

The Data-Centric Revolution in Networking

Scott Shenker

ICSI and U. C. Berkeley
Berkeley, CA
USA

shenker@icsi.berkeley.edu

Summary

Historically, there has been little overlap between the database and networking research communities; they operate on very different levels and focus on very different issues. While this strict separation of concerns has lasted for many years, in this talk I will argue that the gap has recently narrowed to the point where the two fields now have much to say to each other.

Networking research has traditionally focused on enabling communication between network hosts. This research program has produced a myriad of specific algorithms and protocols to solve such problems as error recovery, congestion control, routing, multicast and quality-of-service. It has also led to a set of general architectural principles, such as fate sharing and the end-to-end principle, that provide widely applicable guidelines for allocating functionality among network entities.

This research and design paradigm has been exclusively *host-centric*; hosts are assumed to know which other hosts (or multicast groups) to contact, and the research focuses on making the resulting host-host communication robust and efficient. However, an increasing number of applications involve accessing particular data objects whose location can't easily be determined within the current Internet architecture. Networking researchers have consequently begun looking at a variety of approaches that are more *data-centric* than host-centric, in that the basic abstractions refer to the name of the data rather than its location.

This data-centric trend is most visible in two areas of networking research: sensornets and distributed hash tables. Data-centrism is natural for sensornets because the

identity of individual nodes is far less important than the data they collect. Traditional networking has toyed with data-centrism in various limited forms (*e.g.*, web redirection, intentional naming), but the recent advent of distributed hash tables has led to a much broader and more explicit engagement with the data-centric paradigm. In both the sensornet and traditional Internet cases, data-centric research initially focused on how to efficiently access data based on logical names. More recent research has used distributed data structures to support more general queries.

Networking researchers have thus unwittingly wandered into the Coddian world of physical data independence, a territory far more familiar to database researchers. This talk will describe our journey to your land.

Permission to copy without fee all or part of this material is granted provided that the copies are not made or distributed for direct commercial advantage, the VLDB copyright notice and the title of the publication and its date appear, and notice is given that copying is by permission of the Very Large Data Base Endowment. To copy otherwise, or to republish, requires a fee and/or special permission from the Endowment

Proceedings of the 29th VLDB Conference, Berlin, Germany, 2003