Distributed Incomplete Pattern Matching via a Novel Weighted Bloom Filter

Siyuan LIU, Carnegie Mellon University
siyuan@cmu.edu
Lei KANG, Hong Kong University of Science and Technology
lkang@cs.wisc.edu
Lei CHEN, Hong Kong University of Science and Technology
leichen@cse.ust.hk
Lionel NI, Hong Kong University of Science and Technology
ni@cse.ust.hk

32nd IEEE International Conference on Distributed Computing
Systems (ICDCS’12), Macau, China, June 2012

LARC Accepted Conference Papers: http://smu.edu.sg/centres/larc/conference-papers/
ABSTRACT

In this paper, we first propose a very interesting and practical problem, pattern matching in a distributed mobile environment. Pattern matching is a well-known problem and extensive research has been conducted for performing effective and efficient search. However, previous proposed approaches assume that data are centrally stored, which is not the case in a mobile environment (e.g., mobile phone networks), where one person’s pattern could be separately stored in a number of different stations, and such a local pattern is incomplete compared with the global pattern. A simple solution to pattern matching over a mobile environment is to collect all the data distributed in base stations to a data center and conduct pattern matching at the data center afterwards. Clearly, such a simple solution will raise huge amount of communication traffic, which could cause the communication bottleneck brought by the limited wireless bandwidth to be even worse. Therefore, a communication efficient and search effective solution is necessary. In our work, we present a novel solution which is based on our well-designed Weighted Bloom Filter (WBF), called, Distributed Incomplete pattern matching (DI-matching), to find target patterns over a distributed mobile environment. Specifically, to save communication cost and ensure pattern matching in distributed incomplete patterns, we use WBF to encode a query pattern and disseminate the encoded data to each base station. Each base station conducts a local pattern search according to the received WBF. Only qualified IDs and corresponding weights in each base station are sent to the data center for aggregation and verification. Through extensive empirical experiments on a real city-scale mobile networks data set, we demonstrate the effectiveness and efficiency of our proposed solutions.