Research Statement

Ionut Trestian
http://www.cs.northwestern.edu/~ict992

My research deals with designing and analyzing large systems that constantly need to meet scaling demands. Current large scale systems face a huge increase in information generated by consuming customers. Scaling such systems under different constraints can prove difficult. On one hand, using powerful dedicated machines can be expensive. On the other hand, many researchers consider that using standard off the shelf hardware (while achieving lower prices) can increase system complexity and sometimes requires a change in paradigms. Protocols and principles that have brought order to the Internet world now show some of their weaknesses and re-engineering is required to improve or maintain performance in the new settings. One can only note that some of the most important techniques that have emerged break the layering paradigm by making decisions based on information available in other layers.

Although there have been a number of significant results related to the social behavior of users when using Internet applications and services no significant research effort can be seen on applying these results on the actual design of systems. For now, one can only conjecture the benefits of using social information in different aspects of system functionality. My work aims at addressing this: using social information such as application usage patterns, communication patterns in the design and implementation of systems with the ultimate goal of scalability. In a broader context my work might help the emergence of new, social paradigms that could improve the design and functionality of systems.

The following sections outline my past and future research directions.

1. Past research directions

In order to scale existing systems one needs to understand how, where and when people are using such them. Trying to understand this, I became fascinated with the question:"What do people do on the Internet?". The usual steps I had to take to answer this question were clear: obtain traffic traces from operational networks and try to determine the applications that people access. With regards to this, some of the work I did comes from an effort to think 'outside the box' and find alternative cheaper approaches for solving networking and systems problems.

Traffic classification and application usage: In collaboration with researchers from Narus Inc(now part of Boeing). I have developed techniques for easily determining application usage across different world regions and across different network address ranges by using queries to search engines. The techniques I have developed are very versatile and I show [SIGCOMM08, ToN10] how they can be used in different scenarios for: traffic classification, determining active IP address blocks, and determining application usage in remote regions. Using my technique one can show differences in application usage in four different world regions (Asia, South and North America and Europe). The method is patent pending and is currently being integrated into some existing Narus products.

Location based services: The Internet is evolving at a fast pace. Technological advances in recent years have brought exciting new capabilities for Internet users - they can access the Internet from almost

anywhere using mobile devices. New services have consequently emerged that expand on the acquired mobility: location based services. Little is known about the way in which mobile phone users use their phones and specifically about the role location plays in terms of the applications that users access online from there. To address this important research problem that could shape the deployment of different types of location based services, together with researchers from Narus Inc. I have conducted the first large scale (280,000 users) study of application usage and mobility across a large metropolitan area [IMC09].

My results show striking differences between applications. Users tend to access bandwidth intensive applications such as downloads or streaming in locations where they spend a large amount of time. Another tendency I observed is that of users to stay informed, connected as they access mail, news, and social networking websites also away from the locations where they spend a large amount of time. Yet, the most surprising result came from uncovering a strong bias towards certain applications at different hotspot locations. I further uncovered that users who access the Internet from these locations will be advantaged in terms of interactions in a location based service system.

2. Future research directions

Mobile network neutrality: In the networking research community, the network neutrality debate has been focused mainly on personal computers as affected devices and p2p as the main affected application. Several systems have consequently been developed so far, as a natural response of the research community, that try to capture an ISP's bias towards p2p flows. The main techniques used to capture such bias consist of conducting either passive measurements or actively sending application probes. The traffic involved being that of p2p applications there is no clear moral line that would distinguish between the two parties (p2p users and ISPs). Indeed, heavy p2p flows can negatively affect available ISP resources and decrease performance for other users. Also the content that users download often times infringes copyright laws. On the other hand the Internet should be regarded by all involved parties as a free network that does not discriminate against individual applications and consequently ISPs should optimize their networks to meet the increase in resource demand of customers. Considering all of the above, systems that passively or actively collect evidences of ISP bias towards p2p applications seem the best solution for this problem as their output mainly points out that such bias exists without effectively taking sides. Sadly, bias shown by Internet providers has proven to be much broader and extensive than pictured above for the p2p case. This bias covers different aspects of the wired and mobile Internet. As the mobile environment has, to the best of my knowledge, been overlooked by the research community so far I took on the task of building a system that addresses network neutrality violations for mobile networks and devices called WindRider [WR].

For this I have developed an application aimed at several popular mobile platforms that captures feedback about the availability and performance (both measured and judged by the user) of Internet services and applications. Designing a system that manages to capture network neutrality violations is a challenging task as I have witnessed from previous work. Doing so for small power constrained devices such as mobile phones can prove even more challenging. One observation that motivates this future work is that computer systems inherently become increasingly complex because of the constant struggle between system designers and malevolent entities. As such, often system designers turn to the end-user as an obvious solution to problems that ultimately affect him. Some examples of problems where the solution can involve also user feedback include: spam (users marking e-mails as spam or legitimate), search (users

bumping up or down search results), content rating (users rating content relevance or quality) etc. I argue that as it is clear that the end-user is also one of the parties affected by mobile operator bias, the users are the ones that can ultimately help solve the network neutrality debate by closing this loop that links users, ISPs, and application or service providers in a natural way, by offering feedback. WindRider has been mentioned by companies such as Google, and Vodafone in their positions with regards to mobile broadband measurement [NAF1, NAF2, VOD].

User-generated content: In ongoing research on user-generated content upload behavior [INFOCOM11] I am dealing with uncovering the hidden ties between different locations that users visit and the content that is uploaded/downloaded from there. Based on this I am currently exploring infrastructure placement for mobile networks. This work is based on one large dataset (two million users across one week) of multimedia messaging communication taken from one large cellular provider located in the United States.

I regard this research as very important because of a few reasons. First, almost any available mobile phone has messaging capabilities thus a messaging based system has the potential to reach billions of mobile users. In fact, messaging is a popular channel for: disseminating information, and media delivery such as pictures, movies, ringtones etc. Second, in emerging regions, Internet connectivity might not exist but cellular connectivity might be available and cellular messaging is a cheap alternative for the mass dissemination of information (for delivering market product prices to farmers for example). Based on the data I have available, I aim to extract communication patterns and based on these patterns analyze how new dissemination techniques can benefit from user mobility and in an efficient manner apply to existing networks. The long-term goal consists of using my research to either develop a simulator for such systems or integrating my work into an existing simulator. Ultimately I will make this available to the research community.

Optimizing content freshness for geo-diverse OSNs: With researchers from the Internet Group at Telefonica Research in Barcelona I am currently working on exploiting structural properties of social networking websites with the purpose of improving the performance of such services. Online Social Network (OSN) applications have become highly popular over the last few years, having large user bases, distributed across the entire planet. It is generally accepted that in order to handle a geographically diverse user base, relevant data should be replicated as close as possible to the end-user, as this reduces the user perceived latency.

The peak-based (i.e., 95-percentile) pricing policy used on WAN links raises an interesting trade-off between freshness of replicated content and bandwidth cost, where freshness characterizes how local copies on replicas capture latest updates. The system we propose [WYT] optimizes the aforementioned tradeoff by scheduling the transmission of contents, leveraging: (i) knowledge of the mapping between social relationships and geographic location, friends being generally geographically close, and (ii) knowledge of timing regularities in the activity pattern of users

References

[SIGCOMM08] **Ionut Trestian**, Supranamaya Ranjan, Aleksandar Kuzmanovic, Antonio Nucci. "Unconstrained Endpoint Profiling (Googling the Internet)" In Proceedings of ACM SIGCOMM 2008, Seattle, WA, August, 2008

[ToN10] **Ionut Trestian**, Supranamaya Ranjan, Aleksandar Kuzmanovic, Antonio Nucci. "Googling the Internet: Profiling Internet Endpoints via the World Wide Web" IEEE/ACM Transactions on Networking Volume 18 Number 2, 2010

[IMC09] **Ionut Trestian**, Supranamaya Ranjan, Aleksandar Kuzmanovic, Antonio Nucci. "Measuring Serendipity: Connecting People, Locations and Interests in a Mobile 3G Network" In Proceedings of USENIX/ACM SIGCOMM IMC 2009, Chicago, IL, November, 2009

[INFOCOM11] **Ionut Trestian**, Supranamaya Ranjan, Aleksandar Kuzmanovic, Antonio Nucci. "Taming User-Generated Content in Mobile Networks via Drop Zones" In Proceedings of IEEE INFOCOM 2011, Shanghai, China, April, 2011

[WYT] Vijay Erramilli, Kevin Huguenin, Nikolaos Laoutaris, **Ionut Trestian**. "WYT: Optimized Consistency for Geo-Diverse Online Social Networks". INRIA Technical Report July, 2010

[WR] WindRider www.cs.northwestern.edu/~ict992/mobile.htm

[NAF1] Mobile Broadband Measurement - Comments of The Benton Foundation, Columbia Telecommunications Corporation, Consumers Union, Native Public Media and New America Foundation https://prodnet.www.neca.org/publicationsdocs/wwpdf/7910naf.pdf

[NAF2] Mobile Broadband Measurement – Comments of Google
https://prodnet.www.neca.org/publicationsdocs/wwpdf/7910naf.pdf

[VOD] Vodafone response to European Commission Questionnaire on the Open Internet and Net Neutrality in Europe http://www.cs.northwestern.edu/~ict992/docs/vodafone.pdf