論文の内容の要旨

論文題目　A STUDY OF PEER-TO-PEER SYSTEMS FOR SPATIAL DATA SHARING

（空間データ共有のためのピアツーピアシステムに関する研究）

氏名　魏　新法

This thesis is about peer-to-peer systems for sharing spatial data. In recent years, tremendous improvements in data gathering techniques have contributed to an unprecedented growth of available spatial data at geographically distributed locations. This has created a strong motivation for the efficient sharing of such data. While peer-to-peer systems becomes an important approach of massively distributed systems not only for file transfers but also for searchable data network, in this thesis we study the methods of sharing geographically distributed data with P2P networks. By now, there are a number of P2P protocols proposed, but most of them are based on distributed hashing tables, such as CAN and Chord which support exact match only and have limited number of search predicates support. When we consider a set of peer nodes as a massively geographically distributed database, several types of search predicates should be provided in addition to exact match search. So in order to efficiently support spatial data sharing peer-to-peer application, we need to design new peer-to-peer systems that broaden the types of query processing, improve the performance and are of fault-tolerance.

This thesis first presents the design and evaluation of GNet, an early work of exploring the possibility of geographical peer-to-peer protocol that targets supporting wide area location-based service. The GNet protocol uses hierarchical geographic address as the identifiers of peer nodes. By combining domain-progressive routing mechanism like plaxton mesh with geographical domain hierarchy, this protocol has the advantages of efficient routing, locality preserving, etc. It supports position-based and especially geographically scoped operations efficiently. Though implementation prerequisites limits its application area, analysis and evaluation results demonstrate

its scalability, query efficiency and load balancing features, which makes it adaptable to certain applications.

As the main contribution of this thesis, DHR-Trees peer-to-peer protocol is presented with its structure design, collaborative multidimensional query method, maintenance method and the cost analysis, and methods to strengthen fault-tolerance of structure and query execution under dynamic network environment. Essentially, DHR-Trees structure is the first peer-to-peer structure that has semi-independent R-Trees structure and supports region-based multidimensional search predicates, such as range queries and nearest neighbor queries as in R-Trees structure, while dealing with network dynamism efficiently as well.

The thesis presents the structure details of DHR-Trees. Instead maintaining a global centralized R-Trees index, each peer owns a semi-independent partial region tree structure, which makes it possible to keep correctness of structure even under dynamic network changes. Each geographically distributed peer node is identified by its Hilbert value on a Hilbert space filling curve, by which the peer's two-dimensional geographical location is mapped to one-dimensional identifier. Peer nodes self-organize into a virtual ring topology, sorted by the identifiers. As the core part of the DHR-Trees' protocol, each peer maintains a routing table, which contains two principal parts: For routing purpose, it holds pointers to a number of nodes in the network; for supporting spatial query purpose, the region information of sub-trees in the DHR-Trees is contained.

Spatial queries are executed in a distributed fashion by collaborative efforts among peers. DHR-Trees mainly provides three spatial query functions: point query, range query and nearest neighbor query. By exploiting region information in routing table, the spatial query evaluation results show that DHR-Trees can execute spatial queries much more efficiently than its competitor, the Squid P2P protocol. Furthermore, the nearest neighbor query, one of most important spatial queries which is unsupportable in Squid, can also be efficiently executed.

DHR-Trees faces network churn problem as well as other P2P systems. To keep the system working properly while nodes join, leave, and fail on their own agenda, each peer node is required to maintain both the ring structure and routing table. To maintain ring structure, it uses similar ring stabilization approach as in Chord protocol. For routing table maintaining, processes includes ping, stabilization, and notification process are run periodically or triggered by routing table change events. Our analysis and evaluation result shows that the overhead of updating routing tables when a new node joins or fails increases nearly logarithmically to the network size. This demonstrates the scalability of DHR-Trees peer-to-peer system.

To improve the fault-tolerance on spatial query support, DHR-Trees proposes two approaches: entry successor list and adaptive bounding rectangle. By introducing successor lists to the entries in the routing table, robustness and resilience are greatly improved. Moreover, to eliminate the frequent updating requirements of the region information in harsh churn environment, we introduce the usage of adaptive bounding rectangle as the replacement of minimum bounding rectangle. This approach decreases the updating overhead and greatly improves the quality of query result under churn.

Through this thesis, two new novel peer-to-peer protocols are provided. Both GNet and DHR-Trees are designed to be architectures for sharing geographically distributed spatial data. In particular, the DHR-Trees can not only index spatial data as in centralized R-Trees, but also be able to handle dynamism in the peer-to-peer network. We believe our approaches can help realization of certain distributed spatial data sharing applications. We hope our works will stimulate more research interest in both peer-to-peer structures and spatial data sharing applications.