ISSN 2250-3153

A Data Path Quality Estimation Based on Delay Time Using Probes

K. Sivasankari, S. Jenila, M. Nizar Ahamed

IT Dept
Periyar Maniammai University, Vallam, Thanjavur.

Abstract- In most of the network data transformation sender nodes choose the link depends on link quality that implies sender could not much concentrate on the traffic and data loss. So this paper mainly focus on both data loss and queuing delay to identify most congested link in a network. Identifying the existence of a dominant congested link is useful for traffic engineering. It also helps us understand and model the dynamics of the network since the behavior of a network with a dominant congested link differs dramatically from one with multiple congested links. The network data congested mainly made by the router overhead and the intruder path selection; by consider these things here introduce a congested identification method by combining hypothesis test with model based approach. Here developing parameter inference algorithms for hidden Markov model and Markov model with a hidden dimension to infer this virtual delay. The process is more efficient than the existing methods and can implement in future for secure high throughput achieving group communication.

Index Terms- bottleneck link, dominant congested link, end–end inference, hidden Markov model (HMM), markov model with a hidden dimension (MMHD).

I. Introduction

Measurement and inference of end–end path characteristics have attracted a tremendous amount of attention in recent years. Properties such as the delay and loss characteristics of an end–end path [7], the minimum capacity and available bandwidth of a path [3], [5], and the stationarity of the network have been investigated. These efforts have improved our understanding of the Internet. They have also proved valuable in helping to manage and diagnose heterogeneous and complex networks.

In this paper, we study a specific end–end path characteristic, namely whether a *dominant congested link* exists along an end–end path. Informally, a dominant congested link is one that produces most of the losses and significant queuing delays on an end–end path. We avoid using the term "bottleneck link" since it has been defined in many different ways in the literature and there is no consensus on its meaning. Later in the paper, we relate our definition of dominant congested link to the notion of bottleneck link.

Identifying the existence of a dominant congested link is useful for traffic engineering. For example, when there are multiple paths from one host to another and all are congested, improving the quality along a path with one dominant congested link may require fewer resources than those along a path with multiple congested links.

Identifying whether a path has a dominant congested link also helps us understand and model the dynamics of the network since the behavior of a network with a dominant congested link differs dramatically from one with multiple congested links. When a dominant congested link exists, identifying the existence of such a link requires distinguishing its delay and loss characteristics from those of the other links. Achieving this goal via direct measurements is only possible for the organization in charge of that network. However, commercial factors often prevent an organization from disclosing the performance of internal links. Furthermore, as the Internet grows in both size and diversity, one organization may only be responsible for a subset of links on an end–end path. Some measurement techniques obtain internal properties of a path by using ICMP messages to query internal routers.

Traceroute and *ping* are two widely used tools in this category. Some more advanced techniques use ICMP messages to measure per-hop capacity or delay and pinpoint faulty links [4]. These approaches, however, require cooperation of the routers (to respond to ICMP messages and treat them similarly as data packets). Contrary to direct measurements using responses from routers, a collection of network tomography techniques infers internal loss rate and delay characteristics using end-end measurements [1]. Most tomography techniques, however, require observations from multiple vintage points.

II. Related Work

A dominant congested link is a link that produces most losses and significant queuing delays on an end–end path. Since most applications (TCP-based or real-time applications) are adversely affected by losses and delays, a dominant congested link is a form of "bottleneck link." One notion of bottleneck link is *tight link*, i.e., the link with the minimum available bandwidth; another notion is *narrow link*, i.e., the link with the minimum capacity [2]. Several studies focus on locating tight or narrow links.

We precisely define dominant congested link and differentiate it from other notions of bottleneck link in Section IV. After identifying a dominant congested link, we further derive an upper bound of the maximum queuing delay of that link, which is an important path characteristic and is complementary to other tools that estimate the available bandwidth or the minimum link capacity of a path [8], [9]. Network tomography infers internal link properties through end–end measurements. A rich collection of network tomography techniques have been developed in the past (see [1] for a review). Many techniques rely on correlated measurements (through multicast or striped unicast probes). More recently, several studies use uncorrelated measurements to detect lossy links [6], estimate loss rates, or locate congested segments that have transient high delays. Most tomography techniques, however, require *many* vintage points, while we only need measurements between two end-hosts along a *single* path.

The work closest in spirit to ours is the loss pair approach that is used to discover network properties. A loss pair is formed when two packets are sent close in time and only one of the packets is lost. Assuming that the two packets experience similar behaviors along the path, the packet not lost in a loss pair is used to provide insights on network conditions close to the time when loss occurs. Our work focuses on determining whether a dominant congested link exists along a path. Furthermore, our model-based approach differs significantly from the loss pair approach: Our approach *infers* the properties of the lost packets by utilizing delay and loss observations jointly and the correlation in the entire observation sequence, instead of using direct measurements from the loss pairs.

III. PROPOSED METHOD

In this paper, we propose a novel *model-based* approach to identify whether a dominant congested link exists along an end–end path using end–end measurements. We periodically send probes from one host to another so as to obtain a sequence of delay and loss values. The key insight in our approach is to utilize the queuing delay properties of the *lost* probes. For example, if one link along the path is solely responsible for all losses, then all lost probes have the property that they "see" a full queue at this link. We interpret a loss as an *unobserved delay* and discretize the delay values. Afterwards, we model the discretized delay sequence of all probes including those with missing values to infer whether a dominant congested link exists.

Our model-based approach has the following advantages. First, it utilizes delay and loss observations *jointly* for inference instead of the common approach of treating them separately. Second, it utilizes the correlation in the *entire* observation sequence instead of the very limited temporal correlation present in back-to-back packets. As we will see, the identification procedure only requires a short probing duration (in minutes).
The following are the main contributions of this paper.

- We present a formal yet intuitive definition of dominant congested link and provide two simple hypothesis tests to identify whether a dominant congested link exists along a path.
- Our model-based approach fully utilizes the information from the probing packets and enables very fast identification. Validation using *ns* simulation and Internet experiments demonstrates that this approach can correctly identify the existence of a dominant congested link in minutes.
- We provide a statistical upper bound on the maximum queuing delay of a dominant congested link once we identify such a link exists.

IV. DOMINANT CONGESTED LINK

In this section, we formally define dominant congested link and relate it to the widely used term "bottleneck link." For ease of reference, the key notation is summarized in Table I.

Consider K links/routers along an end–end path, as shown in Fig. 1. Each link/router is modeled as a droptail queue with a processing rate equal to the link bandwidth, and the maximum queue size equal to the buffer size of the router. Let Q_K denote the maximum queuing delay at queue k, i.e., the time required to drain a full queue. Then Q_K is determined by the buffer size and the link bandwidth of queue k. Probes are sent periodically from the source to the destination in a time interval $[t_1, t_2)$.

TABLE I
KEY NOTATION

Notation	Definition
K	Number of links/queues along the path
Q_k	The maximum queuing delay at queue k
D_t^k	Queuing delay for virtual probe t at link k
D_t	Aggregate queuing delay for virtual probe t over all the links along the path
L_k	Set of virtual probes marked as lost at link k
L	Set of virtual probes with loss marks
F_k	Set of virtual probes that experience the maximum queuing delay Q_k at link k
F	Set of virtual probes that experience the maximum queuing delay at some link along the path

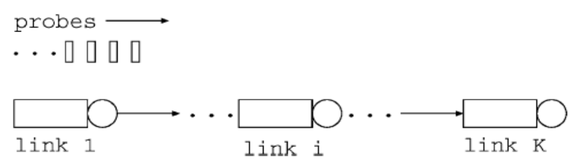

Fig. 1: Periodic probes are sent along a path with K links to identify the existence of dominant congested link.

Definition 1: Link k is a **strongly dominant congested link** in time interval $[t_1, t_2)$ if and only if for a virtual probe sent at any time $t \in [t_1, t_2)$, the following two conditions are satisfied:

$$P(t \in L_k \mid t \in L) = 1 \qquad (1)$$

$$P(D_t^k \geq \sum_{i \neq k} D_t^i \mid t \in F_k) = 1 \qquad (2)$$

In other words, link k is a strongly dominant congested link if and only if it is responsible for all the losses, and if a virtual probe experiences the maximum queuing delay at link k, this delay is no less than the aggregate queuing delays over all the other links. It is easy to see from this definition that a strongly dominant congested link is unique.

The above definition considers both loss and delay, reflecting our sense that a dominant congested link is one that causes most losses and leads to significant queuing delays. Note that the condition on queuing delay is defined over the virtual probes that experience the maximum queuing delay at link k instead of overall virtual probes. This definition accounts for the dynamic nature of the network since even a congested link may sometimes have very low queue occupancy.

Definition 2: Link k is a **weakly dominant congested link** with parameters θ and ϕ, where $0 \leq \theta < 0.5$ and $0 \leq \phi < 1$, in time interval $[t_1,t_2)$ if and only if for a virtual probe sent at any time $t\in[t_1,t_2)$, the following two conditions are satisfied:

$$P(t\in L_k \mid t\in L)\geq 1-\theta \quad (3)$$

$$P(D_t^k \geq \sum_{i\neq k} D_t^i \mid t\in F_k)\geq 1-\phi. \quad (4)$$

In other words, link k is a weakly dominant congested link if and only if a virtual probe is lost at link k with a probability no less than $1-\theta$, and if a virtual probe experiences the maximum queuing delay at link k, this queuing delay is no less than the aggregate queuing delays over all the other links with a probability no less than $1-\phi$. Since $0\leq\theta<0.5$, that is, more than half of the losses occur at a weakly dominant congested link, a weakly dominant congested link is unique. Note that the lower the values of θ and ϕ, the more stringent are the requirements on being a weakly dominant congested link. In particular, the definition of a weakly dominant congested link is the same as that of a strongly dominant congested link when $\theta=\phi=0$. A link identified as a weakly dominant congested link with θ and ϕ is also a weakly dominant congested link with θ' and ϕ', where $\theta'\geq\theta$ and $\phi'\geq\phi$. In particular, a strongly dominant congested link is a weakly dominant congested link with any $\theta\geq 0$ and $\phi\geq 0$.

Dominant Congested Link Versus Bottleneck Link

A bottleneck link is typically defined to be a link with high loss rate, long queuing delay, high utilization, low available bandwidth, or low link capacity.
Several other differences between bottleneck link and dominant congested link are the following.

- Whether or not a link is a dominant congested link is relative. A link with a low loss rate is a dominant congested link as long as it satisfies the corresponding delay and loss requirements, despite the low loss rate.
- By definition, dominant congested link is unique if it exists, while there may exist multiple bottleneck links along a path.
- Neither strongly nor weakly dominant congested link can describe links that do not have losses. Therefore, a link with the lowest capacity, available bandwidth, or highest utilization is not a dominant congested link if no loss occurs at that link.

V. IDENTIFICATION OF DOMINANT CONGESTED LINK

In this section, we first describe two hypothesis tests to identify whether a dominant congested link exists along a path. We then describe how to obtain an upper bound on the maximum queuing delay of a dominant congested link after detecting its presence.

A. Hypothesis Tests

Our hypothesis tests utilize the queuing delays of the virtual probes with loss marks, i.e., virtual probes in L. We next use an example to illustrate why these queuing delays are helpful for dominant congested link identification. Suppose the null hypothesis is that there exists a strongly dominant congested link k. Then, if this hypothesis holds, the queuing delay of any virtual probe in L must satisfy the following two properties. First, by Condition (1), it must be no less than Q_K, the maximum queuing delay at link k. Second, it must satisfy Condition (2) since all probes in F must satisfy this condition and L is a subset of F. If one of the two conditions does not hold, we can reject the null hypothesis.

We next describe the identification methodology in detail. Let W be a random variable representing the discretized end–end queuing delay of virtual probes in L. The discretization is as follows. Let D_0 denote the end-end propagation delay along the path. Let $D_{\max}$ denote the largest end-end delay of all virtual probes sent in the time interval $[t_1,t_2)$ (including those with and without loss marks). The maximum queuing delay is therefore $D_{\max}-D_0$. We divide the range of queuing delay, $[0,D_{\max}-D_0]$ into M equal length bins with bin width $b=(D_{\max}-D_0)/M$. Then, W takes value in {1,2,…,M}, where i corresponds to an actual delay value between (i-1)b and ib. Let $F_W(w)$ represent the cumulative distribution function (CDF) of W. That is, $F_W(w) = P(D_t \leq w \mid t\in L)$ for any virtual probe sent at time $t\in[t_1,t_2)$.

B. Upper Bound of the Maximum Queuing Delay at a Dominant Congested Link

Suppose link k is a strongly dominant congested link. We estimate an upper bound of its maximum queuing delay Q_K as follows. From $F_W(w)$, we find the smallest value D such that $F_W(D)>0$. Since all losses occur at link k, by the definition of $F_W(w)$, $D\geq Q_k$. Therefore, D is an upper bound of Q_K (note that D is a discretized delay value; the corresponding actual delay value is (D-1)b, where b is the bin width). For a weakly dominant congested link k with parameters θ and ϕ, we can obtain an upper bound on its maximum queuing delay Q_K in a similar manner. More specifically, from $F_W(w)$, we find the smallest value D such that $F_W(D)>\theta$, then D can be used as an upper bound of Q_K since $D\geq Q_k$ by Theorem 2 (again the actual delay bound is (D-1)b, where b is the bin width). For link k with a very small value of θ, we can apply the following heuristic to obtain a tighter bound on Q_K.

VI. IDENTIFICATION OF DOMINANT CONGESTED LINK

In this section, we validate the model-based identification method using both *ns* simulations and Internet measurements.

Validation Using ns Simulations

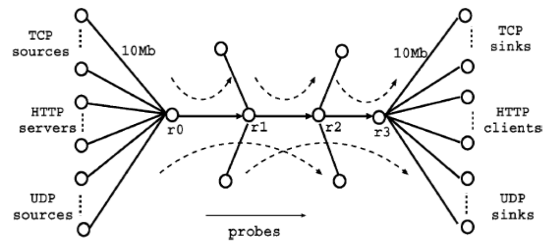

Fig. 2: Topology used in *ns*

We use a topology containing four routers, r_0, r_1, r_2 and r_3, in *ns* simulation, as shown in Fig. 2. Link (r_i, r_i+1) denotes the link from router r_i to r_{i+1}, where $0 \le i \le 2$. The bandwidth and the buffer size of link (r_i, r_i+1) are varied to create different scenarios. All the other links (from a source or a sink to its corresponding router) have bandwidth of 10 Mb/s and buffer size sufficiently large so that no loss occurs. The propagation delay of link (r_i, r_i+1) is 5 ms.

The propagation delay from a source or a sink to its corresponding router is uniformly distributed in [10,20] ms. We create three types of traffic conditions. The first type only has TCP-based traffic (in particular, FTP and HTTP traffic) from router r_i to r_j. The number of FTP flows ranges from 1 to 10, and the HTTP traffic is generated using the empirical data provided by *ns*. The second type only has UDP on-off traffic on link (r_i, r_i+1). The third type has both TCP-based and UDP ON–OFF traffic. The utilization of link (r_i, r_i+1) varies from 28% to 95% in different scenarios.

We only present results under the third type of conditions; results under the other two types are similar (indeed, our scheme relies on virtual queuing distribution and is not sensitive to whether the congestion is caused by TCP or UDP traffic). In each experiment, we send UDP probes periodically along the path from r_0 to r_3 at an interval of 20 ms. Each probe is 10 bytes. Therefore, the traffic generated by the probing process is 4 kb/s, much smaller than the link bandwidths used in the simulation.

VII. CONCLUSION

In this paper, we provided a formal yet intuitive definition of dominant congested link and proposed two simple hypothesis tests for identifying whether a dominant congested link exists along a path. We then developed a novel model-based approach for dominant congested link identification from one-way end–end measurements. Our validation in *ns* simulation and Internet experiments shows that the model-based approach requires only minutes of probing for accurate identification. As future work, we will investigate how to pinpoint a dominant congested link after identifying such a link exists.

REFERENCES

[1] A. Adams, T. Bu, R. Caceres, N. Duffield, T. Friedman, J. Horowitz, F. L. Presti, S. Moon, V. Paxson, and D. Towsley, “The use of end-to-end multicast measurements for characterizing internal network behavior,” *IEEE Commun. Mag.*, vol. 38, no. 5, pp. 152–159, May 2000.

[2] M. Jain and C. Dovrolis, “Pathload: A measurement tool for end-to-end available bandwidth,” in *Proc. PAM*, Mar. 2002, pp. 14–25.

[3] K. Lai and M. Baker, “Measuring link bandwidths using a deterministic model of packet delay,” in *Proc. ACM SIGCOMM*, Aug. 2000, pp. 283–294.

[4] R. Mahajan, N. Spring, D. Wetherall, and T. Anderson, “User-level Internet path diagnosis,” in *Proc. ACM SOSP*, Oct. 2003, pp. 106–119.

[5] B. Melander, M. Bjorkman, and P. Gunningberg, “A new end-to-end probing and analysis method for estimating bandwidth bottlenecks,” in *Proc. IEEE GLOBECOM*, Nov. 2000, vol. 1, pp. 415–420.

[6] V. N. Padmanabhan, L. Qiu, and H. J. Wang, “Server-based inference of Internet link lossiness,” in *Proc. IEEE INFOCOM*, Mar.–Apr. 2003, vol. 1, pp. 145–155.

[7] V. Paxson, “End-to-end internet packet dynamics,” *IEEE/ACM Trans. Netw.*, vol. 7, no. 3, pp. 277–292, Jun. 1999.

[8] A. Shriram and J. Kaur, “Empirical evaluation of techniques for measuring available bandwidth,” in *Proc. IEEE INFOCOM*,May 2007, pp. 2162–2170.

[9] J. Strauss, D. Katabi, and F. Kaashoek, “A measurement study of available bandwidth estimation tools,” in *Proc. ACM SIGCOMM IMC*, Oct. 2003, pp. 39–44.

[10] Wei Wei, Bing Wang *and* Don Towsley, Jim Kurose, “Model-Based Identification of Dominant Congested Links,” *IEEE/ACM Trans. Netw.*, vol 19, no.2, April 2011.

AUTHORS

First Author – K. Sivasankari, PG Student, IT Dept, Periyar Maniammai University, Vallam, Thanjavur
Email id - sivasankari74@gmail.com
Second Author – S. Jenila, PG Student, IT Dept, Periyar Maniammai University, Vallam, Thanjavur, Email id - s.jenila@gmail.com
Third Author – M. Nizar Ahamed, Assistant Professor, IT Dept, Periyar Maniammai University, Vallam, Thanjavur,
Email id - bismiindian_na@yahoo.co.in