Then node A divides R into subranges by the key of node E: $R_A = \{0 \leq key < 36\}$ and $R_E = \{36 \leq key < 50\}$. R_E is attached to the query and forwarded to node E from node A, while R_A is still possessed by node A.

Next, node A searches for another neighbor on its right side in the same way as mentioned above, but at levels lower than the level at which node A forwarded the query to node E. As a result, node A chooses node C at level 1. Then node A divides R_A into subranges by the key of node C: $R_{AA} = \{0 \leq key < 27\}$ and $R_{AC} = \{27 \leq key < 36\}$. R_{AC} is attached to the query and forwarded to node C from node A, while R_{AA} is still possessed by node A.

DOI: 10.1587/comex.4.14
Received October 22, 2014
Accepted December 16, 2014
Published January 23, 2015

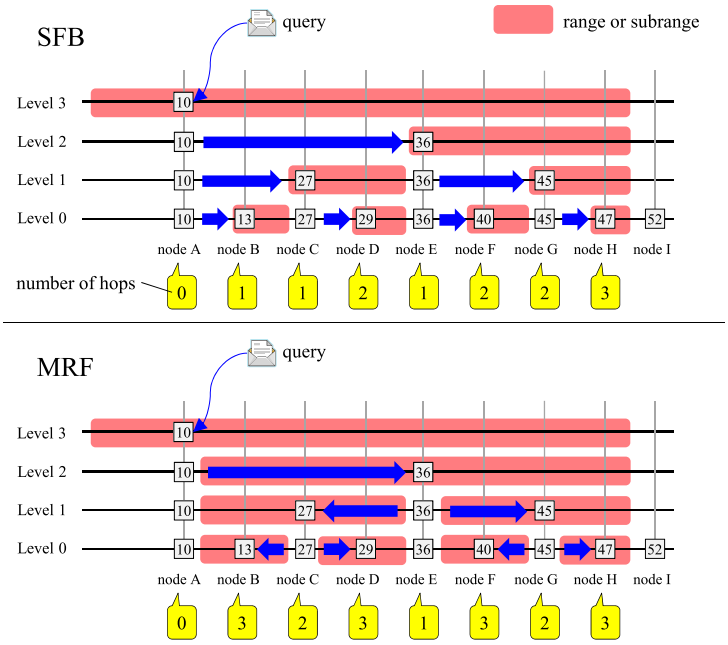


Fig. 1. Difference of multicasting algorithms between SFB and MRF.

Similarly, node A forwards the query to node B at level 0. Node A also executes such process on its left side. Every node within R except for node A will receive the query and execute the same process on the opposite side of the forwarder of the query. For example, node E receives the query from node A on its left side, and forwards the query to the corresponding nodes on its right side. (In Fig. 1, node G and F.)

5 Analytical evaluation

5.1 Qualitative comparison

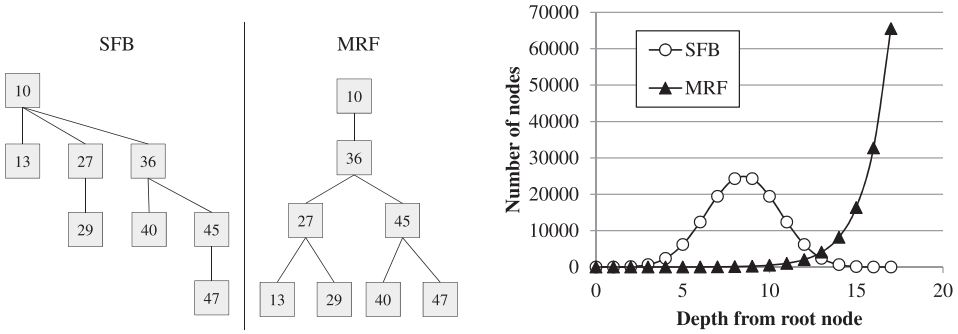
In SFB, each node within the target range receives the same query only once. Hence, SFB requires expected $O(\log N + N_R)$ messages. A query is forwarded from the issuer to one of the nodes within the range with expected $O(\log N)$ hops, and subsequently it is forwarded from the node to every node within the range with expected $O(\log N_R)$ hops. Thus, SFB requires expected $O(\log N + \log N_R) = O(\log N * N_R) = O(\log N)$ hops, given that $N \gg N_R$.

5.2 Difference of tree structures

The expected performances described in the preceding section are same as MRF, but SFB can reduce the actual average number of hops of MRF. Indeed, in Fig. 1, the number of hops from node A to each node within R which is depicted below the node name is relatively smaller than that of MRF. This is caused by the difference of structure of multicasting trees as shown in Fig. 2a.

In SFB, each node which receives a query forwards it to all of possible neighbors within the specified range or subrange. In contrast, in MRF, each node forwards it to at most two neighbors. As a result, SFB composes an unbalanced tree as shown in the left side of Fig. 2a, while MRF composes a balanced binary tree as shown in the right side.

DOI: 10.1587/comex.4.14
Received October 22, 2014
Accepted December 16, 2014
Published January 23, 2015



(a) Difference of tree structures. (b) Number of nodes at each depth.

Fig. 2. Comparison of multicasting trees between SFB and MRF.

To compare analytically, we use following assumptions in this section:

- Skip Graph is composed with ideal membership vector, namely each list at each level consists of a row of evenly spaced nodes on the basis of the number of nodes.
- N_R is exponentiation of 2.
- The leftmost node is the first receiver of the query within the target range.

With these assumptions, the followings can be said: the number of nodes at each depth of the tree of SFB is the same as binomial coefficient, and that of MRF is exponentiation of 2. For example, in Fig. 2a, the number of SFB is 1, 3, 3, 1, and that of MRF is (1,) 1, 2, 4.

When $N_R = 131,072$, the difference of the number of nodes at each depth of the trees is shown in Fig. 2b. Regarding MRF, the number of nodes whose depth is deepest 17 is the largest. On the other hand, the SFB's tree has the largest number of nodes at the depth 8 and 9. This difference makes the superiority of SFB regarding the average number of hops.

5.3 Comparison of average number of hops

From above assumptions, the average number of hops of SFB H_{SFB} can be calculated as below:

$$H_{SFB} = \frac{1}{N_R} \sum_{k=0}^{\log N_R} \{ {}_{\log N_R}\mathrm{C}_k \cdot k \}$$
$$= \frac{\log N_R}{2}$$

For MRF, the average number of hops H_{MRF} can be calculated as below:

$$H_{MRF} = \frac{1}{N_R} \sum_{k=0}^{\log N_R - 1} \{ (k+1) \cdot 2^k \}$$
$$= \log N_R - 1 + \frac{1}{N_R}$$

Fig. 3 illustrates values of H_{SFB} and H_{MRF}, where the horizontal axis represents N_R. From these, if N_R is enough large, it is clear that the average number of hops of SFB is approximately half of that of MRF.

DOI: 10.1587/comex.4.14
Received October 22, 2014
Accepted December 16, 2014
Published January 23, 2015

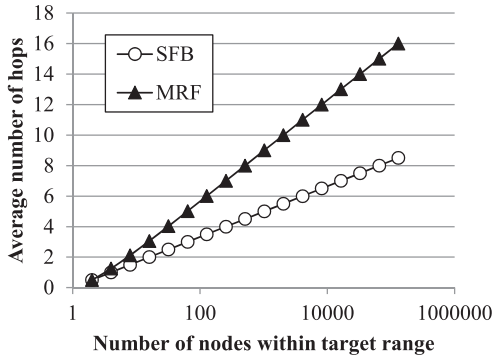


Fig. 3. Average number of hops

6 Conclusion

In this letter, we have proposed SFB which is a method for handling range queries on Skip Graphs. SFB requires only the smaller number of hops and messages compared to the sequential method and the broadcasting methods. In addition, SFB can improve the average number of hops of MRF. This is not only effective for reducing the average latency, but also for improving the churn tolerance because the reduction of hops lowers the probability of message loss by nodes' disappearing.

Future work includes combining SFB and MRF to allow selecting suitable one depending on the situation, because MRF still has an advantage regarding fairness of forwarding load. We also plan to implement SFB and evaluate it by the actual network environment.

DOI: 10.1587/comex.4.14
Received October 22, 2014
Accepted December 16, 2014
Published January 23, 2015