Wireless Cloud Infrastructure for High-Performance Wireless Applications and Services

Xin Wang, Stony Brook

(xwang@ece.sunysb.edu)

Two topics that best fit research interests:

Cloud Self-Monitoring and Autonomic Control, Network Support for the Cloud

Current research activities:

In a Cloud service environment, besides long-term data transfers, many short-term connections may be needed to transmit management information, while no communication is needed for a long period when the providers are busy with certain services. Current routing protocols are normally designed to support long-term traffic. In topology based routing, maintaining the paths proactively or searching for an end-to-end path through flooding would be very expensive if the traffic is sporadic or short-term. Although geographic routing protocols are more scalable by making forwarding decision only based on the local topology, the use of proactive fixed-interval beaconing for location update of a node would result in location outdate and routing failure under high mobility, and create a lot of control overhead when there is no traffic. Restricting path searching to one-hop may result in inefficient paths, while inaccurate destination position can lead to delivery failure. It is also hard to preset the routing parameters optimally for different environments.

In addition to one-to-one transmissions, in a Cloud service environment, multicast transmissions between one sender and multiple receivers are frequently required, e.g., to send search messages or distribute information. This type of traffic is often also short and rare. There is a need to avoid significant control overhead for maintaining multicast paths, and to ensure transmission reliability. The existing multicast protocols are generally inefficient in a service environment, due to their large overhead for route searching, group membership management, and tree/mesh structure creation and maintenance over the dynamic topology. Current geographic multicast protocols also have limited scalability due to the need of inserting the positions of all group members in the packet headers and searching for the positions of all members through location services. Group membership management, which is critical for service collaboration, is also not considered.

We have designed a self-adaptive on-demand hybrid routing scheme (SOHR) which builds efficient paths based on the need of user applications and adapts to various scenarios to provide efficient and reliable routing. In addition, we propose a robust and scalable geographic multicast protocol (RSGM) which can builds virtual multicast infrastructure on demand without need of maintaining state information for more reliable and scalable membership management and packet forwarding in the face of high network dynamics due to unstable wireless channels and node movement. SOHR and RSGM are extremely fit for use in mobile cloud environment to *instantly form various types of virtual distribution infrastructures* (one-to-one, one-to-many, many-to-one) upon service need, which could facilitate flexible, scalable and robust service provision for mobile cloud. In addition, we have designed geography-based service provision framework which can be extended to meet the service management need in Cloud.

We are working on developing a scalable, reliable, and flexible Pub/Sub based information and data management infrastructure to meet various application needs in presence of wireless channel dynamics and mobility. In addition, being invited to participate in the Hawaii project of Microsoft research, we are integrating our research with our educational effort to develop various schemes to support mobile cloud.

Research problems to pursue:

Cloud computing can potentially enable performance-hungry applications on almost any mobile devices by executing a resource-intensive application on a high-performance server or server cluster in a remote Cloud. Although promising, there are some potential barriers to this shift in mobile computing. The issues include the lack of mobile Internet access everywhere, the limited wide-area wireless network bandwidth, and the big delay involved in WAN-level Internet transmission. Instead of relying on a distant cloud, a viable option is to let mobile devices to leverage use of nearby resource-rich computers or servers whenever possible while gracefully falling back to cloud resources when needed.

The pooling of resources has the potential to enable a number of pervasive and ubiquitous services. However, the concurrent use of local resources (including enterprise resources) and remote Cloud is largely unexplored. Specifically, there is a lack of an efficient service management infrastructure to discover, recruit and coordinate the nodal resources. To support wireless services over higher-speed but shorter-range wireless links, the main challenges are the relative unreliability of the wireless medium, the mobility of nodes, and the resource limitations of wireless devices. The service infrastructure has to allow data to be collected and distributed reliably to and from widely distributed nodes, and meet the rate and delay requirements imposed by emergent applications. The service framework has to provide mechanisms to locate and track resources and operational states at a large number of nodes which are possibly mobile and unreliable, enable resource sharing and service collaborations, and maintain service reliability. Despite the fact that each topic of resource monitoring, resource discovery, and task allocation is crucial for forming the service infrastructure, there are limited studies in literature on each topic over dynamic wireless networks and in presence of uncertainty and constant changes of resources, not to mention a combination of the three due to numerous inherent challenges.

The objective of our work is to develop a scalable and robust service infrastructure to enable seamless access of resources from local resources and remote Clouds for high-performance wireless applications and services in presence of node mobility, network and resource dynamics, as well as uncertainty of environment and resource consumption. To better handle node mobility and radio channel dynamics, and consequent changes in network topology and resource availability without incurring excessive management overhead, we will integrate location information into the management infrastructure to significantly reduce control overhead and increase reliability through *various stateless virtual infrastructures*.

Our planned research will consist of four thrusts: (1) *Developing* a *robust and scalable monitoring framework* to track locations, resources and service availability, and liveness of resource providers and clients in a dynamic wireless environment; (2) *Designing an efficient service and resource discovery scheme* to allow the registration of distributed resources, the publication of services, the quick finding of resources from a large pool of servers, and the coordination of a group of servers to support services requested by a client; (3) *Developing a scalable, reliable, and flexible Pub/Sub* based information and data management infrastructure to meet various application needs in presence of wireless channel dynamics and mobility; (4) *Design of an efficient task allocation scheme* to approach optimal resource allocation, while ensuring scalability, reliability and stability of resource management in a dynamic system with inaccuracy and uncertainty in modeling of resource consumption and constant changes of resources; (5) *Testbed implementation of the proposed service infrastructure*, exploiting cloud tool such as Miscrosoft Window Azure platform to enable wireless clients to seamless access resources from local resource pool and remote Clouds.

The collective use of nodal resources and virtualization of services will see wide usage in the fields of commerce, medicine, science, education, entertainment, emergency rescue, among others. The limitation of wireless devices and networks would be greatly mitigated with the pool of resources collected through the proposed resource and service management infrastructure. With efficient discovery, recruiting, and coordination of nodal resources in a dynamic and changing environment, the proposed resource and service management infrastructure will potentially enable more reliable, powerful, pervasive, and new class of mobile Cloud computing and services