SCAN: A Small-World Structured P2P Overlay for Multi-Dimensional Queries

Xiaoping Sun
China Knowledge Grid Research Group
Institute of Computing Technology
Graduate School of Chinese Academy of Sciences, China
sunxp@kg.ict.ac.cn

ABSTRACT
This paper presents a structured P2P overlay SCAN that augments CAN overlay with long links based on Kleinberg's small-world model in a *d*-dimensional Cartesian space. The construction of long links does not require the estimate of network size. Queries in multi-dimensional data space can achieve $O(\log n)$ hops by equipping each node with $O(\log n)$ long links and $O(d)$ short links.

Categories and Subject Descriptors
H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval – Search process.

General Terms
Algorithms, Design, Experimentation.

Keywords
P2P, small-world, multi-dimensional queries.

1. INTRODUCTION
Information retrieval in large-scale distributed environments often involves multi-dimensional data management and queries. CAN overlay supports data partition and query in a *d*-dimensional Cartesian space [4]. It achieves $O(dn^{1/d})$ query hops. This paper introduces SCAN that builds long links in CAN overlays based on Kleinberg's small-world model [2]. It can achieve $O(\log n)$ hops by equipping each node with $O(\log n)$ long links. Compared with previous small-world solution Symphony [3], SCAN approximates Kleinberg's small-world network in multi-dimensional data space without requiring estimate of network size. eCAN also achieves $O(\log n)$ hops [6] by building express ways in CAN overlay. Long link construction depends on the joining process of nodes. In [1], small-world long links are built in Delaunay-graph-based networks. It uses a piggy-backing method in node joining process to add long links. Long links in SCAN can be built during or after the construction of the underlying CAN.

2. ARCHITECTURE OF SCAN
In a *d*-dimensional SCAN, each node is identified by a vector $v<x_1, x_2, ..., x_d>$. x_i is drawn from a real interval $R = [0, H]$ ($H > 1$). The first node holds the complete *d*-dimensional Space R^d. Forthcoming joining nodes split the zones of existing nodes in half along one dimension in a cyclic way. Data objects are identified by vector IDs drawn from R^d and are stored at the node the vector IDs of data objects fall in. To uniformly partition the *d*-dimensional space, a joining node first draws a random ID $v<x_1, x_2, ..., x_d>$, where x_i follows the uniform distribution in $[0, H]$. Then, the node locates the existing node that holds this random ID and split it in one dimension. Nodes use the central point of the range they hold as their node IDs. Each node maintains $O(d)$ short links to their neighboring nodes. Long links are added for nodes in a small-world way to speed up queries. Figure 1 shows a partial view of a 2-*d* SCAN topology.

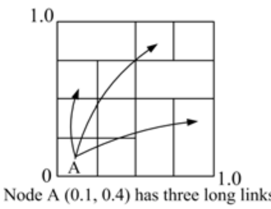

Figure 1. Topology of SCAN.

2.1 Building long links in SCAN
In space R^d, we define the Manhattan distance between two points $v<x_1, x_2, ..., x_d>$ and $u<y_1, y_2, ..., y_d>$ as $d(v,u)=\sum_{i=1}^{d} d_i(x_i, y_i)$. $d_i(x_i, y_i) = \min\{\mathrm{abs}(x_i - y_i), H - \mathrm{abs}(x_i - y_i)\}$ is the coordinate distance in the *i*th dimension. The maximum Manhattan distance $L_{\max}$ is $dH / 2$ because in each dimension the maximum coordinate distance is $H / 2$.

To build long links, a node $v<x_1, x_2, ..., x_d>$ first draws K real numbers $r_1, r_2, ...,$ and r_K following the harmonic distribution in real interval $[0, L_{max}]$. Then, for each r_i, a vector point $l_i<y_1, y_2, ..., y_d>$ is generated as a seed ID at distance r_i from v. Finally, node v locates node v_i that is responsible for l_i and connects v_i as a long link.

Since we do not know the network size, we set $K = \lfloor C\log_2 N \rfloor$ where N is a predefined large integer satisfying $N >> dn^{1/d}$ and $C \geq 1$ is a predefined constant integer. $r_1, r_2, ...,$ and r_K are produced by a harmonic distribution generator in $[0, L_{max}]$. $r_i = L_{max} / 2^x$, where x is a real number randomly drawn from the real interval $[0, \log_2 N]$ for $i = 1, 2, ...,$ and K.

Given a distance r_i from v, there are multiple candidate points for l_i. We randomly generate a real vector $\tau_i<s_1, s_2, ..., s_d>$ so that point $l_i<x_1 + s_1, x_2 + s_2, ..., x_d + s_d>$ is at distance r_i from v (plus is in wrap mode in $[0, H]$). We iteratively generate s_k by regarding the remainder coordinates $s_{k+1}, s_{k+2}, ..., s_d$ as one coordinate. Initially, let $M_1 = L_{max}$, $D_1 = r_i$, and $\delta = L_{max} / d$. To get s_k, following steps are repeated for $k = 1, 2, ...,$ and d:

(STEP 1): If $M_k \leq \delta$, let $s_k = M_k$ and return;

(STEP 2): $I = [0, \delta] \cap [D_k - M_k + \delta, D_k]$;

WWW 2007, May 8–12, 2007, Banff, Alberta, Canada.
ACM 978-1-59593-654-7/07/0005.

(STEP 3): Get a random real number from I as s_k;

(STEP 4): $M_{k+1} = M_k - \delta$; $D_{k+1} = D_k - s_k$.

Long links from v can approach a remote node in two directions along one dimension. We randomly assign s_k a positive or negative sign with probability 1/2. Then l_i<y_1, y_2, ..., y_d> is obtained by $y_k = x_k + s_k$ (k = 1, 2, ..., d and plus is in wrap mode in [0, H]). Node v locates the remote node v_i that holds l_i. After finding v_i, v inserts it into the routing table.

Although $K = \lfloor C\log_2 N \rfloor$ is larger than log n, many seed IDs are actually located in the same node and the expected number of distinct long links is $O(C\log_2 dn^{1/d})$. It is the harmonic distribution of r_i in [0, L_{max}] that makes the long links form a small-world overlay.

2.2 Query routing in SCAN

The size of ranges should be considered in a greedy routing process. Let Z_v = <z_1, z_2, ..., z_d> be the d-dimensional range that the remote node v currently holds, where $z_i = [zx_i, zy_i]$ is the real interval in the ith dimension that v occupies. The range Manhattan distance from node v to the target point t<y_1, y_2, ..., y_d> is $d_r(v,t) = \sum_{i=1}^{d} q_i(z_i, y_i)$. $q_i(z_i, y_i)$ is the dimensional range distance in the ith dimension. $q_i(z_i, y_i) = 0$ if $y_i \in z_i$. Else $q_i(z_i, y_i)$ = min{$d_i(zx_i, y_i)$, $d_i(zy_i, y_i)$}. In each hop, node selects from its links the one with the shortest range Manhattan distance to the target point as the next hop. When the node at distance zero to the target point is reached, the target point is located.

When $C = 2^d$, the expected routing hops is bounded by $O(\log_2 dn^{1/d})$ because each long link can help reduce the distance by half with probability $1/2^d$. When d is large, building $2^d\log_2 N$ long links is prohibitive. Fortunately, when $d > \log_2 n$, $O(\log n)$ query hops can be achieved without using long links. Moreover, we can use $K = 4\log_2 N$ seed IDs to build long links and achieve $O(\log_2 2n^{1/2})$ query hops in most cases as long as $n > (d/2)^{(1/2-1/d)}$, i.e., $dn^{1/d} < 2n^{1/2}$. It is because that when $dn^{1/d} < 2n^{1/2}$, d-dimensional CAN overlays of size n have shorter longest query hops than 2-d CAN overlays of the same size n. Adding the same number of long links, the d-dimensional CAN overlays can still achieve shorter query hops than the 2-d CAN overlays. Using $K = 4\log_2 N$ seed IDs can achieve $O(\log_2 2n^{1/2})$ query hops in 2-d SCAN overlays. Thus, using the same number of seed IDs can also achieve $O(\log_2 2n^{1/2})$ query hops in d-dimensional SCAN overlays of size n if $dn^{1/d} < 2n^{1/2}$.

3. EXPERIMENTS AND CONCLUSIONS

Figure 2 (a) depicts the distribution of query hops in a 2-d Kleinberg small-world mesh (Kleinberg 2D) and a 2-d SCAN with $N = 2^{20}$ (SCAN_2D). Both have 1024 nodes. Kleinberg's 2-d mesh is strictly regular, having shorter average query hops. Figure 2 (b) shows that in a 2-d SCAN with $K = 4\log_2 N$, the average number of long links (curve 2_avg_rt) is bounded by $4\log_2(2n^{1/2})$ (curve 4LOG). The average query hops (curve 2_avg_qr) is bounded by $\log_2(2n^{1/2})$ (curve LOG). Figure 2 (c) and (d) demonstrate that using $K = 4\log_2 N$ seed IDs in SCAN overlays with d = 3, 4 and 5, the average query hops (curves d_avg_qr) and the maximum query hops (curves d_max_qr) are bounded by those of 2-d SCAN overlays. Figure 3 shows that as d increases, CAN overlays without long links can also achieve $O(\log n)$ query hops with routing table size of $O(\log n)$. Experiments demonstrate the effectiveness and the efficiency of SCAN. It can be extended to support multi-dimensional queries based on other distance metrics. Future work also includes applying the load balancing method in one-dimensional ring [7] to SCAN.

4. ACKNOWLEDGEMENTS

This work was supported by the National Basic Research Program (973 project no. 2003CB317000) and the National Science Foundation of China (No. 60503047).

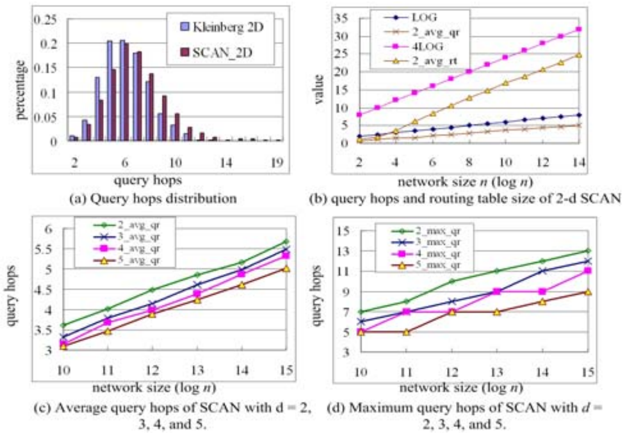

Figure 2. Topology properties of SCAN.

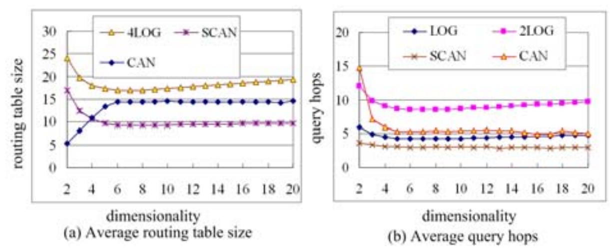

Figure 3. CAN and SCAN overlays with different dimensionality.

5. REFERENCES

[1] F. Banaei-Kashani, C. Shahabi, "SWAM: A Family of Access Methods for Similarity-Search in Peer-to-Peer Data Networks", in *Proceeding of ACM CIKM 2004*, pp: 304 - 313, 2004

[2] J. Kleinberg, "The Small-World Phenomenon: An Algorithmic Perspective", in *Proceeding of 32nd ACM STOC*, pp. 163-170, 2000.

[3] G. Manku, M. Bawa, and P. Raghavan, "Symphony: Distributed Hashing in a Small World", in *Proceeding of the 4th USITS*, pp. 127 - 140, 2003.

[4] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker, "A Scalable Content-Addressable Network", in *Proceeding of SIGCOMM 2001*, pp. 161 - 172, Aug. 2001.

[5] C. Tang, Z. Xu, S. Dwarkadas, "Peer-to-peer Information Retrieval Using Self-Organizing Semantic Overlay Networks", in *Proceeding of SIGCOMM 2003*, pp.175-186, 2003.

[6] Z. Xu and Z. Zhang, "Building Low-maintenance Expressways for P2P Systems", *Tech. Rep. HPL-2002-41*, Hewlett-Packard Labs, Palo Alto, CA, 2002.

[7] H. Zhuge, X. Sun, J. Liu, E. Yao and X. Chen, "A Scalable P2P Platform for the Knowledge Grid", *IEEE Transactions on Knowledge and Data Engineering*, Vol. 17, No.12, pp. 1721-1736, 2005.