Estimation of a Mobile IPv6's Home Agent Load

Mihai Cristian[1], Albert Cabellos-Aparicio[2], Rares Cosma[1], Jordi Domingo-Pascual[2],
Pedro Vale Pinheiro[3], Fernando Boavida[3], Virgil Dobrota[1]

Abstract – The Home Agent (HA) is a key point when considering the performance of Mobile IPv6-based networks. This paper presents the first steps towards characterizing the load of a HA. This may be useful both for researchers that aim to propose novel architectures that improve the performance of the HAs and for ISPs willing to deploy Mobile IPv6. To achieve our goals first we analyzed the internal traffic of a medium-size department. Then we reviewed the existing models and we evaluated their applicability to this particular scenario. Our results showed that the estimated load of a HA serving a medium-size department (around 1500 hosts) is high with a maximum throughput of 262 Mbps.

Keywords: Home Agent, Mobile IPv6

I. INTRODUCTION

Wireless technologies have rapidly evolved in recent years. IEEE 802.11 is one of the most used wireless technologies and it provides up to 54 Mbps of transfer rate in an easy an affordable way. In the current Internet status a user can be connected through a wireless link, but he/she cannot move (i.e. change its access router) without breaking the IP communications. With mobility, a user can move and change the point of attachment to the Internet without losing the network connections. In Mobile IP designed by IETF a Mobile Node (MN) has two IP addresses. The first one, HoA (Home Address), is for identity, while the second one, CoA (Care-Of Address) identifies the MN's current location. The mobile node will always be reachable through its HoA while it will change its CoA according to its movements. A special entity called HA (Home Agent) placed at the home network will maintain bindings between the MN's HoA and CoA addresses. The main limitation of Mobile IP is that communications between the node and its peers are routed through the HA. This means that a HA might be responsible of multiple MNs on a Home Link. The failure of a single HA may then result in the loss of connectivity of numerous MNs. Thus, HAs represent the possibility of a single point of failure in Mobile IP-based networks. Moreover MN's communications through the HA may also lead to either the Home Agent or the Home Link becoming the bottleneck of the system. In addition, the HA's operation such as security check, packet interception and tunneling might not be as optimized in the software as plain packet forwarding. Mobile IP comes into two flavors, Mobile IPv4 [13] and Mobile IPv6 [14]. The last one outperforms the first one in many aspects. For instance Mobile IPv6's clients can communicate directly with their peers. This means that these communications are not forwarded through the HA and this reduces the communications delay and the load. Unfortunately communications to/from the Home Network must be forwarded through the HA. This paper focuses on a Mobile IPv6's HA. The research community has focused on solving these issues proposing novel architectures that improve both the performance and the reliability of the HAs [15,16,17,18]. Although they are very effective, Mobile IPv6 has not been widely deployed yet. This means that the load of a Mobile IPv6's HA is unknown. Hence the proposed architectures might have been evaluated with an unrealistic load. On the other hand ISP's willing to deploy Mobile IPv6 needs an estimation of the expected load that the deployed Home Agents will have. This paper presents the steps towards characterizing the load of a Mobile IPv6 Home Agent. We believe that this is important for both the research community and the industry. First, we analyzed the internal traffic of a medium-size department from UPC (Universitat Politecnica de Catalunya). We assumed that all the inside hosts were Mobile IPv6 nodes that they were away of their Home Network. Thus all the internal traffic had to be processed by a hypothetical Mobile IPv6's HA. Second we reviewed existing models of traffic that applied to this particular scenario. Specifically we focused on the models that characterized the load of Wireless LAN networks [1,2], evaluating its goodness-of-fit for our particular case. Finally we analyzed at which granularity these

models could be applied. The rest of the paper is organized as follows: Section II presents an overall look of the datasets used in terms of internal versus external traffic, transport and application protocols. Section III contains an empirical characterization of the load of our Home Agent. Section IV reviews other models that fit with our estimates for the aggregate traffic. Section V looks to the traffic generated per sub-net and Section VI contains our conclusions.

II. DATA ACQUISITION AND TRAFFIC BREAKDOWN

In this section we presented an overview of the datasets used for this study and we were focused on the structure of the traffic that is relevant for the goal of this paper. We described how the traffic was divided (internal and/or external) and we examined the main components of the traces on the transport and application layers.

A. Internal versus External

The data analyzed comprises NetFlow records from a department router at UPC Barcelona, for six days of traffic, from March 9 to March 14, 2007 both inside the department and to/from the Internet. For the 144 hours of traffic monitored we found a total of 903.7 Gigabytes. Fig. 1 shows that 95% of all flows, octets and packets belong to traffic to and from the Internet and only 5% represents the traffic between hosts inside the department.
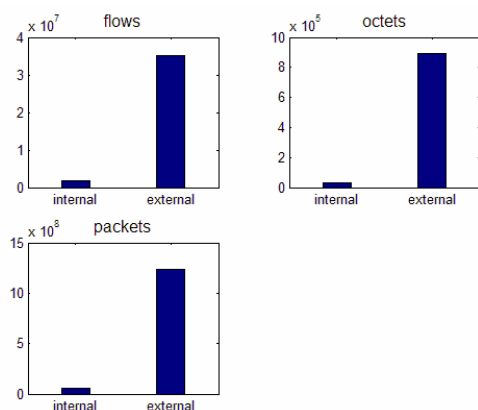

Fig. 1 Internal versus external

We based our study on the assumption that all clients inside the department are Mobile IPv6 clients. Following this the traffic between clients inside the department (labeled internal) should be routed by the Home Agent. Traffic to and from the Internet (labeled external) was delivered to its destination directly using IPv6 extension headers and the Return Routability procedure [11]. As our aim was to characterize the load of the Home Agent we disregarded the external traffic. It is worth noting here that Mobile IPv6 clients could also communicate directly with its peers, even if they were at its Home

Network. However these communications had to be secured and this was achieved by routing them through the Home Agent. Additionally this router did not affect the performance of the communications since the Home Agent and the peers of the Home Network were very close. The number of hosts (Mobile IPv6 clients) inside the department sending and receiving packets was 1547. They accounted for a total of 32.95 Gigabytes of traffic on 1773751 flows. We provided an analysis of this traffic the next sections. These hosts were divided into 7 different subnets, each with a number of hosts ranging from 140 to 220 (subnets 1 to 7). The addresses of the hosts were made anonymous.

B. Traffic Breakdown

In order to efficiently characterize the load of the Home Agent we must first understand the makeup of the traffic that we are analyzing. To do this we chose a similar approach to the one presented in [3, 8], that is examining the traffic on the transport and application layer. In Table 1 we broke down the traffic by transport protocol in terms of flows, octets and packets. Most of the bytes were sent, as expected, using the TCP protocol. The percent of the ICMP traffic increased during the second and third day, Saturday and Sunday, but this is only because the total traffic on these days was less then the average (less department's staff was present during the weekend). The TCP and UDP traffic remained fairly constant during all these days. We found that the bulk of the traffic was sent using the TCP protocol for reasons explained below.

Table 1. Fraction of flows, bytes and packets

Protocol	Flows [%]	Bytes [%]	Packets [%]
17(udp)	28.75	0.81	3.73
6(tcp)	67.87	99.1	95.84
1(icmp)	3.37	0.05	0.42

Next we were focused on the application layer. To classify the traffic inside the department we grouped the applications into several high-level categories, as in Table 2.

Table 2. Applications and their protocols

Category	Protocols
Bulk	ftp, https, tftp, rtip
e-mail	smtp, imap, pop, brutus, pop3s
interactive	telnet, ssh
Name	Dns,netbios-ns
net-manage	dhcp, ntp, epmap, snmp, timed
web	http, https
windows	netbios-ssn, netbios-dgm
authentication	ident
printing	ipp
streaming	hp-pdl-datastr

They have been identified using the flow destination port. Only the important applications are shown, the minor ones being ignored. The main flows found are consistent with the deployed services in the department.

Fig. 2 shows the application usage for the aggregated traffic (all seven subnets). The percentage ratio changes depending on the week day, as explained earlier. We can clearly see that email and interactive (especially SSH in our case) were the most used applications both in terms of flows and packets. In terms of flows email applications represented roughly 50% and SSH only 5%, whereas in terms of octets SSH accounted for almost 60% percent of the traffic,

which indicates that it was used for large file transfers within the period monitored. Other applications showed a normal ratio between the flows and octets percentage. The figure also reflects the findings in Table 1. Most of the application layer protocols found used the TCP protocol (web, email, SSH). This explains the 99% bytes transmitted through TCP, whereas in terms of flows only 67% used this transport protocol. The reason is that network management applications are based on UDP protocol.

Finally it is worth noting here that [3] showed a similar analysis regarding internal enterprise traffic and found the same pattern of large numbers of bytes traversing the network on a small number of flows.
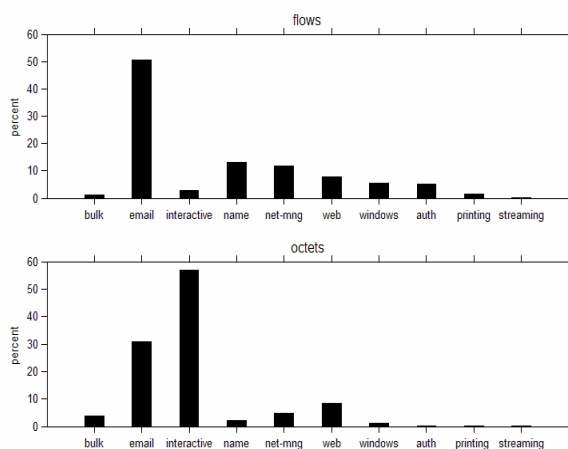

Fig. 2 Application usage in terms of flows and bytes

III. EMPIRICAL ESTIMATION OF THE HOME AGENT'S LOAD

A. Flow arrivals

In this section we aimed to provide an empirical characterization of the traffic that a Home Agent in a medium size department had to process. The department contained a number of 1,547 hosts (assumed as IPv6 clients), which accounted for a total of 32.95 GB of traffic during the one week interval. In order to characterize this load we computed a series of parameters such as number of flows per second, flow size, flow inter-arrival times, throughput in bytes per second and packets per second. We started by looking at the flow arrival process to see if we can observe any patterns that could be helpful in our characterization. Fig. 3 plots the time series of the flow arrivals for the entire network using one hour bins. The plot shows sharp increases in the number of flow arrivals in the morning with peaks at 28,000 flows per hour during weekdays and 9,000 flows per hour during the weekend.
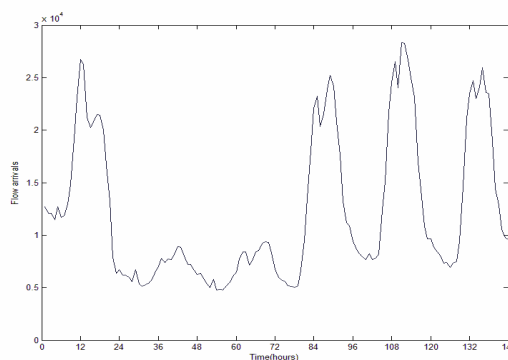

Fig. 3 Flow arrivals pattern

B. Average flows per second

Another important flow-level variable for the traffic load is the number of flows per second. This parameter represents the number of active flows that, for each second, a HA has to handle. This allowed us to see what kind of loads our Home Agent should

expect to process when deployed in a similar environment to that of the department. Fig. 4 plots the CDF of the average flows per second for each of the six days of traffic. The plots show a clear distinction between the weekdays and the weekend as observed before. For the week days we found a mean value of 4.67 flows per second and a maximum of 19.7 flows per second, while during the weekend the mean dropped to 1.9 flows per second. The large difference between the weekday and weekend plots also indicates that modeling this variable using a single parametric distribution is rather difficult. Throughput parameters were also computed in this scenario for all the days using one minute interval for averaging. Table 3 summarizes our findings regarding these particular parameters. For each day we computed the mean and maximum throughput in bits and packets per second (bps and pps) and the mean and maximum active flows per second (fps). The mean and maximum values for the throughput are shown in Mbps.

Table 3. Throughput values for the aggregated traffic

Day	Mean Mbps	Max Mbps	Mean pps	Max pps	Mean fps	Max fps
1	1.39	102.95	217.3	12.66K	4.73	20.8
2	0.03	3.28	18.2	0.29K	1.9	18.41
3	0.04	4.35	22.36	7.54K	1.89	14.98
4	0.68	61.58	123.69	7.33K	4	35.75
5	0.63	262.24	124.26	24.66K	4.67	19.7
6	0.9	123.48	191.58	14.61K	4.27	12.46

We can see that during the six day monitoring period our hypothetical HA had to process mean values of up to 1.39 Mbps and 217.63 pps. Maximum values reached 262.24 Mbps and 24.66 kpps on day 5. We could also distinguish a pattern that holds for all the mean values in the table. Both bps and pps mean values started from a low value in the weekend (days 2 and 3) and increased gradually, peaking on the last day of the week (day 1).
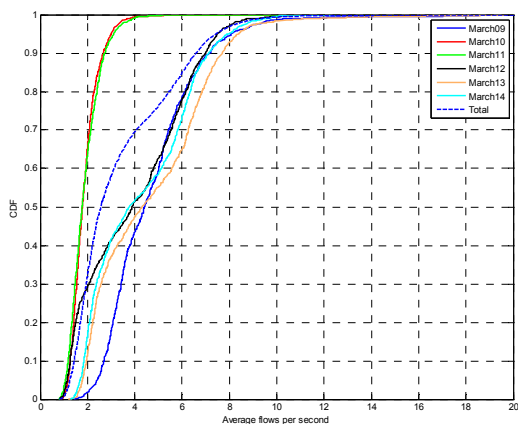

Fig. 4 Distribution of average flows per second

C. Flow inter-arrival time

The next component we looked at the flow inter-arrival time distribution. This parameter is essential for estimating a Kendall queuing model [12] for our deployed HA and was computed in milliseconds. Fig.5 plots the distribution for the six days monitored. We can see that there was still a difference between the weekdays and weekends but it was much smaller than in the case of flows per second or throughput. Mean values for the flow inter-arrival time were 279 ms for the weekdays and 618 ms for the weekend, which was consistent with the plot in Fig.3 where we see a smaller number of flows arriving during the weekend.
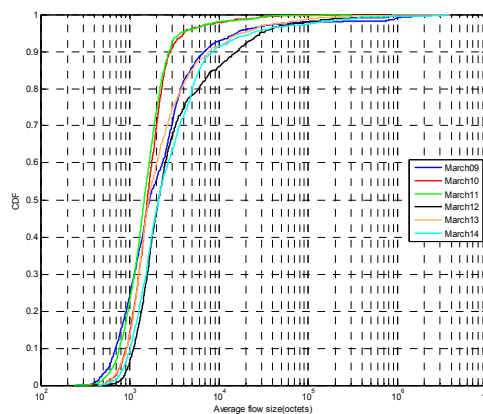

Fig. 5 CDF plot of flow inter-arrivals

D. Flow size

In order to provide a good characterization of the load of our hypothetical home agent it is not enough to look at the flow arrival process but also at the flow sizes to find out how much each flow carries in terms of bytes. Average flow sizes were computed for one minute intervals, for everyday of traffic, by dividing the total number of octets to the total number of flows inside the bin. The average flow size distribution is plotted in Figure 6. Flow sizes showed mean values of 10.4 Kilobytes per flow during the weekdays and 2.47 Kilobytes per flow in the weekend.
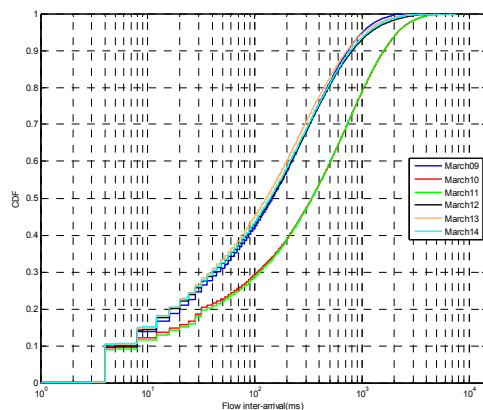

Fig. 6 Average flow size distribution for all days of traffic

Summarizing, this section attempted to provide an empirical characterization of the load of a HA. With a total of roughly 1500 Mobile IPv6 nodes, a HA would have to process a mean value of 3.57 active flows per second and should handle mean throughput values of 0.67Mbps and 120 pps. Maximum throughput values of 262.24 Mbps indicate that high processing power is needed to deploy as HA in a similar scenario. Also we have spotted patterns in the flow arrival process, flow inter-arrival times and the average flow-sizes for hour traffic which indicate that these variables can be modeled using parametric distributions.

IV. DATA MODELING

To the best of our knowledge no studies exist aiming to characterize the load of a Mobile IPv6 Home Agent. Nevertheless other researchers have characterized similar loads. In this section we aim to evaluate if these existing models fit on our empirical data. A similar analysis to ours is found in [1]. However the authors attempted to model the traffic for a campus WLAN, not for a Home Agent. Their approach was based on two levels of modeling: the session and the flow level. Flow arrivals were considered as a cluster process triggered by session arrivals, which was not the case for our study. We focused entirely on the flow level. Additionally, the load characterized in [1], as well as in our analyzed data, was highly dependent on the applications deployed. Their usage differed when looking at a WLAN or a HA load which could led to different load characteristics. We will try to see if the models proposed fit with our empirical data.

We found that flow arrivals pattern (Fig.3) is very similar to the session arrivals pattern depicted in [1]. This leads us to believe that models for other flow variables might apply in our scenario as well. Variables on the flow level were modeled for WLAN traffic using the following distributions: flow-size using a Pareto distribution and flow inter-arrival time using the log-normal distribution.
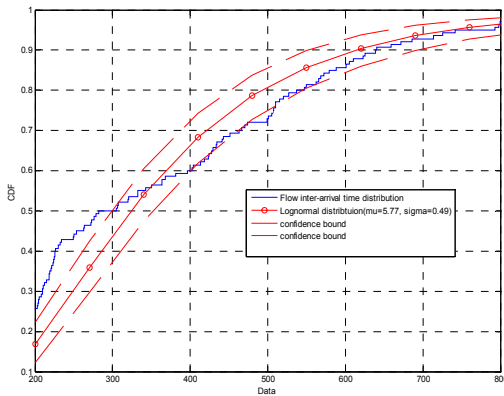

Fig. 7 CDF plot of Lognormal distribution plotted against empirical distribution of flow inter-arrival times

Fig.7 plots the flow inter-arrival empirical distribution against the proposed LogNormal distribution with the following parameters: mean=365.72, mu(log location)=5.77 and sigma(log scale)=0.49, with it's 95% confidence bounds. The empirical distribution remained within the confidence bounds of the LogNormal distribution, but in the tail section, which contained mostly extreme values, it varied from lognormality. When attempting to fit the proposed distribution in [1] for the flow size empirical data we came across the same issues.

Fig.8 plots the empirical distribution of the flow sizes with is confidence bounds against the proposed Pareto distribution with the following parameters: mean=3099.9, k(shape)=0.63, sigma(scale)=1234.18 and theta =635. Again we can see that the proposed model provided a good fit for our empirical data, with the same problem in the tail of the curve where the empirical distribution has a higher skew. We also found that the generalized extreme value distribution provided a slightly better fit especially for the mentioned tail section because this section contained mostly extreme values of the plotted data. Our findings indicate that flow variables for the load of a HA can be modeled using statistical distributions. Existing models found in [1] provided a good fit for our empirical data and could be used to model the load of a hypothetical HA in a scenario with similar parameters (number of hosts and deployed services).
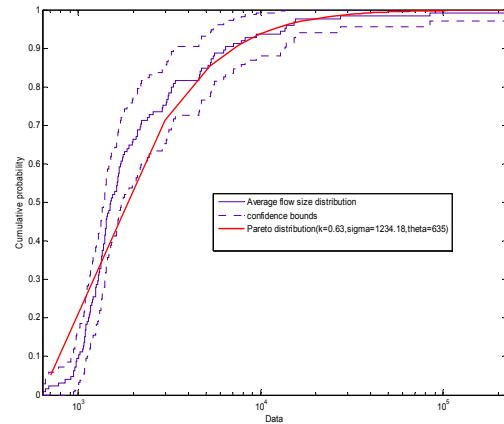

Fig. 8 Generalized Pareto distribution plotted against empirical distribution of average flow sizes

V. SUB-NETWORK TRAFFIC

In this section we tried to find patterns in the per-subnet traffic and fit them to the above mentioned models. This could be useful for adapting the models to suit a scenario with any given number of subnetworks. The same parameters presented in Section III were computed for each of the seven subnetworks in the department using one minute intervals as stated before. Fig.9 plots the average flow size distribution for each of the seven subnetworks using dotted lines and the average flow size for the aggregated traffic using a solid line.
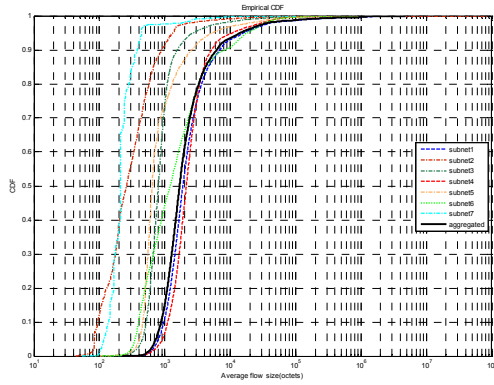
Fig. 9 Average flow-size distributions for all subnets

For this parameter we can see that subnet 1 and 4 followed the Pareto distribution of the aggregate traffic. Other subnets could be grouped together according to their distribution, for instance subnets 5 and 3 had similar flow size distributions, but subnet 6 and subnets 7 and 2 did not resemble the model of the aggregated traffic. In terms of flow inter-arrival times, plotted in Fig.10, we see a higher resemblance between the distribution of the subnets and the LogNormal distribution of the aggregated traffic. As before, subnets 1 and 4 showed the closest match whereas subnet 2 showed a completely different distribution. The large difference between distributions for sub-networks could be explained by the fact that they had different application usage patterns. A similar analysis to the one in Section II revealed that subnets 1 and 4 presented the highest resemblance to the overall application usage (Fig.2), whereas other subnets showed a completely different pattern. Also subnet 1 and subnet 4 had the highest number of clients. This leads us to believe that the flow-level traffic models presented for the aggregated traffic could work for subnets with a higher number of clients. A higher number of clients led to a smoother curve for the evaluated parameters and an application usage pattern that complied with the one presented for the aggregated traffic.
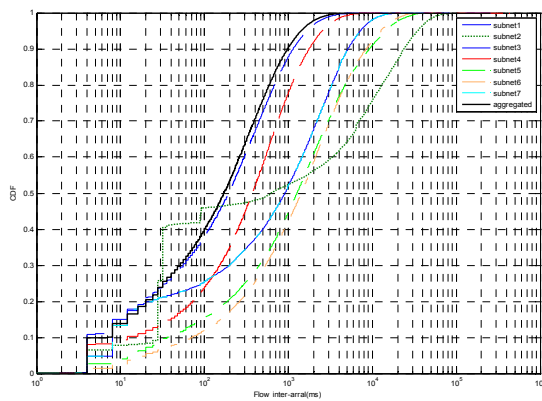

Fig. 10 CDF plot of flow inter-arrivals per sub-network

VI. CONCLUSIONS AND FUTURE WORK

In this paper we presented the first steps towards characterizing the load of a Mobile IPv6's HA. In order to do this we measured the internal traffic of a medium size department and we assumed that all the Mobile IPv6 hosts were away from home network. We evaluated the load that a Mobile IPv6 HA would have to process. Our results showed that in a medium size network comprising roughly 1500 clients a HA is looking at mean values of 3.57 active flows per second and maximum throughput values of 262.24 Mbps and 24.6 Kpps. This indicates that high processing power is needed for a HA deployed in a similar scenario. Next, we found that variables such as flow-size, inter-arrival times and flow arrivals fit currently deployed models for the aggregated traffic even if these models have been developed for characterizing WLAN traffic and not the load of a HA. The following variables have been modeled: average flow sizes using a Pareto distribution and flow inter-arrival times using a Lognormal distribution. In the final part we looked at the per subnet traffic for similar patterns. We found that the same distributions can be used to model the traffic at the subnet level for subnets with a high number of clients. Further work is needed to match subnet traffic patterns with aggregated traffic ones.

REFERENCES

[1] F.Hernandez-Campos, M.Karaliopolus, M.Papadopouli, H.Shen "Spatio-Temporal Modelling of Traffic Workload in a Campus WLAN", WICON 2006
[2] M. Karaliopolus, M. Papadopouli, E.Raftopolous, H.Shen. "On scalable measurement-driven modelling of traffic demand in large WLANs", LANMAN 2007
[3] R.Pang, M.Allman, M.Bennet,"A first look at modern enterprise traffic", IMC 2005
[4] A.Sinha, K.Mitchell, D.Methi. "Flow level upstream traffic behaviour in broadband access networks: DSL versus Broadband fixed wireless", IPOM 2003
[5] C.E. Perkins." Mobile IP', IEEE Communications Magazine, May 1997
[6] Mark Fulmers flow-tools
[7] Z.Sun, D.He, L.Liang. "Internet QoS and traffic modelling", IEE 2004
[8] M.Ploumidis, M.Papadopouli, T.Karagiannis. "Multi-Level applicationbased traffic characterization in a large scale wireless network", WOWMOM 2007
[9] K.Thompson, G.J.Miller, R.Wilder."Wide-Area Traffic Patterns and Characteristics", 1997
[10] C.E.Perkins, D.B. Johnson. "Mobility Support in IPv6", MOBICOM 1996
[11] K.J.Lee, J.Park, H.Kim. "Route optimization for mobile nodes in a mobile network based on prefix delegation", VTC 2003
[12] N.Vandaele, T.Van Woensel, A.Verbuggen. "A queueing based traffic model", March 2000
[13] D.Johnson et al. "Mobility Support in IPv6", RFC 3775, 2004
[14] C. Perkins: IP Mobility Support for IPv4 RFC 3344 (2002)
[15] F. Heissenhuber et al. "Home Agent Redundancy and Load Balancing in Mobile IPv6", Int.Conf Broadband Comm. 1999
[16] H. Deng et al "Load Balance for Distributed Home Agents in Mobile IPv6", IEEE PIMRC 2003
[17] R. Wakikawa et al "Virtual mobility control domain for enhancements of mobility protocols" INFOCOM 2005
[18] A. Cabellos-Aparicio et al., "A flexible and Distributed Home Agent Architecture for Mobile IPv6-based Networks" IFIP Networking 2007