Research Statement

Qixiang Sun

My research interest is in distributed systems, consisting of many networked components that interact with each other to perform tasks. With increased Internet access, distributed systems with millions of participants are now common. Associated with this increase in the number of participants, we also have an enormous amount of information available. As a result, one important research area is how to facilitate information sharing and how to manage the data. Peer-to-peer file-sharing is just one popular example. This new focus on how we share and retrieve information has compelled us, as distributed systems researchers, to build a system by first understanding how information will be used in an application and then design the system according to these needs. In other words, we design systems with information management as a first-order concern rather than an afterthought. I call this new trend of system design the *information-centric approach* to distributed systems. My research centers on this information aspect.

The information-centric approach to distributed systems will be even more critical as new technologies, such as mobile wireless devices in cars and radio frequency identification (RFID) tags, become widely deployed. These technologies not only bring more connectivity and bandwidth to users, but also increase the amount of information generated and consumed. And utilizing these technologies to build useful distributed systems that manage this information is an important problem. For example, how do we build systems that enable passengers in cars to retrieve information from nearby stores? Moreover, such information can be generated in real-time from sources like RFID readers. There are many challenges in building such a system that range from technical details like maintaining the network to modifying user behavior like encouraging participants to cooperate. Some of the desirable properties for a scalable information system are:

- *Local Interactions* - Individual participants should not rely on any global knowledge, but act simply based on locally observable information. Yet as a whole, the system should function harmoniously.
- *Viable Incentives* - Participants are selfish in that they want to achieve their objectives while contributing as little as possible to the system. In order for a system to thrive, we need mechanisms so that it is in a participant's own interest to contribute.
- *Context Sensitive* - Application-level semantics often depend on context such as geographical location. Thus a system must take the context into account and distinguish between different contexts.

In my Ph.D. thesis, I address some of the issues raised above. I designed the Selfish Link-based InCentive (SLIC) protocol, a simple and purely local incentive mechanism in the context of an unstructured search network for file sharing that uses flooding for search queries. The goal of the incentive scheme is to encourage users, in exchange for better service, to share more data, devote more capacity for query processing, and establish more connections to improve the overall system connectivity. My SLIC protocol achieves these goals without relying on a global reputation system or a global economic system. In SLIC each participant simply rates its neighbors and then divides its own resource among the neighbors according to how "good" the neighbors are. The notion of "good" is context sensitive. In the case of flooding-based search network, I used the number of hits for a query. In other context, the quality of search hits might be more appropriate. Via simulation, I demonstrated that SLIC indeed creates a viable incentive structure. Malicious behaviors are also punished quickly. I have also studied SLIC under conditions where participants actively change their network connectivity in the hope of improving their service.

Another area that I have worked on is disseminating information in ad-hoc inter-vehicle networks that only relay messages wirelessly between cars. The specific application that I focused on is delivering an alert, e.g. a road condition, to cars within a geographic region for a duration. This application requires the local interactions, viable incentives, and context sensitivity properties that I mentioned before. First of all,

because of the mobility of cars, making decisions based only on local knowledge is critical. Secondly, the context of a geographical location is a key aspect of the problem. And lastly, we must have incentives for cars to relay alerts for the benefit of others. I developed a simple solution, the Bidirectional Perimeter-based Protocol (BiPP), that addresses the local interaction and context sensitivity aspects of the problem. BiPP guarantees the deliver of alerts to all "reachable" cars without excess broadcast overhead. As the name BiPP implies, I solve the challenge of efficiently maintaining an alert for a time period by exploiting any two-way traffic on a road and intelligently setting up a perimeter for notifying new cars. Although BiPP does not handle incentives, we can side-step the problem by relying on tamper-proof hardware with identifications and trusted programs in cars.

In the future, I plan on continuing to study large-scale distributed systems where managing information is a critical component. There are many emerging problems where local interactions, viable incentives, and context sensitivities are important. One example scenario arises from a municipal or community wireless environment (Muni-WiFi). In Muni-WiFi, a city, like San Francisco, deploys a broadband wireless infrastructure to provide city-wide coverage. A Muni-WiFi deployment connects a vast array of technologies like wired clients, wireless roaming components, geographical localization systems, and sensors such as traffic sensors or even RFID readers.

I believe Muni-WiFi is a rich area of research and a promising future direction because we have to integrate, manage, and deliver large volumes of information. Moreover, there are many useful applications that can significantly improve our daily life. One application is to disseminate relevant information according to geographical location and time, in a publish/subscribe manner. For instance, near a ball park during a game, wireless users in a car or people with PDAs would be interested in receiving scores and traffic data. On the other hand, near the financial district, users would prefer stock quotes and news. Another application of Muni-WiFi is to search for data about an area that a user is currently in. If a user, as an illustration, is interested in lunch, he or she can "query" nearby neighborhoods for restaurants, menus, seating availabilities, etc..

Supporting various applications in a Muni-WiFi environment has many challenges. One challenge is what kind of underlying support architecture or middle-ware to build. One solution might be to form a peer-to-peer network among users' home computers on the wired infrastructure for information sharing, and the wireless devices then simply contact their "home" computers for relevant data. An alternative is to develop ad-hoc networks based on geographic location and interests. Exploring applicability of different middle-ware architectures with respect to end-user applications is of fundamental importance.

Another challenge is to manage the information in the system intelligently. A system where users must explicitly specify every piece of information that they are interested in will not be practical. For example, if I can only get traffic information by pre-specifying the travel route every time I get into a car, then the system is too tedious to use. Instead, we should automatically identify a car's location based on its wireless transmission signal and deliver appropriate traffic information directly without any user intervention. Similarly, if a client with a wireless device is walking on the sidewalk near stores, then we should directly route information from RFIDs about merchandise in the stores to the client's device, in case the client is interested. Also, after the client walks beyond the stores, the client's device should automatically stop displaying the data. Devising mechanisms for automatically generating relevant information and delivering the data is an interesting and important problem. As with any information gathering, synthesis, and dissemination, there are also many issues revolving around security and privacy. We have to protect users' private data while still be able to suggest and deliver useful information.

In short, I believe that by studying large-scale distributed systems from an information-centric perspective, we will improve our ability to build better systems that serve the needs of the users. I think the municipal wireless environment scenario is practical, and a promising "driving application" of research for the near future.