Measuring Mobile Peer-to-Peer Usage: Case Finland 2007

Mikko V.J. Heikkinen, Antero Kivi, and Hannu Verkasalo

TKK Helsinki University of Technology, P.O. Box 3000, FI-02015 TKK, Finland
{mikko.heikkinen,antero.kivi,hannu.verkasalo}@tkk.fi

Abstract. We study the development of mobile peer-to-peer (MP2P) traffic and the usage of MP2P applications in Finland during 2005-2007. Research data consists of 1) traffic traces measured from three Finnish GSM/UMTS networks covering the Internet-bound mobile data traffic generated by 80-90% of Finnish mobile subscribers (N > 4,000,000), and 2) usage log files collected with a dedicated Symbian handset monitoring application (N = 579). In the traffic trace measurement, we notice almost zero P2P file sharing traffic for handsets, but 9-18% of unidentified traffic, part of which possibly being P2P traffic. Potentially a notable growing trend for computer-based P2P file sharing traffic is visible in GSM/UMTS networks, BitTorrent and eDonkey being the most popular protocols. In the panel study, only Fring, a client to several P2P-based communication services, has significant usage and data volume levels.

Keywords: Measurement, mobile, peer-to-peer, traffic, usage.

1 Introduction

Peer-to-peer (P2P) traffic has been rapidly growing in the Internet during past years [1]. On the other hand, mobile devices and laptops are increasingly using GSM/UMTS (Global System for Mobile communications / Universal Mobile Telecommunications System) connections to access the Internet [2]. Our aim is to study the development of mobile P2P (MP2P) traffic and the use of MP2P applications in Finland during 2005-2007 from two perspectives: 1) by analyzing traffic traces measured at the GSM/UMTS networks of three major Finnish mobile operators, and 2) by investigating results from a panel study conducted with a handset monitoring application running in Nokia Symbian S60 operating system. In the traffic trace measurement, we identify P2P file sharing traffic based on TCP/UDP port numbers and use TCP fingerprinting to differentiate between computers and handsets. In the handset monitoring, we identify MP2P applications based on logged application name, and record the number of usage sessions and transferred data volume. Our approach is restricted to the end-user viewpoint.

No commonly accepted definition for MP2P exists. According to Steinmetz and Wehrle [3], P2P is “a system with completely decentralized self-organization and resource usage”. Androutsellis-Theotokis and Spinellis [4] attribute the following

S.B. Moon et al. (Eds.): PAM 2009, LNCS 5448, pp. 165–174, 2009.

characteristics to P2P systems: distribution, node interconnection, self-organization, resource sharing, adaptation to failures and transient populations, maintenance of acceptable connectivity and performance, and absence of global centralized servers or authority. We use the term "mobile" to describe both laptop computers and handsets with a data transfer connection to a GSM/UMTS mobile network. We cover both P2P systems consisting only or partially of mobile nodes, and mobile clients to P2P systems consisting only of fixed nodes. Of the handset applications we analyzed, only SymTorrent acts as a fully functional peer in a P2P system, other applications act as clients to P2P systems, i.e. relay traffic via intermediating peers or servers without implementing full peer capabilities.

Previous traffic measurement studies on MP2P in GSM/UMTS mobile networks are concentrated on the evaluation of using specific applications in limited test scenarios [5]-[9]. According to the best of our knowledge, our study is the first holistic examination of MP2P usage in GSM/UMTS mobile networks.

Our paper is structured as follows: first, we conduct a brief literature study on previous research. Then, we present our research methods, and the results of our measurements. Finally, we conclude our findings.

2 Previous Research

P2P traffic can be identified using various alternative methods. The simplest method is to investigate the port numbers in TCP or UDP packet headers [10], [11]. Some P2P applications use static port numbers to relay traffic, making their usage analysis fairly straightforward. However, an increasing number of P2P applications randomize or let their users randomize the port numbers in use. Therefore, in some studies various statistical methods have been developed to identify P2P traffic [12]-[16]. Other studies identify application-specific signatures in packet payload [17]-[19]. Guo et al. [20] analyze the log files generated by centralized components of a P2P network.

The point of measurement varies significantly in previous P2P measurement studies. Some analyzed a border point of an Internet Service Provider (ISP) network [10], [11], [13]. Others investigated a border point of an academic network [15], [17], [18]. Henderson et al. [19] collected packet-level traces and syslog messages from several access points to an academic Wireless Local Area Network (WLAN). Wu et al. [16] obtained traces from a company providing a streaming service based on a P2P network.

The reported metrics in the studies can be grouped into several categories. Traffic-based metrics include bandwidth consumption and traffic volume [11], [18], [19]; connection and session durations and latencies [10], [11], [18]; packet and flow level distributions [15], [16]; traffic patterns over time [10], [11], [13], [16]-[20]; and upstream and downstream traffic comparisons [10], [11] [16], [19]. Peer-related metrics consist of geographical distribution of peers [10]; number of downloads, uploads and shares by peers [8]; number of peers over time [14], [16], [20]; and peer connectivity and locality [10], [11].

3 MP2P Usage Measurements

3.1 TCP/IP Traffic Measurements

We use TCP/IP traffic measurements to collect traces of IP traffic in GSM/UMTS mobile networks. The measurements were conducted simultaneously at the networks of three major Finnish GSM/UMTS mobile network operators during two weeks in September-October 2005, 2006, and 2007. The measurements took place at a point between the Gateway GPRS Support Node (GGSN) and the Internet, at each of the three networks measured. In total, the measurements included the Internet-bound packet-switched data traffic of approximately 80-90% of Finnish mobile subscribers, i.e. over 4,000,000 subscribers. The measurements resulted in traces with the packet headers of Internet and transport layer protocols, whereas all application layer protocol headers were sanitized from the data.

The underlying operating systems of the end-user devices generating the traffic, for example Symbian and Windows, are identified by using a method called TCP fingerprinting [21]. Different operating systems are recognized by identifying idiosyncrasies in the implementation of their respective TCP/IP stacks. No application layer data is needed, as the method only uses certain TCP and IP header fields. Operating system identification using TCP fingerprinting is a fairly reliable method, but it leaves some of the traffic unidentified.

Different application protocols are identified from the traffic traces using transport protocol (TCP, UDP) port numbers. Furthermore, the analysis of P2P traffic by Karagiannis et al. [22] is used to classify specific port numbers as P2P traffic. Port-based identification of P2P traffic has some limitations. First, as many P2P applications select the used port numbers dynamically, identification based on the default port might leave a lot of traffic unidentified. Second, some P2P applications also use port 80 (HTTP) to masquerade their traffic as web traffic in order to, for example, pass simple firewalls or gain higher priority. Due to the limitations of the process, only the use of the following file sharing protocols is identified: BitTorrent, Direct Connect, eDonkey, FastTrack, Gnutella and Napster. Several applications can use the protocols and their variations.

Table 1. TCP and UDP port numbers used in P2P file sharing protocol identification

	TCP	UDP
BitTorrent	6881-6889	-
Direct Connect	411, 412	411, 412
eDonkey	4661, 4662	4665
FastTrack	1214	1214
Gnutella	6346, 6347	6346, 6347
Napster	6699-6702	6257

Table 1 summarizes the TCP and UDP port numbers used for identification. We also depict traffic generated by web and email protocols with the following TCP port numbers: HTTP (80), HTTPS (443) and HTTP alternate (8080) for web, and SMTP (25), POP3 (110), IMAP (143), IMAP/SSL (993) and POP3/SSL (995) for email.

3.2 Handset Monitoring

We utilize a handset-based research method in collecting data from end-users [23]. The method provides data on the actual usage of mobile applications and services, as the measurements are conducted directly in the device. End-users participating in the study install a client application on their device having a Nokia Symbian S60 operating system. The client runs as a background process, observing user actions and storing usage data into memory. The data collected includes application usage, data sessions, communication activities, memory status and Uniform Resource Locator (URL) traces, among others. The data is collected at the level of events and sessions, including accurate time stamps and identifiers for participating end-users. The data is transmitted daily to centralized servers for analysis. The method is deployed in controlled panel studies, to which typically a few hundred panelists are recruited.

The panel lasted for 1-2 months between November 2007 and January 2008. The panelists (i.e. users participating in the study) are provided with €20 vouchers as a compensation to potential data transfer costs they have to bear due to research and are entered to prize draws.

The main shortcoming of the method is the adverse selection of panelists. Typically technologically enthusiastic persons or people motivated by the prize draws participate in this kind of research panels. In addition, the Symbian device penetration is still well below 20% in the Finnish market [2], therefore the panelists could potentially be characterized as early-adopter users.

The panelists are recruited from the subscriber bases of three major Finnish mobile operators, targeting only consumer customers. 579 panelists from whom at least three weeks of usage-data is collected are included in the dataset. 44% of panelists are using Nokia S60 3rd edition devices, and 56% use older 2nd edition devices. 25% of the panelists have WLAN functionality in their handsets. 79 % of the panelists are male. The most dominant age groups are 20-29 years (38%) and 30-39 years (30%). Most panelists (77%) are employed. Over half (58%) of the panelists have a usage based data plan, the rest have a quota based plan (31%) or a flat rate plan (11%).

In the panel measurements, we identify four MP2P applications: Fring, iSkoot, MobileMule and SymTorrent. We have no knowledge whether the applications were installed by the users before or during the panel. Also, we do not know which applications the user had installed but did not use during the panel. We identify the applications by analyzing a list containing all the applications the panelists had used. The decision whether to include an application for detailed analysis is based on its recorded name. The detailed analysis consists of an Internet search, the aim of which is to determine whether the application can be considered a MP2P application. If we had used a different classification scheme, for instance included push-to-talk and instant messaging applications which use servers for both control and media traffic, we would have classified significantly more applications as MP2P applications.

4 Results

4.1 TCP/IP Traffic Measurement Results

According to the data provided by the mobile operators to Statistics Finland [24], [25], the volume of packet-switched data traffic in Finnish mobile networks has been growing rapidly during the recent years (2005: 34,000 GB, 2006: 100,000 GB, 2007: 500,000 GB). The rapid growth in data traffic volumes is partly explained by increased penetration of UMTS capable handsets [2], expansion of UMTS network coverage to smaller cities, and High-Speed Downlink Packet Access (HSDPA) upgrades to mobile networks. These developments have been accompanied by introduction of alternative flat rate mobile data subscriptions and heavy marketing of data cards and Universal Serial Bus (USB) data modems.

Our results from the measurement weeks in Falls 2005, 2006, and 2007 are consistent with the yearly figures reported by Statistics Finland. A fourfold increase in overall traffic volumes was observed between Falls 2005 and 2006, as both computer and handset (Symbian) originated traffic grew in proportion. However, between Falls 2006 and 2007 the traffic by computers grew by a factor of fourteen, whereas the growth of handset traffic was more moderate and merely tripled. The obvious difference between our results and the figures from Statistics Finland is explained by the fact that traffic volumes in Finnish mobile networks started to grow rapidly in late Summer 2007, likely due to the operators' aggressive marketing of flat rate mobile broadband subscriptions bundled with HSDPA-capable USB dongles for laptop computers. Overall, our measurements show that the relative share of computer traffic in mobile networks has grown from 70-75% in 2005-2006 to over 90% in 2007, whereas the traffic share of handsets running Symbian operating system has dropped from around 15% to about 4% of all traffic.

The profile of handset and computer traffic by application protocol is presented in Fig. 1 and 2. Clear differences in the application protocol profile of computers and handsets can be observed. Handset traffic is dominantly web browsing, whereas the share of email is significant albeit decreasing in relative terms. Other protocols, especially P2P protocols, are marginal traffic-wise. For computers, the relative share of web browsing and email traffic have been decreasing at the expense of the traffic that could not be identified, which amounts to almost 60% of all traffic in the latest traces. A small share (4-5%) of P2P traffic is also observed.

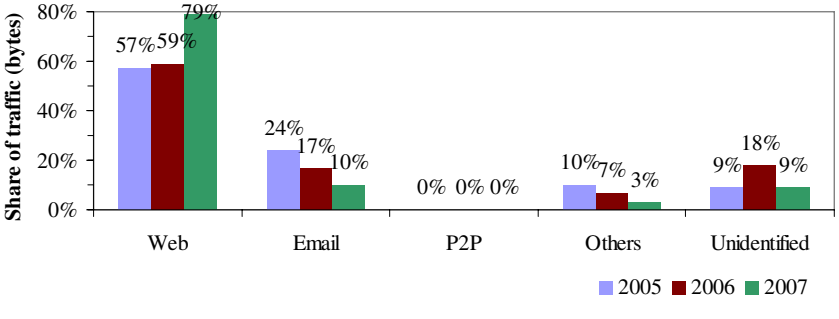

Fig. 1. Handset traffic by application protocol

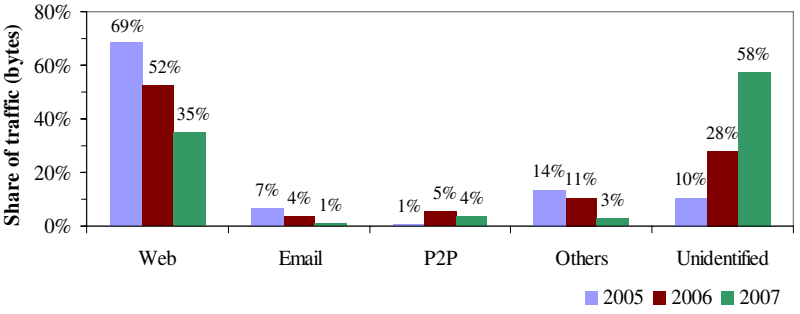

Fig. 2. Computer traffic by application protocol

The notable growth of unidentified traffic merits a further comparison with other traffic classes. Moreover, as P2P protocols are among the protocols typically using non-default port numbers, a first hypothesis would be that some of the unidentified traffic would in fact be P2P traffic. The diurnal distribution of computer traffic by application protocol for the measurement done in 2007 is presented in Fig. 3.

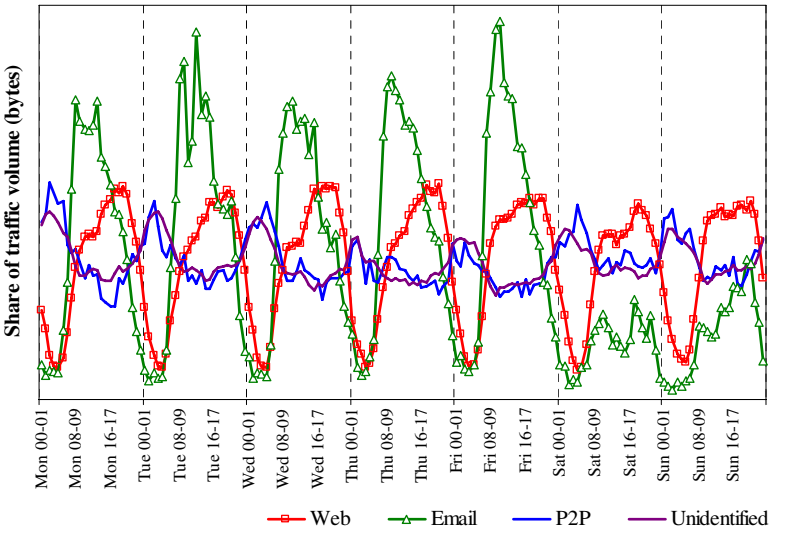

Fig. 3. Diurnal distribution of computer traffic by application protocol in 2007

The distributions of web and email traffic clearly reflect patterns of human behavior, i.e. high activity throughout waking hours and dramatically decreasing usage during the night. Moreover, the use of some applications (e.g. email) concentrates on office hours during the working days, whereas others (e.g. web browsing) are more free-time oriented. On the contrary, both P2P and unidentified traffic follow very similar patterns through the week, and the peak hours for both traffic types occur during the night.

Typically in fixed access networks P2P traffic is uniformly distributed over the whole day [1], as large media files get downloaded at the limit of available capacity without the need for any human input. In mobile networks, phone calls and non-P2P data usage consume most of the network capacity during the day, but the remaining capacity is consumed by P2P applications during the night. Further investigation of the traffic traces shows that the share of uplink traffic for P2P traffic (58%) is similar to unidentified traffic (52%), and significantly higher than for other applications (e.g. email: 32%, web: 13%). All this suggests that the true share of P2P traffic in the mobile network is considerably higher than the level proposed by pure port based identification.

A breakdown of P2P traffic by protocol is presented in Fig. 4. BitTorrent displays a growing trend; Direct Connect, Gnutella and Napster exhibit a decreasing trend; and eDonkey shows variation. However, as unidentified traffic potentially consists mostly of P2P traffic, the protocol profile is a result of the identification method, not of the true usage. In other words, the portion of traffic using non-standard port numbers is unidentified, therefore potentially biasing the distributions significantly.

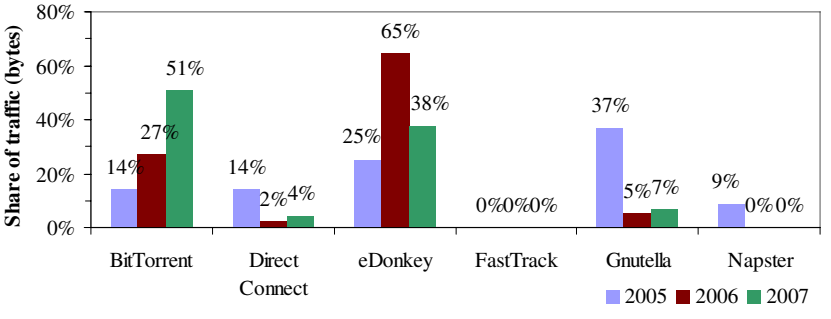

Fig. 4. Breakdown of P2P traffic by protocol

4.2 Handset Monitoring Results

In the panel study, three major categories of P2P applications are found: Voice over IP (VoIP) clients, instant messaging, and file sharing. Most of the identified P2P applications receive very little usage, typically having only a few users, which rarely activate the P2P application on a regular basis (see Table 2). Clearly MP2P applications are not popular. The average number of usage sessions per month per user is relatively high. The usage session is defined to consist of minimum 15 seconds from the activation of the application until its closure.

Table 3 presents the results of the data volume analysis in both GSM/UMTS and WLAN. MP2P applications experience very little actual usage. According to the results, people probably only experiment with MP2P applications. Only Fring exhibits wider usage, typically generating 4% of the total packet data volume of its users. On average Fring generates 1 MB per month per user. 12% of Fring traffic takes place on GSM networks (10% in EDGE and 2% in GPRS networks), and the share of UMTS networks is 16%. The share of WLAN traffic in Fring usage is 72%, possibly indicating cost avoidance behavior.

Table 2. Application usage by panelists

Application name	Used	Used every month	Used every week	Used twice per every week	Sessions per month per user
Fring	4.5%	3.5%	1.2%	1.9%	5.1
MobileMule	0.3%	0.3%	0.0%	0.2%	5.8
iSkoot	0.2%	0.0%	0.0%	0.0%	4.4
SymTorrent	0.2%	0.2%	0.0%	0.0%	6.0

Table 3. Data volume by panelists

Application name	Avg. proportion of the user's data volume	Avg. MB per month per user
Fring	4.0%	1.01
MobileMule	0.0%	0.03
SymTorrent	0.0%	0.04
iSkoot	0.0%	0.01

5 Conclusions

In the traffic trace measurement, we observe almost zero P2P file sharing traffic for handsets, but 9-18% of unidentified traffic, part of which possibly being P2P traffic. A growing trend for computer-based P2P file sharing traffic is visible in GSM/UMTS networks, BitTorrent and eDonkey being the most popular protocols. Again, a significant growing portion (10-58%) of traffic is left unidentified by our port number based identification method, potentially suggesting a noticeable increase of P2P traffic using random port numbers. Diurnal analysis of the traffic partially confirms this behavior. Only Fring, a client to several P2P-based communication services, has significant usage and data volume levels in the panel study.

Relative P2P traffic growth and changes in P2P protocol distributions in our traffic trace measurements is dependent on global P2P trends. For instance, some of the studies discussed earlier depicted a coherent trend of growth in BitTorrent usage which is also visible in our study. The other trend seen in our study common to other studies is the relative growth of unidentified traffic potentially consisting of masqueraded P2P traffic. The absolute traffic growth is probably explained by the following significant changes in the Finnish mobile market: the active marketing of flat rate data tariffs in 2006 and USB data modems in 2007 to the consumer market by major mobile operators.

In the panel study, many kinds of metrics on service usage can be derived straight from the device. The shortcomings of the panel study include the adverse selection of panelists and the amount of data available. If the dataset contained more P2P data sessions, the data traffic patterns could be studied in detail, including the distinction between WLAN and cellular P2P use.

Further research could include applying more advanced P2P traffic identification methods to traffic trace measurements, such as Domain Name Server (DNS) based analysis and detection according to statistical identification functions. The more

advanced methods could facilitate the detection of other types of P2P protocols than file sharing using random port numbers, for instance Skype [26]. While applying these methods, the potential peculiarities posed by the GSM/UMTS and WLAN networks, such as longer access delays, should be taken into consideration. The panel results could be refined by having a more precise approach to MP2P application identification and by analyzing possible correlations between respondents' demographics and usage profiles.

Acknowledgments. We would like to thank Markus Peuhkuri and Timo Smura for their assistance in analyzing the data, and Heikki Kokkinen for his comments on a draft version of this paper. This research has been conducted as part of the COST IS0605 framework.

References

1. Haßlinger, G.: ISP platforms under a heavy peer-to-peer workload. In: Steinmetz, R., Wehrle, K. (eds.) Peer-to-Peer Systems and Applications. LNCS, vol. 3485, pp. 369–381. Springer, Heidelberg (2005)
2. Kivi, A.: Mobile Data Service Usage Measurements: Results 2005-2007. Technical report, TKK Helsinki University of Technology (2008)
3. Steinmetz, R., Wehrle, K. (eds.): Peer-to-Peer Systems and Applications. LNCS, vol. 3485, pp. 9–16. Springer, Heidelberg (2005)
4. Androutsellis-Theotokis, S., Spinellis, D.: A Survey of Peer-to-Peer Content Distribution Technologies. ACM Computing Surveys 36, 335–371 (2004)
5. Hofeld, T., Tutschku, K., Andersen, F.U.: Mapping of File-Sharing onto Mobile Environments: Enhancement by UMTS. In: Proc. IEEE Pervasive Computing and Communications Workshops, pp. 43–49 (2005)
6. Hossfeld, T., Tutschku, K., Andersen, F.U.: Mapping File Sharing onto Mobile Environments: Feasibility and Performance of eDonkey over GPRS. In: Proc. IEEE Wireless Communications and Networking Conference, pp. 2453–2458 (2005)
7. Hoßfeld, T., Binzenhöfer, A.: Analysis of Skype VoIP Traffic in UMTS: End-To-End QoS and QoE Measurements. Computer Networks 52, 650–666 (2008)
8. Matuszewski, M., Beijar, N., Lehtinen, J., Hyyryläinen, T.: Content Sharing in Mobile P2P Networks: Myth or Reality. Int. J. Mobile Network Design and Innovation 1, 197–207 (2006)
9. Matuszewski, M., Kokkonen, E.: Mobile P2PSIP: Peer-to-Peer SIP Communication in Mobile Communities. In: Proc. Fifth IEEE Consumer Communications & Networking Conference, pp. 1159–1165 (2008)
10. Plissonneau, L., Costeux, J.-L., Brown, P.: Analysis of peer-to-peer traffic on ADSL. In: Dovrolis, C. (ed.) PAM 2005. LNCS, vol. 3431, pp. 69–82. Springer, Heidelberg (2005)
11. Sen, S., Wang, J.: Analyzing Peer-to-Peer Traffic Across Large Networks. IEEE/ACM Transactions on Networking 12, 219–232 (2004)
12. Guha, S., Daswani, N., Jain, R.: An Experimental Study of the Skype Peer-to-Peer VoIP System. In: 5th International Workshop on Peer-to-Peer Systems (2006)
13. Karagiannis, T., Broido, A., Faloutsos, M., Claffy, K.C.: Transport Layer Identification of P2P Traffic. In: Proc. Internet Measurement Conference, pp. 121–134 (2004)

14. Ohzahata, S., Hagiwara, Y., Terada, M., Kawashima, K.: A traffic identification method and evaluations for a pure P2P application. In: Dovrolis, C. (ed.) PAM 2005. LNCS, vol. 3431, pp. 55–68. Springer, Heidelberg (2005)
15. Schmidt, S.E.G., Soysal, M.: An Intrusion Detection Based Approach for the Scalable Detection of P2P Traffic in the National Academic Network Backbone. In: Proc. Seventh IEEE International Symposium on Computer Networks, pp. 128–133 (2006)
16. Wu, C., Li, B., Zhao, S.: Characterizing Peer-to-Peer Streaming Flows. IEEE J. on Selected Areas in Communications 25, 1612–1626 (2007)
17. Bleul, H., Rathgeb, E.P.: A simple, efficient and flexible approach to measure multiprotocol peer-to-peer traffic. In: Lorenz, P., Dini, P. (eds.) ICN 2005. LNCS, vol. 3421, pp. 606–616. Springer, Heidelberg (2005)
18. Gummadi, K.P., Dunn, R.J., Saroiu, S., Gribble, S.D., Levy, H.M., Zahorjan, J.: Measurement, Modeling, and Analysis of a Peer-to-Peer File-Sharing Workload. In: Proc. ACM Symposium on Operating Systems Principles, pp. 314–329 (2003)
19. Henderson, T., Kotz, D., Abyzov, I.: The Changing Usage of a Mature Campus-Wide Wireless Network. Computer Networks 52, 2690–2712 (2008)
20. Guo, L., Chen, S., Xiao, Z., Tan, E., Ding, X., Zhang, X.: A Performance Study of Bittorrent-Like Peer-to-Peer Systems. IEEE J. on Selected Areas in Communications 25, 155–169 (2007)
21. Smith, C., Grundl, P.: Know Your Enemy: Passive Fingerprinting. Technical report, The Honeynet Project Know Your Enemy Whitepapers Series (2002)
22. Karagiannis, T., Broido, A., Brownlee, N., Claffy, K.C., Faloutsos, M.: File-Sharing in the Internet: A Characterization of P2P Traffic in the Backbone. Technical report, University of California, Riverside (2003)
23. Verkasalo, H., Hämmäinen, H.: A Handset-Based Platform for Measuring Mobile Service Usage. INFO 9, 80–96 (2007)
24. Official Statistics of Finland: Telecommunications 2006. Statistics Finland, Helsinki (2007)
25. Official Statistics of Finland: Telecommunications 2007. Statistics Finland, Helsinki (2008)
26. Baset, S.A., Schulzrinne, H.: An Analysis of the Skype Peer-to-Peer Internet Telephony Protocol. Technical report, Columbia University, New York (2004)