NetQuest: A Flexible Framework for Internet Measurement

Lili Qiu
Joint work with Mike Dahlin, Harrick Vin, and Yin Zhang

The Internet connects over 200 million hosts and nearly a billion users worldwide. How to better understand and thereby control this enormous, decentralized, and constantly evolving infrastructure is a major challenge faced by today's researchers, engineers, and network operators. As an essential means of achieving better understanding, network measurement has become crucial to a variety of existing and emerging network applications, such as ISP performance management, traffic engineering, content distribution, overlay routing, and peer-to-peer applications.

While much progress has been made in network measurement, several significant challenges remain. First, large-scale network management applications often require the ability to efficiently monitor the whole network. The quadratic growth of the number of network paths with the number of network nodes makes it impractical to measure every path. Second, existing techniques are often tailored to specific application needs, and thus lack the *flexibility* to accommodate applications with different requirements.

To address these challenges, we develop NetQuest, a measurement framework that would enable large-scale continuous network monitoring. NetQuest consists of two key components: experimental design and network inference.

We apply Bayesian experimental design to design measurement experiments that maximize information gain about network path properties given resource constraints (e.g., probing overhead). Bayesian experimental design is built on solid theoretical foundations, and has found numerous applications in scientific research and practical applications, ranging from software to testing, to medicine, to biology, and to car crash test. However, to the best of our knowledge, it has never been applied to designing large-scale network monitoring experiments. Making the experimental design applicable to such context involves addressing several challenges. First, how to apply the Bayesian experimental design to constructing measurement experiments. Second, there are a number of different Bayesian experimental design schemes, and it is not clear which ones are applicable to networking context. Third, the traditional Bayesian experimental design is targeted for a single application. In our environment, there can be many different applications with different requirements.

To address the issues, we investigate different Bayesian design schemes, and use extensive experimental evaluation to identify the design schemes applicable to network monitoring. In addition, we achieve flexibility by designing measurement experiments that maximize information gain for different objectives and constraints. In particular, our approach can support the following design requirements: (i) augmented design for conducting additional experiments given existing observation, (ii) differentiated design for providing better resolution to certain parts of the network, and (iii) joint design for supporting multiple users who are interested in different parts of network.

Based on the observation obtained from the measurement experiments, we then use inference techniques to accurately reconstruct the global view of the network without requiring complete information. We also leverage prior knowledge obtained from infrequent probing or from dimensional reduction through network embedding to further reduce monitoring overhead and enhance inference accuracy.

Our results show that our measurement framework based on Bayesian experimental design can estimate network-wide average path delay within 15% error by monitoring only 5% paths. It achieves similar degree of accuracy for estimating individual path properties by monitoring 15% paths. In addition, we demonstrate the flexibility of our measurement framework in supporting continuous monitoring, providing differentiated monitoring, and satisfying the requirements of multiple users.

To summarize, while Bayesian experimental design has found many applications in other scientific fields, to the best of our knowledge, this is the first time that it is applied to large-scale network measurement. Building on top of Bayesian experimental design and inference techniques, we develop a unified framework within which a large class of network performance inference problems can be modeled, solved, and evaluated. Our framework is flexible, and can accommodate different design requirements.