Large Scale Distributed Systems and Peer-to-Peer Technologies

Zoltán Ács, Ádám Agócs, Attila Balaton, Sándor Laki, Tamás Lukovszki

Department of Information Systems,
Faculty of Informatics, Eötvös Loránd University
{acszolta,agocs,balcsi4,lakis,lukovszki}@inf.elte.hu

Abstract

Peer-to-peer networks and social networks are very popular large scale applications. In this project we investigate the properties of such networks. Based on empirical exploration and mathematical modeling we develop efficient methods to ensure scalability, robustness, and reliability of such systems. We focus on the following areas:

- *Mobile peer-to-peer applications, mobile social networks*
- *Peer-to-peer content distribution*
- *Geolocation*

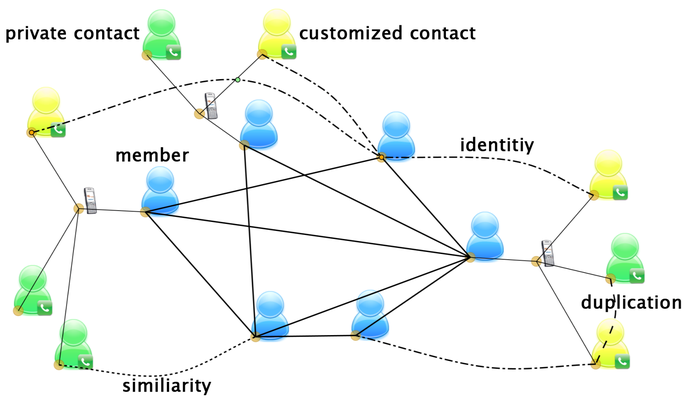
1. Mobile based social networks

Online social networks:

- More than 0.5 billion Facebook users (Feb. 2011)
- Facebook and LinkedIn are among the 10 most visited web sites (Feb. 2011)
- Goals: managing social realtionships, friends, sharing some content, etc...

Mobile based social networks:

- Hardware and software capabilities of mobiles allow them to access social networks
- Phonebooks of the mobiles also define social relationships
- Synchronizing the phonebook and the social network is benefitial
  - Identifying persons in the phonebook that are members of the social network,
  - update the data (semi-) automatically
- Resource requirement (and thus, scalability) mainly depends on the total number of similarities
  - A member of the network is *similar* to a phonebook contact if they have some equal fields, e.g. email, phone number, name, etc...



Our contributions:

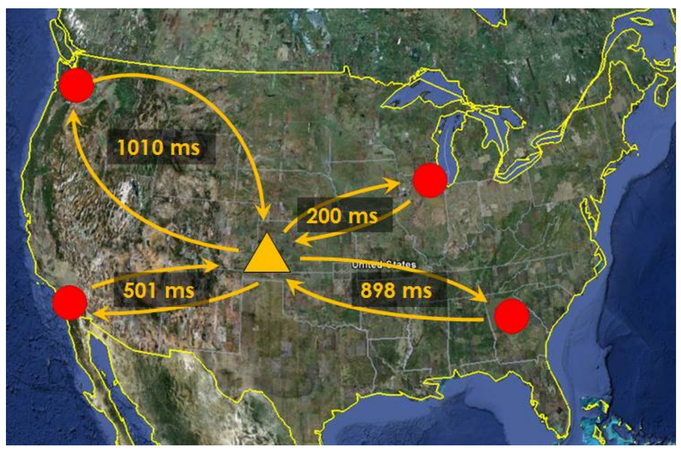
- Modeling and analyzing the structure of such networks, i.e. distribution of in-degree, out-degree, phonebook sizes, and similarities [1]
- Estimating the total number of similarities and scalability of such networks based on the above model. Showing the accuracy of our estimation mathematically and by measurements [2, 3]
- Examining registration delays in such networks and showing stability conditions based on queueing model [4]
- Our results are integrated in the mobile based social network Phonebookmark, which is property of Nokia Siemens Networks

2. IP Geolocation

- Localization of Internet hosts opens space for applications, from targeted, location aware content provision to localizing illegal content
- It provide benefits for network research and development, e.g. visualize the results of Internet measurements, exploring network topology, policy routing and pricing strategies
- Determining geographical location of Internet hosts by an IP address poses many challenges, since there is no direct relationship between the IP address of a network device and its geographic location

Geolocation methods:

- Registry based approaches store organizational information assigned to IP domains, or try to infer location information from DNS names. Usually the accuracy of these services is insufficient due to the lack of reliable information
- Active geolocation techniques approximate the geographic distance based on delay and topology measurements and then use triangulation-like methods



Our contributions:

- We focus on techniques mostly based on active probing
- In [5] we propose a path latency model for improving the accuracy of router localization
- In [6] we introduces a novel probabilistic geolocation approach based on the detailed statistical analysis of the relationship between network delay and geographic distance. Using this method a "prototype" service, called Spotter, was built which is available at http://spotter.etomic.org.
- In [7, 8] we show how this method can be used to examine the geographic aspects of the Internet that have been hidden so far (e.g. route similarity, link length distributions, geographic anomalies in network design, etc.).
- We designed and built a Web Service based measurement plane, called SONoMA (Service Oriented NetwOrk Measurement Architecture), http://sonoma.etomic.org [9]

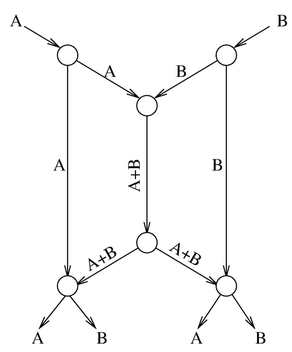
3. P2P content distribution and network coding

P2P content distribution:

- Peer-to-peer based content sharing is considered as one of the most efficient and popular content distribution solution

Network coding:

- The network nodes are able to combine arriving packets and transmit the combined packets over the links
- For a source node to a set of destination nodes, the rate of the multicast can achieve the cut bound (the minimum of the cuts separating the source from a destination) by using network coding. In general, this is not possible, if the packets can simply be routed or replicated.
- Network coding attracted a huge amount of theoretical and practical research since its first appearance
- Random network coding is a very popular. It builds random linear combinations of the packets at the nodes. Decoding is possible after receiving enough linearly independent combinations



Our contributions:

- We have proposed a method for increasing download efficiency in BitTorrent, in case of rare pieces by using network coding [10]
- We have introduced a method for early detection of peer-to-peer traffic [11]

References

[1] Ekler, P., Lukovszki, T.: Experiences with phonebook-centric social networks. In: Proc. 7th IEEE Consumer Communications and Networking Conference (CCNC). (2010) 412–416

[2] Ekler, P., Lukovszki, T.: The accuracy of power law based similarity model in phonebook-centric social networks. In: Proc. 6th International Conference on Wireless and Mobile Communications (ICWMC). (2010) 209–214

[3] Ekler, P., Lukovszki, T., Charaf, H.: Evaluating dynamically evolving mobile-based social networks. Acta Cybernetica **19(4)** (2010) 735–748

[4] Ekler, P., Lukovszki, T., Charaf, H.: Examining registration delays in mobile based social networks. In: Proc. 11th IEEE International Symposium on Computational Intelligence and Informatics (CINTI). (2010)

[5] Laki, S., Mátray, P., Hága, P., Csabai, I., Vattay, G.: A model based approach for improving router geolocation. Computer Networks **54(9)** (2010) 1490–1501

[6] Laki, S., Mátray, P., Hága, P., Sebők, T., Vattay, G., Csabai, I.: Spotter: A model based active geolocation service. In: Proc. IEEE INFOCOM. (2011)

[7] Mátray, P., Hága, P., Laki, S., Vattay, G., Csabai, I.: The network geography of the internet. In: Proc. IEEE INFOCOM. (2011)

[8] Mátray, P., Hága, P., Laki, S., Vattay, G., Csabai, I.: On the spatial properties of internet routes. In: To be published. (2011)

[9] Hullár, B., Laki, S., Stéger, J., Csabai, I., Vattay, G.: Sonoma: A service oriented network measurement architecture. In: Proc. TridentCom. (2011)

[10] Ekler, P., Lukovszki, T.: Extending mobile bittorrent environment with network coding (short paper). In: 8th IEEE Consumer Communications and Networking Conference (CCNC). (2011)

[11] Hullár, B., Laki, S., György, A.: Early identification of peer-to-peer traffic. In: Proc. IEEE ICC. (2011)

National Development Agency
www.ujszechenyiterv.hu
+36 40 638 638

HUNGARY RENEWS

The Project is supported by the European Union and co-financed by the European Social Fund (grant agreement no. TAMOP 4.2.1./B-09/1/KMR-2010-0003).