Minimum Delay Routing in Multihop Wireless Networks

Maggie X. Cheng[1], Xuan Gong[1], and Peng-Jun Wan[2]

[1] Department of Computer Science,
Missouri University of Science and Technology, Rolla, MO 65401
{chengm,xgwt2}@mst.edu
[2] Department of Computer Science,
Illinois Institute of Technology, Chicago, IL 60616
wan@cs.iit.edu

Abstract. End-to-end delay is an important QoS metric in multihop wireless networks such as sensor networks and mesh networks. Along with throughput, end-to-end delay determines the user-experienced data transmission time. End-to-end delay refers to the total time it takes for a single packet to reach the destination. It is a result of many factors including the length of the route and the interference level along the route, and therefore both the routing scheme and the MAC layer scheduling scheme can affect end-to-end delay. We assume a deterministic scheduling scheme is used at the MAC layer. Since the actual delay depends on the MAC layer scheduling algorithm, at the network layer we try to reduce the interference on the path instead of the actual delay time. To find the routing solution that minimizes path interference, a sufficient condition on conflict-free transmission is established, which helps to quantify the interference on a link. A linear program based on the sufficient condition is developed to compute the routing solution. Through simulation, we show that the proposed routing scheme can effectively reduce end-to-end delay.

1   Introduction

With the increasing application of wireless mesh networks and sensor networks, multihop wireless networking technology is expected to not just provide multihop connectivity in locations where wired networks cannot reach, but also to support user traffic with certain service guarantees. End-to-end delay is one of the major metrics for quality of service. The user-perceived data transfer time is a combined effect of both data rate and end-to-end latency. For transferring a small file, the dominating factor is end-to-end latency; for transferring a large file, the dominating factor is data rate. In a typical sensor network, where small packets generated by sensors need to be periodically reported to the base station, end-to-end delay plays a more important role. This paper aims to address how to achieve the minimum end-to-end delay for regular traffic through routing.

We assume a deterministic scheduling scheme is used at the MAC layer. For the network in Fig. 1, the route in (a) leads to longer delay than the route in

(b) since (b) is the shortest path and there is no interference from other data flows. But the shortest path does not always lead to the minimum delay. In fact, end-to-end delay is a result of both the number of hops on the path, and the interference level along the path. The shortest path leads to the minimum delay only if the shortest path is the least interfered path.
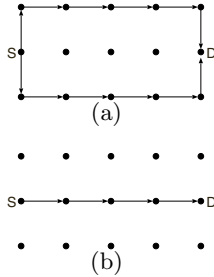

Fig. 1. (a) With maximum throughput routing, latency is 6 slot-time; (b) With minimum delay routing, latency is 4 slot-time

To see how delay increases with interference, we compare the delay from the two scenarios in Fig. 2. If there is only one data flow from source S1 to destination D1, end-to-end latency is 6 slots; However, if there are other transmissions nearby, the end-to-end latency of the same flow can be increased to 10 slots. Results are obtained by using the FCFS scheduling policy, so a packet is scheduled to use the next available slot as soon as it arrives; each slot is used to transmit one packet.
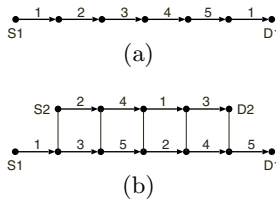

Fig. 2. (a) With a single data flow, latency is 6 slot-time; (b) When other transmitters are active, the latency becomes 10 slot-time. Numbers on links are slot numbers. There are 5 distinct slot numbers.

End-to-end delay is a result of many factors including the length of the route and the interference level along the route, and therefore both the routing scheme and the MAC layer scheduling scheme can affect end-to-end delay. The scope of this paper is on routing only. Without knowledge of the MAC layer scheduling algorithm, we can only indirectly reduce delay. In this paper, we relate end-to-end delay to the accumulated interference along the path, and compute the routing solution that minimizes path interference. We quantify the interference level on a link to be the number of mutually conflicting transmissions involving

this link. Therefore the interference on a link is closely related to the scheduling delay on this link, and the accumulated interference on a path is related to the end-to-end delay on the path. In fact, the accumulated interference defined this way is the upper bound of the end-to-end delay on the path by some TDMA scheduling scheme. By minimizing the accumulated interference through routing, we can indirectly achieve the minimum delay routing.

The rest of the paper is organized as follows. In Section 2, we briefly survey the related work on interference modeling and delay optimization in recent years; in Section 3, we present the necessary background for interference modeling in the optimization model, and establish a sufficient condition for conflict-free transmission; in Section 4, we present the linear programming based routing scheme; in Section 5 we show the effectiveness of the proposed scheme through extensive simulation. Finally, Section 6 concludes the paper.

2   Related Work

End-to-end delay is an important performance metric in multihop wireless networks. Delay optimization has been achieved through routing path selection, multiple access control, and sometimes joint design of physical layer and MAC layer.

MAC layer solutions explore the broadcast nature of wireless transmission and improve the end-to-end delay through transmission scheduling or transmission probability control. Chatterjee et al. [1] presented when the routing tree is given for a sensor network, how to determine the time slot of each node such that the maximum latency to send a packet from a node to the sink is minimized. Chaporkar et al. [2] addressed the MAC layer multicast problem as an instance of the stochastic shortest path problem and developed an optimal transmission strategy for minimum delay multicasting. Sarkar et al. [3] addressed the energy-delay tradeoff problem and formulated the problem as a constrained optimization problem that achieves the minimum energy while satisfying the constraint on average packet delay. The optimization problem was solved by using dynamic programming formulation, from which a closed form expression for the optimal sleep duration is derived. Pereira et al. [4] addressed delay optimization problem for a random access MAC protocol. They presented an accurate analytical model to derive the optimal transmission probability of each mobile node that minimizes the delay.

Network layer solutions consider delay as one of the constraints while computing the routing path. In [5], Sivrikaya et al. presented an algorithm to compute the minimum-delay path for networks with STDMA. Wan et al. [6] presented approximation algorithms for minimum latency aggregation in sensor networks. The minimum latency problem is to compute an aggregation tree for sensor nodes so that the makespan of the aggregation schedule is minimum. The authors proved the problem is NP-hard and proposed efficient approximation algorithms. Li et al. [7] studied how to select the routing path with the minimum end-to-end delay in multi-radio wireless mesh networks and developed routing protocols

for both single-channel and multi-channel wireless mesh networks. Alzahrani and Woodward [8] proposed a localized QoS routing algorithm by using statistics collected locally and avoided the overhead incurred in most global routing schemes.

Network performance is inherently related to parameters of multiple layers. In addition to those separate-layer solutions mentioned above, there are some research work using cross-layer schemes for delay optimization. Cui et al. [9] considered joint design of network layer, MAC layer and physical layer for delay optimization. They optimized the routing flow, TDMA slot assignment and MQAM modulation rate and power on each link to minimize worst-case packet delay. The cross-layer optimization problems are approximated by convex optimization problems and efficiently solved. Xia et al. [10] used fuzzy logic system in cross-layer design and considered physical layer, data link layer and application layer together. Pakdehi et al. [11] introduced cross layer design between MAC layer and physical layer to optimize the overall system throughput while preserving packet average delay time. Xiao [12] investigated joint design of network-coding and channel-coding to achieve the optimal delay performance through the tradeoff design between the network layer and the physical layer.

Delay optimization is the aim of this paper, while interference modeling is the means to achieve this goal. Interference modeling is a critical part of this paper. Throughput and delay optimization both rely on accurate modeling of wireless interference. Interference modeling was first studied in the context of throughput optimization. The most related work includes [13, 14, 15, 16, 17]. Jain et al. [13] first used conflict graphs to model the effect of wireless interference under a simplified protocol model; Qiu et al. [15] continued to use conflict graphs to model interference under IEEE 802.11 interference model; Further in [16], Qiu et al. proposed a physical interference model, which is based on measured interference rather than distance between nodes. In addition to interference modeling, Padhye et al. [18] focused on the estimation of interference and studied the effect of interference on aggregated network throughput based on IEEE 802.11 model. The interference modeling method used in our paper is different from all previous work in that it is based on the IEEE 802.11 protocol model but uses a sufficient condition instead of the accurate condition on cliques to capture all mutually conflicting relations among transmissions.

3   Preliminaries

The discussion in this section is in the context of a TDMA scheme at the MAC layer. We assume each slot can be used to transmit one package on a link and get acknowledgement from the receiver. To make sure all transmissions are conflict free, it is important that all active links that are mutually conflicting with each other use different slots. In other words, no two links can use the same slot if they interfere with each other.

A collision domain is defined as a group of links that are *mutually* conflicting with each other. In previous work [13], a conflict graph has been used to model

the conflicting relation between wireless links. A conflict graph representing a wireless network can be built in polynomial time as follows: we use vertices to represent wireless links, and then put an edge between two vertices if the wireless links they represent interfere with each other. To list all collision domains in a network requires to build a conflict graph first and then to find all cliques in the conflict graph. Although to build the conflict graph can be done in polynomial time, to find all cliques in the graph is an NP-hard problem. To avoid solving an NP-hard problem, we will find a sufficient condition to replace the clique condition.

Suppose link (k, l) is a two-hop neighbor of link (i, j) (Adjacent links are considered as one-hop neighbors). Let $N2_{ij}$ denote the group of two-hop neighbors $\{(k, l)\}$ that satisfy: the distance from its one end to link (i, j) is one, and the distance form the other end to link (i, j) is at most three via a different path. If there is no alternative path to link (i, j), then the distance is counted as ∞.

For each link (i, j), we define CD_{ij} to be the set of links that includes the following links:

1. link (i, j) itself, and
2. all one-hop neighbors (adjacent links) of (i, j), and
3. the two-hop neighbors of (i, j) defined in $N2_{ij}$.

Obviously CD_{ij} has polynomial number of links in it as defined above, and there are $O(V^2)$ CD_{ij}'s in a network. Therefore to list all links in $\{CD_{ij} | (i, j) \in E\}$ takes only polynomial time.

The following theorem presents a sufficient condition to achieve conflict-free transmission.

Theorem 1. *If the following condition is satisfied for all link $(i, j) \in E$, then all transmissions in the wireless network is conflict free: At any time t, among all links defined in CD_{ij} there is at most one link active at time t.*

The detailed proof of the theorem is omitted here due to the space limit. Briefly speaking, it is a sufficient condition because the set $\{CD_{ij} | (i, j) \in E\}$ includes all cliques on the conflict graph. If there is at most one link active in CD_{ij}, and this condition holds for any link (i, j), and then there is at most one link active in any clique.

The implication of this theorem is that for any scheduling algorithm at the MAC layer, at any time slot, as long as it has at most one link active in CD_{ij} for any link (i, j), then there is no conflict. This conclusion leads to another theorem.

Theorem 2. *Let I_{ij} be the total number of active transmissions on links defined in CD_{ij}. Given a routing path $\mathcal{P}$, there exists a conflict-free TDMA scheduling scheme such that the end-to-end delay achieved by it is bounded from above by $\sum\limits_{ij \in \mathcal{P}} I_{ij}$.*

I_{ij} is an indicator of the interference level on link (i, j), so $\sum\limits_{ij \in \mathcal{P}} I_{ij}$ is the accumulated interference on the routing path. This theorem relates the number of

active transmissions in a link's collision domain to delay, and allows us to use the accumulated interference on the path as the upper bound of end-to-end delay. Due to the space limit, the design of the TDMA algorithm and the proof for the upper bound is omitted. Following this direction, the minimum delay routing becomes minimum interference routing.

4   Minimum Interference Routing (MIR)

4.1   The Model

Following Theorem 2, we use the total interference along the path as an indicator of delay. This is a better measure of delay than the number of hops, since the same path can experience different amount of delay if the congestion level is different.

We define variable $X_{ij,f} = 1$ if link (i, j) is on the routing path of flow f. If the flow can be split, $X_{ij,f}$ can be fractional and hence represent the probability of using link (i, j) for flow f. $R_{ij,f}$ is the data rate of link (i, j) allocated to flow f. R_f, as input constant, is the source rate of flow f. Each flow is specified by a *(source, destination)* pair. Since the source and destination of each flow can be arbitrarily set, it is possible that one node serves as the source for multiple flows with each having a different destination, so we use $R_{i,f}$ to denote the source rate of node i for flow f. Since our ultimate goal is to reduce the end-to-end delay of a unit flow, we set constant R_f the same for all flows. B is the wireless link capacity. I_{ij} roughly gives some indication of the interference level at the location of link (i, j) caused by its own transmission and conflicting links' transmissions. I_f is the total interference of all links along the routing path of flow f.

In equality (4c), we approximately quantify I_{ij} as the total "busy time" of link (i, j) caused by the transmission on link (i, j) itself and nearby transmissions. We include all links in CD_{ij} as defined in section 3, instead of all links within two hops from (i, j). The reason is that some two-hop neighbors of link (i, j), namely, $(k1, l1)$ and $(k2, l2)$, if they are not mutually conflicting with each other, can be active at the same time slot. Therefore the number of slots that (i, j) cannot use due to their interference should not be counted twice. By including the transmission on link (i, j), we have considered the effect of path length; by including the transmissions on other links in CD_{ij}, we have considered the interference that link (i, j) received from nearby transmissions.

The minimum interference routing (MIR) can be modeled as follows:
Minimize

$$\sum_f I_f \tag{1}$$

Subject to
(1) Flow conservation

$$\sum_{j \in N_i} (R_{ij,f} - R_{ji,f}) = R_{i,f}, \quad \forall i, f \tag{2a}$$

$$R_{ij} = \sum_f R_{ij,f}, \quad \forall \texttt{link}(i,j) \tag{2b}$$

$$R_{i,f} = R_f, \quad \forall f, i \text{ is } Source(f) \tag{2c}$$

$$R_{i,f} = -R_f, \quad \forall f, i \text{ is } Sink(f) \tag{2d}$$

$$R_{i,f} = 0, \quad \forall f, i \neq Source(f), Sink(f) \tag{2e}$$

(2) Bandwidth constraint

$$r_{ij} + \sum_{l \in N_i, l \neq j} r_{il} + \sum_{k \in N_j, k \neq i} r_{jk} + \sum_{(k,l) \in N2_{ij}} r_{kl} \leq B, \quad \forall \, \texttt{link}(i,j) \tag{3a}$$

$$r_{ij} = R_{ij} + R_{ji}, \quad \forall \texttt{link}(i,j) \tag{3b}$$

(3) Interference modeling

$$R_{ij,f} = R_f \cdot X_{ij,f}, \quad \forall \, flow \, f \tag{4a}$$

$$I_f = \sum_{link(i,j)} I_{ij} \cdot X_{ij,f}, \quad \forall \, flow f \tag{4b}$$

$$I_{ij} = r_{ij} + \sum_{l \in N_i, l \neq j} r_{il} + \sum_{k \in N_j, k \neq i} r_{jk} + \sum_{(k,l) \in N2_{ij}} r_{kl}, \quad \forall \texttt{link}(i,j) \tag{4c}$$

$$0 \leq X_{ij,f} \leq 1, 0 \leq R_{ij} \leq B. \tag{4d}$$

4.2   Linearization

The above formulation is not linear since both I_{ij} and $X_{ij,f}$ in constraint (4b) are variables. A close approximation to the above formulation is to minimize the total interference from all links, i.e.,

Minimize

$$\sum_{link(i,j)} I_{ij} \tag{5}$$

Subject to   All constraints but (4b).

Without constraint (4b) the alternative formulation is linear. Minimizing total interference from all links can indirectly reduce the end-to-end delay.

An iterative heuristic to solve the nonlinear program (1) is provided as follows:
1. solve the alternative linear program (5) to get $X_{ij,f}$ and I_{ij}, set objective value $Z_0 = \sum_f I_f$;
2. use I_{ij} as link weight; for each source, use the shortest path algorithm to find the path with the smallest total interference;
3. update $X_{ij,f}$ and I_{ij}; set $Z_1 = \sum_f I_f$;
4. if $Z_1 < Z_0$, set $Z_0 = Z_1$ and repeat step 2-4; otherwise stop.

4.3   Routing Path Reconstruction and Rounding

If the flow is allowed to split, then $X_{ij,f}$ is real-valued, so the above linear program can be solved in polynomial time; If the flow cannot be split, then $X_{ij,f}$ is a 0-1 integer variable, in which case we first relax it to a fractional linear program and then use the following algorithm to reconstruct the routing paths. The fractional $X_{ij,f}$ indicates the fraction of flow f that is carried on link (i, j).

PATHRECONSTRUCTION$(G(V, E), X)$

1   **for** each flow f, initialize $D_f = 1$, let $s = Source(f)$, $d = Sink(f)$
2   **do** find the minimum value $x_{min} = \min_{ij}\{X_{ij,f} \neq 0\}$
3   construct a path p_f $(s \rightsquigarrow d)$ using links (i, j) with $X_{ij,f} \geq x_{min}$
4   update $D_f = D_f - x_{min}$, $X_{ij,f} = X_{ij,f} - x_{min}$ for each link (i, j) on p_f
5   update the collection of paths $P_f = P_f \cup p_f$, $flow(p_f) = x_{min}$
6   iterate lines 2–5 until $D_f = 0$

Picture 3 shows $X_{ij,f}$ on each link, obtained from solving the relaxed linear programming problem. For instance, in Fig. 3(a), link (SA) has 60% of flow and link (SD) has 40% of flow for flow $s \rightsquigarrow d$. We first get path p1=s-D-C-B-d with 0.4 unit of flow, then get path p2=s-A-C-E-d with 0.6 unit of flow. After we reconstructed all routing paths for flow f, select the path p_f^* with the largest $flow(p_f)$, set all links (i, j) on p_f^* with $X_{ij,f} = 1$, and set all other links with $X_{ij,f} = 0$. By this way, we have selected the path with the largest fraction of flow as the main routing path and shut down other branches. Fig. 3(b) shows the result after rounding.
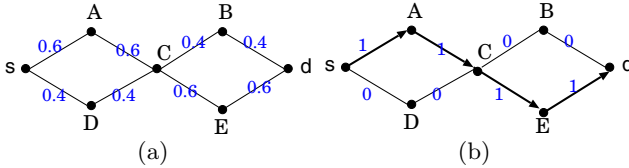

Fig. 3. Routing path reconstruction based on the computed fractional values of $X_{ij,f}$

5   Simulation

We now compare the proposed Minimum Interference Routing (MIR) with Shortest Path Routing (SPR) when they use the same MAC layer scheme. Since the actual delay time depends on the specific MAC layer scheduling scheme, we assume the most commonly used FCFS scheduling scheme is used.

In the simulation study, we use 50 nodes deployed on a 150×150 square region, with node transmission range 30. 10 out of the 50 nodes are randomly selected as source nodes, and all source nodes transmit to a common receiver (sink node). Each source node generates a packet and we observe the end-to-end

latency of the single packet. In FCFS, the packet arrival order is random. A relay
node schedules a packet as soon as it arrives; when deciding which slot to use, a
relay node chooses the next available slot to transmit the packet if it does not
conflict with other scheduled transmissions. FCFS is one of the most commonly
used scheduling policy in practice. Since the packet arrival order is an important
factor to FCFS, for each given network topology we run 25 tests on 25 random
arrival orders. We tested over 30 network topologies.

SPR chooses the shortest path (in hops) to reach the destination, while MIR
chooses the path that has the minimum interference along the path, with the
objective function specified in 1 in section 4. Simulation results show that MIR
always outperforms SPR. The performance gain is between 18% to 29% when
FCFS is usedfor scheduling. Fig. 4 shows the performance data. Delay is mea-
sured in the number of slots, and each slot is equivalent to the transmission time
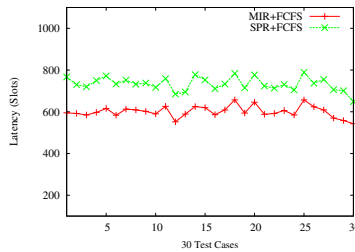for one packet.

Fig. 4. MIR vs. SPR comparison

6   Concluding Remarks

In this paper, we studied how to achieve the minimum end-to-end delay in a
multihop wireless networks. We presented an optimization model in which the
impact of wireless interference is considered. A sufficient condition for conflict-
free transmission is established, and a linear program model for minimizing end-
to-end delay is developed using this condition as a constraint. The simulation
results show that model-based optimization did achieve shorter delay than the
existing routing scheme Shortest Path Routing.

The sufficient condition is used in a global optimization framework in this
paper, but it can also be used locally for dynamic scheduling, in which the
condition only needs to be satisfied within the 2-hop neighborhood of a link.
Moreover, it can also be used for resource reservation and admission control as
part of QoS provisioning in wireless networks. We will address this issue in the
future work.

References

1. Chatterjee, P., Das, N.: A cross-layer distributed tdma scheduling for data gathering with minimum latency in wireless sensor networks. In: Wireless VITAE 2009, pp. 813–817 (May 2009)
2. Chaporkar, P., Sarkar, S.: Minimizing delay in loss-tolerant mac layer multicast. In: Third International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks, WIOPT 2005, pp. 358–367 (April 2005)
3. Sarkar, M., Cruz, R.L.: A mac layer power management scheme for efficient energy delay tradeoff in a wlan. The International Journal of Computer and Telecommunications Networking 51, 1–6 (2007)
4. Pereira, M., Bernardo, L., Dinis, R., Oliveira, R., Carvalho, P., Pinto, P.: Delay optimization on a p-persistent mac protocol for a multi-packet detection in sc-fde system. In: Wireless Communications and Networking Conference, WCNC 2010, pp. 1–6. IEEE, Los Alamitos (2010)
5. Sivrikaya, F., Yener, B.: Minimum delay routing for wireless networks with stdma. Wireless Networks 15(6), 755–772 (2007)
6. Wan, P.J., Huang, S.C.H., Wang, L., Wan, Z., Jia, X.: Minimum-latency aggregation scheduling in multihop wireless networks. In: MobiHoc 2009, pp. 185–194 (2009)
7. Li, H., Cheng, Y., Zhou, C., Zhuang, W.: Minimizing end-to-end delay: A novel routing metric for multi-radio wireless mesh networks. In: IEEE INFOCOM 2009, pp. 46–54 (April 2009)
8. Alzahrani, A., Woodward, M.: End-to-end delay in localized qos routing. In: 11th IEEE Singapore International Conference on Communication Systems, ICCS 2008, pp. 1700 –1706 (November 2008)
9. Cui, S., Madan, R., Goldsmith, A., Lall, S.: Cross-layer energy and delay optimization in small-scale sensor networks. IEEE Transactions on Wireless Communications 6(10), 3688–3699 (2007)
10. Xia, X., Ren, Q., Liang, Q.: Cross-layer design for mobile ad hoc networks: energy, throughput and delay-aware approach. In: Wireless Communications and Networking Conference, WCNC 2006, vol. 2, pp. 770–775. IEEE, Los Alamitos (2006)
11. Pakdehi, A.G., Ashtiani, F.: Cross-layer optimization of adaptive modulation and coding preserving packet average delay time. In: IEEE GLOBECOM 2008, pp. 1–5 (December 4, 2008)
12. Xiao, M.: Cross-layer design of rateless random network codes for delay optimization. In: IEEE International Conference on Communications, ICC 2010, pp. 1–6 (May 2010)
13. Jain, K., Padhye, J., Padmanabhan, V.N., Qiu, L.: Impact of interference on multihop wireless network performance. In: MobiCom 2003: Proceedings of the 9th Annual International Conference on Mobile Computing and Networking, pp. 66–80. ACM, New York (2003)
14. Rangwala, S., Gummadi, R., Govindan, R., Psounis, K.: Interference-aware fair rate control in wireless sensor networks. In: SIGCOMM 2006: Proceedings of the 2006 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications, pp. 63–74. ACM, New York (2006)
15. Li, Y., Qiu, L., Zhang, Y., Mahajan, R., Zhong, Z., Deshpande, G., Rozner, E.: Effects of interference on wireless mesh networks: Pathologies and a preliminary solution. In: HotNets 2007 (November 2007)

16. Qiu, L., Zhang, Y., Wang, F., Han, M.K., Mahajan, R.: A general model of wireless interference. In: MobiCom 2007: Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking, pp. 171–182. ACM, New York (2007)
17. Li, Y., Qiu, L., Zhang, Y., Mahajan, R., Rozner, E.: Predictable performance optimization for wireless networks. In: SIGCOMM 2008: Proceedings of the ACM SIGCOMM 2008 Conference on Data Communication, pp. 413–426. ACM, New York (2008)
18. Padhye, J., Agarwal, S., Padmanabhan, V.N., Qiu, L., Rao, A., Zill, B.: Estimation of link interference in static multi-hop wireless networks. In: IMC 2005: Proceedings of the 5th ACM SIGCOMM Conference on Internet Measurement, p. 28. USENIX Association, Berkeley (2005)