Journal of Web Engineering, Vol. 6, No.2 (2007) 121-130

ALTERNATIVE PATH SELECTION IN RESILIENT WEB INFRASTRUCTURE USING PERFORMANCE DEPENDENCIES

VLADIMIR I. ZADOROZHNY

University of Pittsburgh, Pittsburgh, PA
vladimir@sis.pitt.edu

LOUIQA RASCHID

University of Maryland, College Park, MD
louiqa@umiacs.umd.edu

Received March 12, 2007
Revised April 30, 2007

We propose an approach to efficiently identify and substitute alternate paths in resilient Web infrastructure using overlay networks for reliable information access. Our approach is based on scalable topology-independent analysis of network behavior to identify dependencies among paths in the overlay network. Such dependencies can be characterized as non-random associations between client/server pairs and will be measured using correlation and mutual information metrics. We demonstrate that these metrics reflect physical topology characteristics, e,g., the overlap of BGP paths.

Key words: Overlay network, alternative path, performance dependency, latency profile
Communicated by: D. Lowe

1 Introduction

We address the challenge of maintaining a resilient Web infrastructure that utilizes overlay networks to seamlessly detect and recover from path outages and periods of degraded performance. Consider two motivating examples.

- Resilient Overlay Network (RON) [1] is an application-layer overlay on top of the existing Internet routing substrate. The overlay nodes monitor the liveness and quality of the Internet paths among themselves, and they use this information to decide whether to route packets directly over the Internet or by way of other RON nodes, optimizing application-specific routing metrics.

- Telecontinuity Service for telecommunication disaster protection [17] provides for rapid restoration of telephone service using robust and intelligent routing of calls during a disaster. Telecontinuity Service creates a dual network of dedicated nodes: Points of Presence (POPs). Note that the Telecontinuity service works under assumption that part of the routing infrastructure is available and alternative inter-POPs paths are utilized to perform efficient traffic re-direction during a disaster.

We provide an approach for the resilient design of Web environment, based on the concept of alternate path substitution at the level of the overlay network. Our approach is based on monitoring of paths in overlay networks, in order to identify dependencies among paths. Monitoring the behavior of

wide area networks includes many challenges. While there has been significant work on the network community to develop models of wide area networks, these techniques do not address the challenge of scalable monitoring to identify dependencies among paths.

One approach to discover paths dependencies is to develop a model of path topology at corresponding lower network layers; this can be expensive. Detecting and monitoring such dependencies can impose considerable overhead on the lower network layers [6]. In this paper we propose an approach for scalable and efficient dependency handling based on *topology independent* analysis of the network behavior. We designed a distributed catalog *AReNA* that discovers paths dependencies from network latency information. We empirically show using real network latency data that we can utilize topology independent metrics e.g., correlation and mutual information, to identify path substitutability in a scalable manner.

2 Related Work

There has been extensive research to develop metrics and models for wide area monitoring. A common objective is to predict access latencies (end-to-end delay). This research includes Internet distance and points of congestion [2, 3, 8, 11, 14]. There has been research on route aggregation based on IP prefixes exchanged via the Border Gateway Protocol (BGP) as well as research to exploit BGP information for intelligent routing and to monitor and predict performance [4, 5]. The Network Weather Service (NWS) [16, 18] is a tool that provides dynamic resource performance forecasts for wide area networks and for the computational grid. More recently, Ganglia [12] has developed promising techniques for scalable performance monitoring for clusters and the grid. Global Network Positioning (GNP) [7] presents an approach to the round-trip transmission and propagation delay prediction problem. It is based on modelling the Internet as a geometric space (e.g. a 3-dimensional Euclidean space) and characterize the position of any host in the Internet by a point in this space. An example of monitoring at the overlay level is the MCoop project [13]; it uses BGP routes expressed as paths via Autonomous Systems (ASes) to predict latencies. An Autonomous System (AS) is a collection of IP networks and routers under the control of one or more entities that present a common routing policy. Autonomous Systems can be grouped into three categories, depending on their connections and operation. A multihomed AS is an AS that maintains connections to more than one Internet Service Provider (ISP). A stub AS refers to an AS that is only connected to a single ISP. A transit AS is an AS that provides connections through itself to the networks connected to it.

While such prediction models are both accurate and valuable, their primary objective was understanding the behavior of wide area networks. It was not to develop techniques for scalable monitoring of large numbers of logical paths over the Web and to identify alternate logical paths of an overlay network for reliable information access. Scalability also motivates the need for a complementary methodology based on passive performance gathering that does not rely as heavily on complete (and perhaps expensive) knowledge of the underlying network topology and network behavior.

3 Path Substitutability and Dependency

Consider the architecture of Figure 1 that represents a hierarchy of WAN overlays. The highest level consists of POPs, which are network nodes located at a multitude of geographically dispersed locations. POPs are responsible for the network monitoring and traffic forwarding over the lower

network layers. A key functionality of the resilient infrastructure is efficient inter-POP link testing and alternative path selection in the POP network. Apparently the quality of inter-POP links depends on the link topology at corresponding lower network layers. Physical topology characteristics, e,g., the overlap of BGP routes, will impact the choice of alternative paths among POPs.

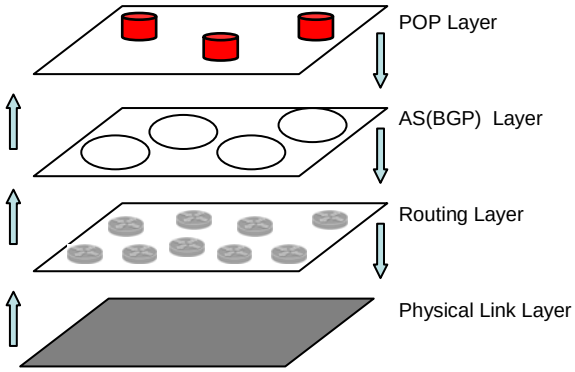

Figure 1: POP overlay and lower network layers

Next we introduce path dependencies. Consider a POP graph where nodes are POPs and edges are POP links (Figure 2). Assume that link S1.S2 is overloaded or "broken". Since this link is a part of *Path1 (S1.S2.S3)*, the infrastructure should choose an alternate path to substitute for Path1. *Path2 (S1.S6.S3)* may be an alternate path. However, we may discover that there is a path dependency between Path1 and Path2. For example, both paths may share the same ASes at the BGP layer. In this case the problems with Path1 most probably will cause a problem with *Path2*. Maintaining information about path dependencies is crucial for efficient implementation of the resilient infrastructure.

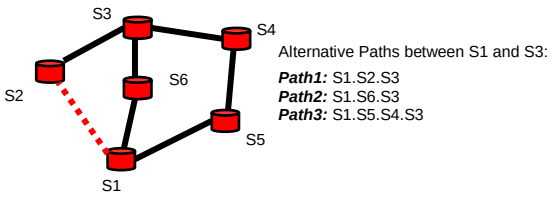

Figure 2: Alternative path selection in POP graph

In general, non-random associations, such as path dependencies, can be accommodated to locate a wide class of relationships within and between parts of the network that can be used to improve network performance aspects. Factors evaluated may include geographic diversity (regional and local), seasonal fluctuations (summer, winter, etc.), daily and hourly effects (day, night, weekend,

workday, holiday, etc), traffic type, and network configuration variations. These data can be used to determine routing configurations, optimum number, density and location of network routers; and optimum bandwidth by location.

One approach to discover dependencies is to develop a model of path topology at corresponding lower network layers. This, however, can be expensive [6]. Alternatively, dependencies can be observed from the network behavior. For example, if it is determined that the networks in the Boston area tend to be overloaded during the mid-morning hours of Tuesdays and Wednesdays and it is further determined that all east coast networks are significantly correlated with this factor, then the default routing should be reconfigured to avoid east coast area networks during that time.

Detecting and monitoring dependencies can impose considerable overhead on the lower network layers [6]. We propose an approach for scalable and efficient dependency handling based on topology independent analysis of the network behavior. We designed a distributed catalog *AReNA* [19, 20] that discovers paths dependencies from network latency information. In [10, 21] we proposed *Latency Profiles* (LPs) as a conceptual model to characterize the behavior of sources over a WAN. LPs are time-dependent latency distributions that capture the changing latencies clients experience when accessing a server. *Individual Latency Profiles* (iLPs) represent a particular client server pair; latencies are measured by client applications or middleware and gathered passively on a continuous basis. Our empirical analysis of LPs confirmed the significance of network topology and recurrent behavior over time. For example, we observed a repetitive latency behavior over a single week, with different days having different latency distributions. Thus, latency profiles can be utilized to predict latencies that clients should expect in response to requests, using historical data and recurrent behavior patterns. AReNA uses measures such as mutual information and correlation to find similarity relationships among iLPs. Such similarities typically indicate performance dependencies between corresponding paths.

Major components of *AReNA* include Data Gathering, Data Analysis, and Latency Prediction modules. AReNA environment includes three types of nodes: clients, content servers, and performance monitors (PMs). Clients continuously download data from content servers and passively construct individual iLPs. PMs aggregate non-randomly associated iLPs from multiple clients. *AReNA* utilizes *Relevance Networks* (RN) to scalably identify non-random associations in a large collection of iLPs. Relevance networks are clusters of non-randomly associated iLPs, i.e., with the similarity measure between a pairs of iLPs is above a certain threshold. *AReNA* constructs a relevance network using either mutual information or correlation as a similarity measure, and observes the changes to the number of identifiable clusters (networks), associations (edges), etc., as the threshold changes.

The AReNA Visualizer allows users to observe the evolution of the distributed environment that is being monitored via the animated Relevance Networks. Figure 3 gives an example of correlation RNs generated by AReNA for two threshold values. Each node is a clent/server pair and an edge between two nodes represents a non-random association. The thickness of an edge reflects the strength of the association. As we increase the threshold weaker associations disappear and the number of edges decreases.

Thus, Relevance Networks provides a bird's eye view of aggregate performance patterns. The visualization technique of the relevance network can be used to tune the threshold and identify stable patterns. Intuitively, a pattern is stable if for a range of thresholds the number of associations or the number of clusters does not change significantly. By observing the changes of the RN, as the threshold

is changed, AReNA can determine how strongly a cluster is associated, compared to the entire graph or to other clusters. This enables efficient, adaptable and scalable data analysis and summarization.

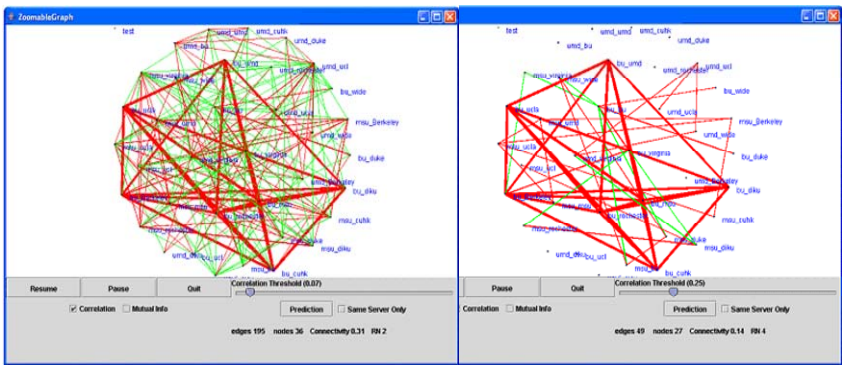

Figure 3: Relevance Networks generated by *AReNA* with increase of the Correlation Threshold.

After constructing clusters from a set of *iLPs*, *AReNA* improves the prediction quality of an iLP using observations of other, non-randomly associated *iLPs*. In related research [20], we demonstrated that high MI and high Correlation corresponds to iLP pairs with low relative error of prediction. In this paper, we will apply techniques from *AReNA* to identify alternate paths in POP graphs.

4 Latency Profiles and Similarity Metrics

Given two POP nodes p and q, an object of size b, and a temporal domain T, an *individual latency profile* is a function $iLPp,q : T \times N \rightarrow R$. $iLPp,q(t, b)$ represents the end-to-end delay for a request sent by node p to a node q at time t. Due to the stochastic nature of the network, $iLPp,q(t, b)$ is clearly a random variable. More generally, latency profiles can be time varying functions that show some regularity, such as a repetitive latency pattern. A repetitive latency behavior over a single week will include different days having different latency distributions, or similar latencies may be observed at the same time of day.

To capture the path dependencies, we define a similarity function $\Sigma: PP \times PP \times T \rightarrow SM$, where PP is the set of all POP pairs, T is a set of finite time regions (possibly intervals), and SM is a domain of similarity measure values (typically a real number between 0 and 1). Σ is a function that measures, given two latency profiles, their similarity over $\tau \in T$. We define two measures of similarity, namely an information theoretic measure, mutual information, and a statistical measure, correlation [20,21]. Mutual information between two *iLPs* X and Y is defined as

$$MI(X,Y) = \sum_{i,j} Pij \lg \frac{Pij}{PiPj}$$

where Pi,j, Pi, Pj are joint and individual probabilities of the latencies X and Y, respectively. A higher mutual information between two iLPs means that those iLPs are non-randomly associated. Conversely,

a mutual information of zero means that the join distribution of iLPs holds no more information than their individual distributions. Correlation between two iLPs *X* and *Y* is as follows:

$$Corr(X,Y) = \frac{1}{n-1}\sum_{i,j}(\frac{xi-\overline{X}}{Sx})(\frac{yi-\overline{Y}}{Sy}),$$

where $\overline{X}$, $\overline{Y}$ are expected values of random variables *X* and *Y* , and *Sx, Sy* are standard deviations of *X* and *Y*. The correlation coefficient as defined above measures the degree of the linear association between two variables. A higher correlation between two iLPs can also indicate that those iLPs are non-randomly associated. In general, there is no straightforward relationship between correlation and MI [9]. While correlation captures linear dependence, mutual information is a general dependence measure.

Example 1. Consider two individual latency profiles X and Y and their joint probability distribution XY as follows:

$$X = \begin{bmatrix} 1 & 2 \\ 0.5 & 0.5 \end{bmatrix}, \quad Y = \begin{bmatrix} 2 & 3 \\ 0.75 & 0.25 \end{bmatrix}, \quad XY = \begin{bmatrix} (1,2) & (1,3) & (2,2) & (2,3) \\ 0.5 & 0 & 0.25 & 0.25 \end{bmatrix}$$

Then MI (X, Y) = 0.31, and Corr (X, Y) = 0.57.

5 Experiments

We designed an experiment to validate that the metrics correlation and mutual information for pairs of iLPs are a good reflection of path substitutability of the POP network. For the experiment, we gathered end-to-end network latencies for multiple POP paths. We then used the BGP topology to determine several metrics. *Overlap(P1,P2)* reflects the degree of overlap at the BGP level for POP paths *P1* and *P2*. We also identify metrics *SCount* and *FCount*. A (larger) value of *SCount* reflects more situations of link failure where path *P2* can be successfully substituted for path *P1*, while a (larger) value of FCount reflects where path *P2* cannot be successfully substituted for *P1*. Below we provide more accurate definitions of above metrics. The experiments will demonstrate that the observed correlation and MI discovered by *AReNA* can be efficiently utilized for path substitutability.

We collected the experimental data over the CNRI Handle testbed [15] and the PlanetLab testbed [9]. We considered group of clients and servers ASes, with partially overlapping BGP routes. Clients periodically downloaded the corresponding digital content using HTTP requests. For this experiment, we deployed multiple clients within different subnetworks of an AS; typically these were university ASes and they were located in the Americas, Europe, and Australasia. We gathered experimental data over several months in 2003-2004. In this section we reports on the experimental results for 100 client/server pairs corresponding to alternative paths.

For each client/server pair we maintained a *log* file with the following: BGP path between the client and the server; the request time-stamp; time for the first set of bytes to arrive (TTF); the total download time (DL). For each two client/server pairs we evaluated three similarity function: correlation, mutual information (as defined in the previous section) and overlap which we explain next. Consider two clients C1 and C2 downloading data from two servers *S1* and *S2* correspondingly.

Assume that *P1* and *P2* are the BGP paths associated with *C1-S1* and *C2-S2* pairs: The overlap between the paths *P1* and *P2* is defined as

$$overlap(P1, P2) = |P1 \cap P2| / |P1|,$$

where $|P1 \cap P2|$ is the number of common nodes (ASes) between *P1* and *P2*, $|P1|$ is the number of nodes in *P1*. Consider an example with $P1=\{AS1, AS2, AS3, AS4\}$ and $P2=\{AS1, AS2, AS5, AS6, AS7\}$. Then

$$P1 \cap P2 = \{AS1,AS2\}, \ |P1 \cap P2| = 2, \ |P1| = 4.$$

This implies

$$overlap(P1, P2) = (2/4)*100 = 50\% .$$

Note that *overlap* relation is not commutative, i.e., $overlap(P2,P1) \neq overlap(P1,P2)$.

We also define two measures of substitutability between two BGP paths *Pi* and *Pj* sharing the same sources node. First measure is number *SCount* of successful substitutions of *Pi* by *Pj* provided that nodes of *Pi* are failing. Second measure is number *FCount* of non-successful substitutions of *Pi* by *Pj* provided that nodes of *Pi* are failing. Table 1 illustrates these measures for above *P1* and *P2* paths assuming that *P1* fails.

Failed AS of P1	***Can P1 be substituted by P2?***
AS1	No *(FCount = FCount +1)*
AS2	No *(Fcount = FCount +1)*
AS3	Yes *(SCount = SCount +1)*
AS4	Yes *(SCount = SCount +1)*

Table 1: Explanation of SCount and FCount measures

As we observe from the table, *SCount* increments if we successfully substitute *P1* by *P2* (i.e., failed AS is not a part *P2*). Otherwise we increment *FCount*. Thus for above example *SCount = 2* and *FCount = 2*. Note that this substitutability relation is not symmetrical. If we assume that *P2* fails and try to substitute it with *P1* the numbers would be as follows: *SCount = 3; FCount = 2*.

In our experiment we reported the trends in *SCount* and *FCount* measures as a function of the BGP paths dependencies for all client/server pairs. Thus, for every two pairs *CS1* and *CS2* we estimated topological dependency *overlap(CS1, CS2)*, as well as observed dependencies based on correlation (*Corr*) and mutual information (*MI*) between CS1 and CS2. In order to estimate *Corr* and MI measures we generated latency profiles *iLPi* for each client server pair CSi and CS2 as described in section 4. The paths were split in groups sharing the same client. SCount and FCount measures were evaluated for each path with respect to each path in its group. Thus we avoided substitutability checks between the paths with different sources. We group pairs of paths based on their *SCount* and *FCount* values. The length of BGP paths in our experiments were ranging from 2 to 6 *ASes* with *SCount* and *FCount* values of 1, 2, 3 and 4.

Figure 4 and 5 reports on the experimental results. We plotted *SCount* and *FCount* measures for all pairwise dependencies. In order to reflect the trends better we also plotted exponential fitting curves for each of considered dependencies.

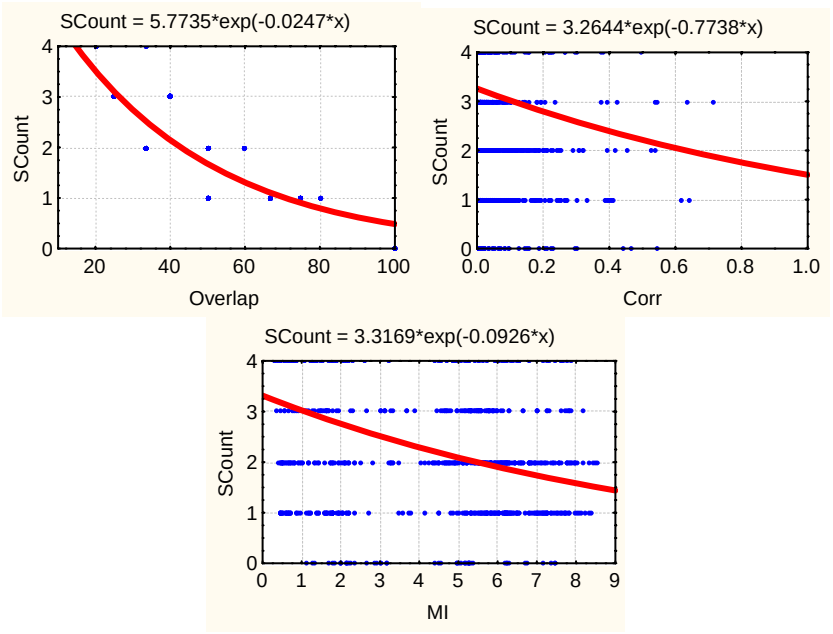

Figure 4: Impact of topological and observed dependencies on successful substitutions

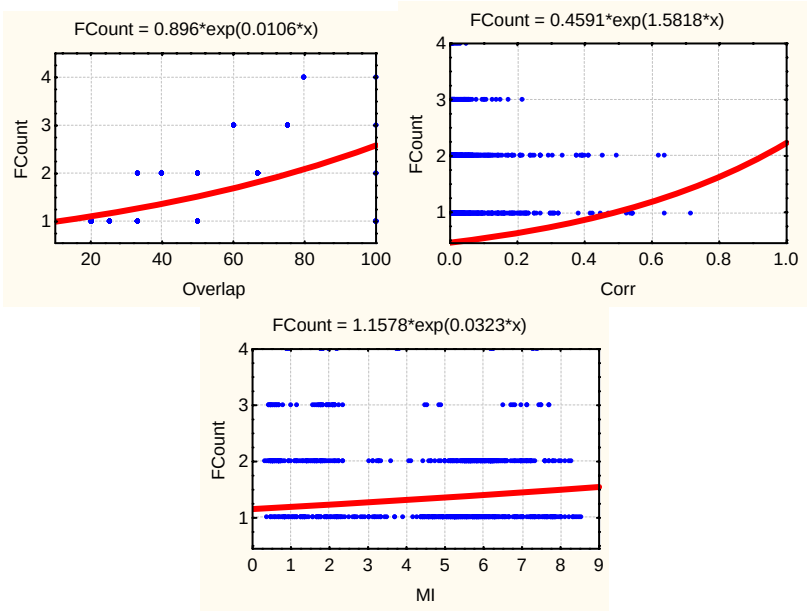

Figure 5: Impact of topological and observed dependencies on non-successful substitutions

Figure 4 reports on *SCount* measure. First, we consider behavior of the overlap metrics shown in the top left graph in Figure 4. Apparently, when there is low overlap between pairs of paths, the *SCount* is high since the chances of common AS link being broken are lower then for the higher

overlap. For those pairs of paths with higher overlap, the *SCount* decreases as expected. We also plot the values of Correlation and MI, for the corresponding pairs of paths. While the values of Correlation and MI do show some dispersion, we observe a significant trend. For higher values of *SCount* (=4), the Correlation between the pairs of paths is low. As expected, when the *SCount* is low (=1), indicating more overlap in the BGP topology of pairs of paths, the Correlation is higher (Figure 4, top right graph). We observe a similar trend for MI (Figure 4, bottom graph).

In general *FCount* graphs (Figure 5) demonstrate opposite trends as expected. For the pairs of paths with higher overlap, Correlation and MI the FCount decreases. However, the impact of correlation on FCount seems to be stronger then that of MI. Note that this is different from what we observed for SCount. This observation allows us to suggest considering both Correlation and MI impacting SCount and FCount measures in order to make a decision on path substitutability. This sounds as a promising approach that requires more research.

6 Conclusions

In this paper we proposed an approach to identify and substitute alternative paths in resilient Web infrastructure using overlay networks. Our approach is based on scalable and efficient dependency handling using *topology independent* analysis of the network behavior. We designed a distributed catalog *AReNA* that discovers paths dependencies from network latency information. We empirically showed that utilizing topology independent metrics e.g., correlation and mutual information, we can identify alternative path substitutability in a scalable manner. We believe the approach reported in this paper can yield a methodology to apply the same principles of measurement and assessment to the Internet in general.

Acknowledgement

We would like to thank Avigdor Gal for his valuable contribution to *AReNA* research.

References

1. D. Andersen, H. Balakrishnan, M. Frans Kaashoek, R. Morris. Resilient Overlay Networks. *Proc. of 18th ACM SOSP, 2001*
2. P. Francis, S. Jamin, V. Paxson, L. Zhang, D. Gryniewicz, Y. Jin. An Architecture for a Global Internet Host Distance Estimation Service. *Proc. of IEEE InfoComm*, 1999
3. S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavin, L. Zhang. On the Placement of Internet Instrumentation. *Proc. of IEEE InfoComm*, 2000
4. B. Krishnamurthy, J. Wang. On Network-aware Clustering of Web Clients, *Proc. of SIGCOMM'*02
5. Z. Mao, C. Cranor , F. Douglis, M. Rabinovich, O. Spatscheck, J. Wang. A Precise and Efficient Evaluation of the Proximity between Web Clients and their Local DNS Servers, *USENIX Annual Technical Conference*, 2002

6. A. Nakao, L. Peterson, A. Bavier. A routing underlay for overlay networks. *Proc. of ACM SIGCOM*, 2003
7. E. Ng, Z. Hui. Towards Global Network Positioning. *Proc. of ACM SIGCOMM Internet Measurement Workshop*, 2001
8. V. Padmanabhan, L. Subramanian. An Investigation of Geographic Mapping Techniques for Internet Hosts, *Proc. of SIGCOMM*, 2001
9. PlanetLab home page. http://www.planet-lab.org
10. L. Raschid, H.-F. Wen , A. Gal , V. Zadorozhny. Latency Profiles: Performance Monitoring for Wide Area Applications. *Proc. of IEEE Workshop on Internet Applications* , 2003
11. D. Rubenstein, J. Kurose, D. Towsley. Detecting Shared Congestion of Flows via End-to-end Measurement, *Proc. of ACM SIGMETRICS*, 2000
12. F. Sacerdoti, M. Katz, M. Massie, D. Culler. Wide Area Cluster Monitoring with Ganglia. *Proc. of the IEEE Cluster 2003 Conference*, 2003
13. S. Srinivasan and E. Zegura. M-coop:A Scalable Infrastructure for Network Measurement. *Proc. of IEEE Workshop on Internet Applications,* 2003
14. M. Stemm, S. Seshan, R. Katz. A Network Measurement Architecture for Adaptive Applications. *Proc. of IEEE InfoComm,* 2000
15. S. Sun, L. Lannom, Handle System Overview. IRDM/IRTF Draft, http://www.idrm.org/idrm_drafts.htm, 2001
16. M. Swany, R. Wolski. Multivariate Resource Performance Forecasting in the Network Weather Service. *Proc. of SC*, 2002
17. TeleContinuity home page. http://www.telecontinuity.com/
18. R. Wolski. Dynamically Forecasting Network Performance to Support Dynamic Scheduling Using Network Weather Service, *Proc. 6th High-Performance Distributed Computing Conerencef*, 1997
19. V. Zadorozhny, A. Gal, L. Raschid, Q.Ye. AReNA: Adaptive Distributed Catalog Infrastructure Based On Relevance Networks. *Proc. of VLDB*, 2005
20. V. Zadorozhny, L. Raschid, A. Gal, Q. Ye, H. Murthy. Using Non-random Associations for Predicting Latency in WANs. *Proc. of WISE*, 2005
21. V. Zadorozhny, A. Gal, L. Raschid, Q. Ye,. Wide Area Performance Monitoring Using Aggregate Latency Profiles. *Proc. of ICWE,* 2004