Research Statement

Research Philosophy

I like to conduct my research on problems that are both practically imperative and theoretically significant. My approach to research is first to understand user requirements, identify underlying technical challenges, and then design protocols and architectures that can improve current systems. I also actively seek possible paradigm shifts that can drastically expand and change the understanding and view of the existing systems. My research goal is to investigate this interaction between network infrastructure and video streaming applications and services, and tries to provide insights on how to improve the performance of both aspects.

Ph.D. Research In University

My Ph.D. research includes design and analysis of protocols and architectures to improve the performance of current systems, and to enable new services and applications. Specifically, it includes the following aspects:

1. Mobile and Pervasive Computing

- **Coordinated data transmission in mobile ad hoc networks** Data aggregation in mobile ad hoc networks is challenging since there is no fixed path from the sources to the destination. This research tackles data aggregation problem in two phases: data collection and data transmission. A novel signaling protocol is designed for coordinated data collection. Also, an application layer solution which uses an advanced data availability query scheme is proposed to further eliminate data redundancy on the data forwarding path.
- **Enable live video streaming in vehicle to vehicle (V2V) networks** V2V network is a special mobile ad hoc network in which mobile nodes do not have power constraints, and are equipped with considerable computing and storage capability. Streaming video in V2V networks is very challenging since 1) a V2V network is highly dynamic due to high vehicle speed; 2) V2V network can be consistently partitioned due to limited vehicle instrumentation ratio. This research designs *V3*: a vehicle-to-vehicle video streaming architecture. V3 incorporates a novel signaling scheme to continuously trigger vehicles into video sources. It also adopts a store-carry-and-forward routing scheme to transmit data in highly partitioned network environments.

2. Resilient Overlay Multicast

- **Scalable video streaming with time shifting and video patching:** This research considers the problem of how to enable continuous live video streaming to a large group of clients in overlay networks. One major challenge for this application is to maintain video continuity when the overlay network structure is disrupted. Two techniques are designed to achieve continuous video streaming in case of node leave or node failure: 1) A time-shifting video server which broadcasts not only the original stream but also streams with different time-shifting values. Clients that suffer service disconnection can then join a video channel of the time-shifted stream to retrieve the lost video content. 2) A live video patching scheme to allow a client to catch up with the progress of the original video program. Video patching technique requires a client to receive twice the rate of a video stream. The impact of video patching to the structure of the application layer multicast tree is studied.

- **A Cooperative patching architecture in overlay networks** There are two major limitations in the time-shifting video server approach: 1) Potential performance degradation, the video server is the only point to provide patching streams for clients. In the case where multiple patching requests are received, the server has to delay some requests when there is no free channel available; 2) Server design complexity, the server not only has to broadcast the live stream, but also has to serve time-shifted video streams on demand. This research proposes a client-based *cooperative patching* architecture to achieve continuous live video streaming in overlay networks. In this design, each end host keeps the video that has been played out for a certain time before discarding it. An end host maintains a list of *patching parents*, and retrieves lost data from one of the patching parents. This architecture relieves server load by shifting video patching responsibility to the client side. Video content is replicated in multiple locations across the overlay network to provide fast and timely data recovery.
- **Path diversified routing in content distribution networks (CDN)** In a CDN, replicated servers are deployed strategically to provide scalable content delivery service. With the development of multiple-description coding (MDC), these servers can cooperatively serve a client's request to provide resilient data service. Different descriptions from different servers are sent along disjoint overlay paths to the video client. This approach is robust since a link failure in one path can only affect a single description of video content. This research investigates the effect of path diversity on robustness and quality of video communication in content distribution networks (CDN). A set of server selection schemes to achieve path diversity in MDC-enabled CDN are proposed and evaluated.

3. Large Scale and Quality Conscious Data Distribution

- **Scalable on demand video streaming service architecture** Due to the bandwidth-intensive nature of video streams, distributing on-demand video content to a large number of users in real time is very difficult. A widely adopted approach is to batch the clients requesting for the same video and serve them using one multicast stream. Another commonly used solution is to replicate video content into multiple locations and serve client requests from a nearby server. This research investigates the design of server selection techniques for a system of replicated batching VoD servers. The key to the performance of a replicated VoD system is the problem of server selection, this work argues that server channel and buffer usage information is necessary in making right server selection decisions. A set of server selection algorithms that use increasingly more information are proposed. A directory-server based architecture to support the proposed algorithms is also described.
- **Quality of service (QoS) based anycasting in Differentiated Service (DiffServ) networks** In anycasting, one end point is specified while the other is selected from a set of end points. QoS based anycasting requires selection of both an end point and a path that satisfies the QoS constraint. This research proposes a probing-based protocol to discover network resources and make reservations for requests in DiffServ networks. To reduce probing overhead, a set of schemes using passive and active caching are proposed. This research work also evaluates the impact of partial probing, in which only a subset of end points are probed, on overall system performance. I am also in charge of a DiffServ Testbed project. This testbed is composed of

two DiffServ domains, and multiple software routers. Multiple video streams with different QoS requirement travel through the DiffServ, and user perceived performance is recorded and evaluated under various network traffic load condition. This project is a group effort from OIT of abc. This work was successfully demonstrated in Internet 2 conference on November 2000.

Research at Google Video

Mr. Last is a research engineer of the video team at Google. His research focuses on providing reliable and scalable video streaming service to millions of clients. More specifically, his research addresses the following challenging problems: 1) Bandwidth management: GOOGLE videos typically stream at several hundred kbps data rate, supporting millions of playbacks everyday poses an excessive load on the backbone networks; 2) Storage management: each video file can be tens of mega bytes up to several gigabytes, to provide high video availability to end users, these videos often need to be replicated to multiple locations; 3) User perceived performance optimization: end users expect great video programs, and they want the content with nice image quality and minimum latency.

Bandwidth management Currently, the only widely available data transfer scheme is through unicasting. Supporting mere several hundred streams simultaneously can easily congest any network channel, let alone thousands of them. One approach we use is using large scale server farms combining with a sophisticated load balancing scheme to distribute the load evenly. The other approach is to distribute the video content to different geographical locations to even out the bandwidth usage.

Video storage management As mentioned above, to distribute the network traffic globally, video contents are stored in more than one locations, this requires more disk spaces everywhere, what is more, copying huge amount of video data across different locations also poses lots of overhead traffic. To address this problem, we use a multi-hierarchy caching scheme which acts as a middle layer to the replication system. In this way, we can achieve better bandwidth distribution without over-replicating our video repository.

Reach to more users With the involvement of computing technology, platforms other than PCs are now equipped with video capabilities, such as cell phones, pocket computers, etc. They can also receive video data through heterogeneous network media. To make google video service accessible to all the users, we encode video content to multiple formats and support streaming of those based on the client side system. To adapt to different network channels, multiple transport protocols are used to stream video content, such as HTTP and RTSP. Lots of my research are focused on improving system performance under new communication paradigms, and developing new services over evolving network infrastructures.

User perceived performance optimization There is a great variation on the video playback patterns. For example, soccer related videos such as Euro Cup are very popular in Europe, while not so hot here in the states. Placing such videos to servers in the Europe instead of in America can achieve better user experience. I devote my research on study the user viewing patterns, such as their geographical location, bandwidth capacity, preferences etc., and integrate them to our video serving system to improve user experience.