Presentation/Poster Abstracts

SAHARA Retreat
June 4-6, 2003

- **The Impact of BGP Dynamics on Intra-Domain Traffic** (Sharad Agarwal)

  BGP convergence and stability properties have been extensively studied, but little work has been done to study its impact on traffic within a single autonomous system. Frequent shifts of large amounts of traffic can be detrimental to traffic engineering and network provisioning from an ISP's standpoint. It can also impact end-to-end performance for customers. Therefore, we correlate BGP routing table changes with packet traces to study how BGP dynamics affect traffic fan-out within the Sprint network. Despite 133 BGP routing table changes on average per minute, we find that these changes do not cause more than 6% variation in the traffic fan-out. This limited impact is mostly due to the relative stability of network prefixes that receive the majority of traffic. We observe that about 0.05% of BGP routing table changes affect intra-domain routes for prefixes that carry 80% of the traffic.

- **Root Cause Analysis of BGP dynamics** (Matthew Caesar, Lakshminarayanan Subramanian)

  The lack of a good understanding of the dynamics of interdomain routing has made efforts to address BGP's shortcomings essentially a black art. To gain more insight into these dynamics, we need to answer two questions: (i) Why does a routing change occur? (ii) Where does a routing change originate? We attempt to answer these questions by collecting routing updates from multiple vantage points and inferring the type and location of an event that triggers a routing change. We first develop a taxonomy of events that cause routing changes and divide them into several equivalence classes based on how they affect routing tables. We then use a two step methodology to infer properties of observed events. First, we classify observed route changes into groups that are likely to have been caused by the same event. Then, we correlate updates within each group to narrow down the suspect set of AS's that might have caused the event. As validation, we show that our algorithm can detect the following well-known events: (a) difficulties with the UUNET backbone on 10/3/02, (b) difficulties with the AT&T backbone on 8/24/02, (c) the NIMDA worm on 9/18/01, (d) the SQL Slammer worm on 1/24/03, (e) and the Code Red worm on 7/19/01. We analyzed several months of BGP update traces and found that our techniques can accurately pinpoint the location of calamitous routing events (e.g. session resets) which affect many prefixes. Aside from these known events, our scheme detected certain pairs of AS's which underwent a large number of repeated session resets.

- **Tomography-based Overlay Network Monitoring** (Yan Chen)

  Overlay network monitoring enables distributed Internet applications to detect and recover from path outages and periods of degraded performance within several seconds. However, for an overlay network with n end hosts, existing systems either require $O(n2)$ measurements, and thus lack scalability, or can only estimate the latency but not congestion or failure rates. Unlike many network tomography systems, we propose a tomography-based monitoring system to characterize end-to-end losses rather than individual link losses. We capture the basis set of k path segments that can fully describe all the $O(n2)$ paths. Then we selectively monitor k paths to estimate the loss rates of k basis, and consequently the loss rates of all other paths. Given a power-law degree Internet topology, if the majority of end hosts are on the overlay network, there are only $O(n)$ links, thus $k = O(n)$. Otherwise, for reasonably large n (e.g., 100), we find that k is still in the range of $O(n)$ by extensively studying synthetic and real topologies. It is probably due to the moderate hierarchical nature of Internet routing. Our technique is tolerant to topology measurement errors, and is adaptive to topology changes.

Preliminary simulation results show that we achieve high path loss rate estimation accuracy with only $O(n)$ measurements. For instance, we find over 99% of lossy paths with a false positive ratio under 2%.

- **Estimating Shared Congestion among Internet Paths** (Weidong Cui, Sridhar Machiraju)

  Recent work on media streaming has proposed to exploit *path diversity*, i.e., the use of multiple end-to-end paths, as a means to obtain better performance. The best performance is achieved when the various paths are independent in the sense that the two paths do not share a *Point of Congestion (PoC)*. However, topologies used in media streaming applications do not meet the assumption of Inverted-Y or Y topologies made by prior work on detecting shared PoC. In this paper, we propose a new technique called *CD-DJ (Correlating Drops and Delay Jitter)* which solves this problem. CD-DJ is better than earlier solutions for three main reasons. First, CD-DJ overcomes the clock synchronization problem and can work with most topologies relevant to applications. Second, it provides applications with an estimate of the fraction of packet drops caused by shared PoCs. This information is more useful than a "yes/no" decision for media streaming applications because they can use it to choose a path based on the level of shared congestion. Third, CD-DJ makes the estimation by correlating bursts of packet drops in conjunction with the correlation of delay jitter in a novel way. A key contribution of our work is our evaluation methodology. We use a novel overlay-based method to evaluate our technique extensively using about 800 hours of experimental traces from Planetlab, a global overlay network. Our results indicate that CD-DJ calculates estimates which are at least within a factor of 0.8 of the actual fraction of shared drops for $80 - 90\%$ of the flows. We also illustrate the advantage of using CD-DJ with a simple streaming video application.

- **Performance Characteristics of TCP Throughput for Peer Selection in Peer-to-Peer Networks** (Weidong Cui, Li Yin)

  Peer selection is one of the key challenges in today's peer-to-peer netorks. Although this problem is similar to the mirror server selection in the Internet, mirror server selection algorithms cannot be applied to peer selection directly because peers are heterogeneous and have limited upload bandwidth. New study about performance characteristics of TCP throughput in peer-to-peer networks is desired for the design of efficient peer selection algorithms. Our paper studies the performance of TCP throughput for peer selection using traced collected in a real peer-to-peer system. Based on the collected traces, we analyze the TCP throughput in the peer-to-peer system, the impact of heterogeneous peers and the performance of various peer selection techniques We compare three types of probing based peer selection techniques including RTT-based probing, size-based probing and time-based probing. We conduct extensive trace-based experiments and show that RTT-based probing can achieve 60% optimal, size-based probing can achieve 70%-80% optimal, and time-based probing can achieve 80%-90% optimal. In addition, we also study parameter selection for size-based and time-based probing under file download of different sizes. We show that the download time of optimal time-based probing is always less than that of optimal size-based probing.

- **TAPAS: A Research Paradigm for the Modeling, Prediction and Analysis of Non-stationary Network Behavior** (Almudena Konrad)

  This presentation introduces two research methodologies for the efficient development of models and performance analysis techniques to be applied to network measurements. The first methodology consists on the preconditioning of measurement data in order to fit traditional mathematical models. This methodology emerges from the fact that data measurements experience complex patterns and time-varying path characteristics due to internal network components, and applying traditional models to this data generates poor models. We develop two data preconditioning models and present a novel approach that enables network researchers to quickly select the most accurate modeling and analysis method for a given wired or wireless network path and network characteristic of interest (e.g., delay, loss, or error process). We show that traditional modeling approaches, such as Discrete Time Markov Chains (DTMC) are limited in their ability to model time-varying characteristics. This problem is exacerbated in the wireless domain, where fading events create extreme burstiness of delays, losses, and errors on wireless links. We present a wireless simulator (WSim) that provides the two data preconditioning models developed from our preconditioning modeling methodology and a feedback algorithm to inform the application on events at other network layers. We develop a second methodology for the performance analysis of multi-layer networks. We argue that for the correct evaluation of today's networks, we must draw performance conclusions from a detailed study of the cross-layer protocol interactions. In particular, we studied the interactions between the Transmission Control Protocol (TCP), a reliable end-to-end transport layer protocol, and the Radio Link Protocol (RLP), a reliable link layer protocol for the wireless connection in the GSM (Global System for Mobile communications) network. Each protocol has its own error recovery mechanisms and by studying the interactions of these protocols, we can improve the performance of the wireless GSM system. We have developed a multi-layer tracing tool to analyze the protocol interactions between the layers.

- **Infrastructure Primitives for Overlay Networks** (Karthik Lakshminarayanan)

  In response to the increasing rigidity of the current Internet infrastructure, many companies and researchers have turned to infrastructure-based overlay networks to meet specific application requirements. However, these overlay networks are mostly independent efforts, sharing nothing but the underlying IP infrastructure. To reverse this trend, we propose the design of a shared generic overlay infrastructure that can support a wide range of overlay functionalities. This proposed overlay infrastructure exports two basic primitives: *path selection* and *packet replication*. We use these primitives to build a network weather service (WS) that is able to accurately measure the performance characteristics between any two overlay nodes. Applications can then achieve their desired service by combining the infrastructure primitives with the information provided by the WS and special-purpose computational nodes. We evaluate our prototype using the PlanetLab testbed. We show that the WS can measure the RTT of a virtual link with an accuracy of 95% in over 98% of the cases, and the unidirectional loss rate of a virtual link with an accuracy of 90% in over 89% of the cases.

- **Single Sign-on to Confederated Wireless Networks** (Yasuhiko Matsunaga, Ana Sanz Merino, Takashi Suzuki)

  Thanks to low deployment cost and high expectation of profit, public WLAN hotspot services have been launched by many providers, such as startup companies and telecom operators. Each of these WLAN operates independently in a small area with its own authentication mechanism, causing users to maintain multiple and heterogeneous credentials to access them. To solve this problem, we propose a cooperative service provisioning model where providers confederate to offer wider coverage, allowing seamless roaming of users between them using the same credentials. Various authentication and authorization models can be used by the networks, increasing the flexibility of our approach.

  In such a service provisioning model, users could encounter service providers with which they have a weak or no trust relationship. Security mechanisms are needed to protect users from exposing their sensitive information to such service providers. Our solution for this problem is an adaptive authentication framework, where the authentication model is dynamically determined based on available user credentials and the capabilities of the network access server. An essential part of this framework is a policy-based mechanism on the client side that provides access control to user information.

  To illustrate our solution we have created a single sign-on testbed consisting of four authentication servers and two user terminals with policy engines. We have considered two different authentication models in our architecture: RADIUS and Liberty Alliance. The measured authentication delay ranges from 0.6 to 1.9 seconds depending on the authentication scheme, small enough for WLAN users.

- **Compound L2/Web Authentication for Securing Public Wireless LAN** (Yasuhiko Matsunaga, Ana Sanz Merino, Takashi Suzuki)

  Public Wireless LANs are different from other commercial public wireless services in that they adopts open network authentication schemes. In other words, users are allowed to access the network without having a pre-shared secret with the wireless service provider. While this enables flexible service authorization and payment methods, it causes several security vulnerablities. Such security threats includes theft of service by spoofing MAC/IP address, eavesdropping, message alteration, denial-of-service attack, etc.
  We propose a compound layer-2/web authentication to avoid these security threats. In our scheme, a user first gets authenticated with IEEE 802.1x guest (anonymous) account and establish a shared secret with the network. Then the user performs a web authentication with embedding layer 2 session key digest in the message. The user gets full access to the external network after web authentication. Although our scheme does not provide a protection for DoS attacks, it can prevent all other security threats listed above.
  A prototype system is developed to evaluate our compound authentication scheme. With several modifications in 802.1x/web client and a RADIUS server, it is confirmed that our scheme works with a standard wireless LAN access point and can provide cryptographically-protected access in public wireless LAN systems.

- **Inter-domain Radio Resource Management Protocol Design** (Yasuhiko Matsunaga)

  Wireless LAN systems use unlicensed frequency bands shared by various public and private systems, and their radio resources are managed independently. As the number of wireless LAN systems increases and geographic coverage overlaps, lack of radio resource management coordination may cause significant performance degradation.
  To overcome this situation, we propose a radio resource broker that enables coordinated radio resource management between wireless LAN systems. The radio resource broker collects radio link configuration and statistics from different domains, and optimizes radio resource usage by changing frequency channel or transmission power at the base stations. The radio resource broker can also perform network-initiated handover for dynamic load balancing across domains.
  The inter-domain radio resource management protocol is designed as a private SNMP management information base (MIB), and being implemented on standard Linux workstations. Although our current research is targeted on wireless LANs, the protocol can be applied to other wireless systems.

- **Geographic Routing without Location Information** (Ananth Rao)

  For many years, scalable routing for wireless communication systems was an enticing but elusive goal. Recently, several routing algorithms that exploit geographic information (e.g., GPSR) have been proposed to achieve this goal. These algorithms refer to nodes by their location, not identifier, and use those coordinates to route greedily, when possible, towards the destination. However, there are many situations where location information is not available at the nodes, and so geographic methods cannot be used. In this paper we define a scalable coordinate-based routing algorithm that does not rely on location information, and thus can be used in a wide variety of ad hoc and sensornet environments.

- **Routing Dynamics in Simultaneous Overlay Networks** (Mukund Seshadri)

  End-host controlled routing of flows is likely to become a significant presence in the Internet due to deployment of overlay networks. We consider the situation when many such overlay networks are deployed and share some physical network space. In particular, we identify scenarios which lead to instability of overlay routing based on best-path selection, and propose simple changes based on randomization to improve stability and performance in these scenarios. We also facilitate the reduction of measurement overhead caused by these overlay networks.

- **Load Balancing in p2p Systems** (Sonesh Surana)

  Most P2P systems that provide a DHT abstraction (CAN, Chord, etc) distribute objects among "peer nodes" by choosing random identifiers for the objects. This could result in an O(log N) imbalance. Besides, P2P systems can be highly heterogeneous, i.e. they may consist of peers that range from old desktops behind modem lines to powerful servers connected to the Internet through high-bandwidth lines. In this work, we address the problem of load balancing in such P2P systems.
  We explore the space of designing load-balancing algorithms that uses the notion of "virtual servers". We first present a spectrum of three schemes that differ primarily in the amount of information used to decide how to re-arrange load. We study how to adapt these load-balancing schemes in systems that experience (i) continuous insertions and deletions of objects, (ii) skewed object arrival patterns, and (iii) nodes joining and leaving the system.

- **A Study of Multi-layer Failure Restoration Schemes** (Fang Yu)

  In July 18, 2001, the Baltimore Tunnel Fire severed several IP links simultaneously and caused certain connections 10 times slower than normal. The cause of the disaster is that some IP links share the same fiber span in the optical network. So, when the fiber is cut, all of them fail at the same time. In this poster, we study the impact of multiple link failure on the OSPF routing protocol at IP layer. In detail, we will study the OSPF convergence time, the duration of loopy or invalid routes, and the route fluctuations it generates. From the simulation results, we observe that multiple link failure will cause more problems even compared to high degree single node failure. In addition, different combinations of link failures will yield dramatically different effects.

- **Fast Failure Detection in Overlay Networks** (Shelley Qian Zhuang)

  In this paper we study the tradeoffs between the detection time of a node failure, the control message overhead, and the probability of false positives in peer-to-peer networks. To facilitate this study we consider three simple schemes to maintain liveness information. We then evaluate these keep-alive schemes in the context of a d-regular network, and of a real peer-to-peer lookup protocol, Chord. Our main findings is that by carefully designing the keep-alive algorithms it is possible to significantly reduce the detection time in the system (e.g., up to two orders of magnitude in the case of Chord) without any additional message overhead.