MD-AVB: A Multi-Manifold Based Available Bandwidth Prediction Algorithm

Pei Zhang, Changqing An, Zhanfeng Wang*, and Fengyuan Ma

Abstract: The performance of Internet applications is heavily affected by the end-to-end available bandwidth. Thus, it is very important to examine how to accurately predict the available Internet bandwidth. A number of available bandwidth prediction algorithms have been proposed to date, but none of the existing solutions are able to achieve a high level of accuracy. In this paper, a Multi-manifold based Available Bandwidth prediction algorithm (MD-AVB) is proposed, based on the observation that the available bandwidth space on the Internet is multi-manifold and asymmetrical. In the proposed algorithm, the available bandwidth space is divided into multiple lower-dimensional domains iteratively, and each domain is embedded separately to predict the available bandwidth. Experiments on HP S^3 datasets demonstrate that the proposed algorithm is more accurate than existing approaches.

Key words: available bandwidth space; performance prediction; multi-manifold; asymmetry

1 Introduction

The performance of many Internet applications is heavily affected by the available bandwidth and this is particularly the case for bandwidth-sensitive applications, such as multimedia, cloud computing, and P2P file sharing. If the available Internet bandwidth information between nodes can be obtained or predicted, this can help these bandwidth-sensitive applications to avoid network congestion and improve their performance.

One way to predict the available bandwidth is based on the network path or Autonomous System (AS) routing, which is adopted by BRoute[1] and SABI[2]. However, some of these methods, such as BRoute, cannot meet the real-time requirement of network applications, and the prediction accuracy is also insufficient. Therefore, there is an urgent need to design an available bandwidth prediction algorithm with a higher prediction accuracy and smaller measurement overheads.

In order to satisfy this requirement, another way to measure the available bandwidth has been proposed based on space embedding, as in the PathGuru method[3]. Researchers have designed a number of network coordinate systems based on the embedding theory. Although these approaches can achieve a superior prediction accuracy compared with methods based on the network path, most of them ignore the non-metric nature of the available bandwidth space. This non-metric feature means that the datasets of the available bandwidth may be high-dimensional, and thus they cannot fully meet the requirements of the embedding theory, which is limited to datasets featuring low dimensionality.

In this paper, by analyzing the characteristics of available Internet bandwidth space, it is found that it is multi-manifold and asymmetric in nature. In response, a multi-manifold based available bandwidth prediction algorithm is proposed, denoted as MD-AVB. The proposed algorithm divides the available
bandwidth space into multiple lower-dimensional domains iteratively to ensure that any two arbitrarily chosen network nodes must have a subspace domain that contains them. An embedding algorithm based on singular value decomposition is then employed to predict the available bandwidth for each domain.

The remainder of this paper is organized as follows. Section 2 gives a short survey of related work about available bandwidth models and techniques for bandwidth prediction. In Section 3, the related concepts and characteristics of the available Internet bandwidth are introduced. Section 4 provides details of the MD-AVB algorithm, and Section 5 evaluates its performance. Finally, Section 6 summarizes the work.

2 Related Work

Despite it being an important Internet metric, the prediction of available bandwidth has received much less attention compared to that of path latency. Key et al.\cite{4} studied how to find the best path based on the quantitative relationship of the bandwidth bottleneck between pairs of hosts, but they did not consider the prediction accuracy for the bandwidth bottleneck itself. Other researchers have attempted to model the available Internet bandwidth space using space embedding theory. Ng and Zhang\cite{5} were the first to propose using the Global Network Positioning system (GNP) to predict network distance, with a method that embedded Internet delays into a Euclidean space. Based on this idea, methods based on the network path or AS routing have been proposed to predict the available bandwidth. Dabek et al.\cite{6} proposed the Vivaldi algorithm, which used a distributed architecture to realize space embedding. Hu and Steenkiste\cite{1} found that most of the Internet’s bottleneck links were located on the edge of the network, and proposed an available bandwidth prediction mechanism called BRoute, which was based on the source and destination node tree, and routing information. BRoute required a lot of data processing and incurred a large performance cost. Shriram\cite{2} proposed the Scalable Available Bandwidth Inference (SABI) method based on routing path, which considered the topology of the network, but neglected to account for the fact that the bottleneck of the network can appear at any hop in a route.

The space embedding-based methods focus on mining the nature of the dataset, regardless of the actual network topology. Ramasubramanian et al.\cite{7}, in their Sequoia system, used the tree space not only to achieve the latency forecast, but also to forecast the available bandwidth. Song et al.\cite{8} presented a Distributed Sequoia (D-Sequoia) method, which proved that constructing a distributed tree can achieve the same precision compared with performing global embedding. However, such a mechanism cannot handle problems of asymmetric bandwidth. Xing et al.\cite{3} presented PathGuru, a method of predicting available bandwidth based on ultra-metric space, which gave minimal error for the nodes in the access network, but only achieved a low prediction accuracy for nodes that did not satisfy the requirement of an ultra-metric space. Xing et al.\cite{9} further optimized PathGuru; they used measurement results from some specific landmarks to infer the available bandwidth and were able to deal with available bandwidth asymmetry.

As outlined above, an increasing number of researchers have tried to use space embedding models to predict the end-to-end available bandwidth. However, existing solutions lack precision due to the global embedding methods. The deficiency lies in the fact that available bandwidth is a non-metric space, such that the datasets do not meet the embedded theory’s requirement for low-dimensionality. Furthermore, the available Internet bandwidth space is essentially multi-manifold, and a successful prediction algorithm must be designed based on this characteristic. In order to further improve the prediction accuracy, our focus in this paper is on these characteristics of Internet bandwidth space.

3 Analysis of the Available Internet Bandwidth Space

Available bandwidth is one of the main metrics used to evaluate the performance of the Internet. The end-to-end available bandwidth represents the minimum capacity available among all links on a given path. The available bandwidth of an Internet path is thus determined by the minimum bandwidth along its link; this is known as the path bottleneck. Therefore, the study of available bandwidth can be transformed into the study of available bandwidth bottlenecks. Available Internet bandwidth prediction needs to estimate the available bandwidth between nodes using just the nature and model of available bandwidth, without direct measurements.

In this section, based on theoretical and experimental
analysis, we introduce the HP $S^3$[10] dataset for use in predicting available Internet bandwidth, and show that Internet bandwidth space has the features of being multi-manifold and asymmetric. These characteristics are the basis for the design of the available bandwidth prediction algorithm presented in the following Section 4.

### 3.1 Dataset

HP $S^3$ project[10] continuously measures path performance metrics, such as the latency, available bandwidth, and loss rate between Internet nodes. The available bandwidth measurement results from this dataset are used as the survey object in this paper. These data are highly meaningful, since we know that the available bandwidth measurements are costly at Internet scale. The source data includes the source address, destination address, delay, and bandwidth. By extracting the bandwidth data between the source and destination addresses, we obtained a $400 \times 400$ available bandwidth matrix. In addition, a $400 \times 400$ random available bandwidth matrix dataset was generated for comparison, in which each number in the matrix is a random number between $(0, 10^4)$, and the number on the diagonal of the matrix is $0$.

### 3.2 Multi-manifold feature of the available bandwidth space

The available Internet bandwidth space is essentially multi-manifold. To verify this, we analyzed the different positions of the datasets, since the intrinsic dimensions of different datasets would be the same if the data were not to conform to the feature of multi-manifold space. Differences in dimensions, on the other hand, would equate to a multi-manifold characteristic[11]. Therefore, whether the available Internet bandwidth space is multi-manifold or not can be determined by examining the dimensionality of the different parts of the datasets.

We used Principal Components Analysis (PCA)[12] and Isomap[13], two classical linear and nonlinear methods respectively, to analyze the dimension characteristics of the dataset[14]. The PCA algorithm reveals the dimensions of the dataset by determining whether the cumulative contribution rate of the first $n$ main components can reach a certain threshold after the projection, whereas Isomap performs the geodetic distance calculation and doubly centers the distance matrix before the principal component decomposition.

Three sub-datasets were extracted from different parts of HP $S^3$ and denoted as HP1, HP2, and HP3. The size of the sub-datasets was randomly selected as $100, 80,$ and $100$, respectively. The comparison of HP1 and HP3 is as sub-datasets of different parts of the same scale, while HP2 was used as a further check to show that the size of the sub-dataset does not affect the results. Details of the sub-datasets are given in Table 1.

Figure 1 displays the cumulative distribution of the dimensions of the three sub-datasets (HP1, HP2, and HP3) of HP $S^3$ through PCA analysis, where the horizontal axis represents the dimension and the vertical axis represents the cumulative distribution of the first $n$ dimensions.

The experimental results show that these three sub-datasets do not match in the dimension characteristic, and that they therefore belong to different manifolds.

The Isomap analysis for the dimensions of these three sub-datasets (HP1, HP2, and HP3) is illustrated in Fig. 2. It can be seen from Fig. 2 that when the residuals are different, their dimensions are also different; thus, the available Internet bandwidth space is multi-manifold.

The reasons for the formation of a multi-manifold characteristic in an actual network were also considered. Due to the underlying heterogeneity of the network topology, the Internet feature space cannot be represented simply by a linear or non-

| Table 1 Three analytical sub datasets. |
|-------------------------------|------------------|
| Dataset | Size   |
|--------|--------|
| HP1    | $100 \times 100$ |
| HP2    | $80 \times 80$   |
| HP3    | $100 \times 100$ |

![Fig. 1 PCA analysis of the dimensions of datasets.](image-url)
linear model, it requires a more complex structure. In addition, in large-scale datasets the error of assuming a single feature model becomes more obvious. Therefore, our analysis shows the superiority of a multi-manifold model of the available Internet bandwidth space.

3.3 Asymmetry of the available bandwidth space

The available Internet bandwidth space has the feature of asymmetry. The model is shown in Fig. 3, where the available bandwidths between $H_i \rightarrow H_j$ and $H_j \rightarrow H_i$ are different. If the available bandwidth between Internet nodes $H_i$ and $H_j$ is represented, it requires a double-tuple including the outgoing and incoming available bandwidth for each node; in what follows, $(b^\text{out}_i, b^\text{in}_j)$ represents the outgoing and incoming available bandwidth bottlenecks of node $i$, respectively.

PathGuru\cite{3} is an example of predicting the available bandwidth $\text{avb}(H_i, H_j)$ between network nodes $H_i$ and $H_j$, which is represented as $\text{avb}_{\text{PG}}(H_i, H_j)$ in the following equations:

$$\text{avb}_{\text{PG}}(H_i, H_j) = \min\{b^\text{out}_{H_i}, b^\text{in}_{H_j}\}$$ (1)

$$\text{avb}_{\text{PG}}(H_j, H_i) = \min\{b^\text{out}_{H_j}, b^\text{in}_{H_i}\}$$ (2)

To illustrate this feature, the asymmetric coefficient of the available bandwidth asymmetry $(H_i, H_j)$ is defined as

$$\text{asymmetry}(H_i, H_j) = \frac{\max a\{\text{avb}(H_i, H_j), \text{avb}(H_j, H_i)\}}{\min\{\text{avb}(H_i, H_j), \text{avb}(H_j, H_i)\}}$$ (3)

From this it can be seen that the bigger the asymmetry $(H_i, H_j)$, the greater the asymmetry of the available bandwidth between nodes $H_i$ and $H_j$.

We take the HP $S^3$ dataset as an example to demonstrate the asymmetry feature of the available bandwidth between Internet nodes (see Fig. 4).

In Fig. 4, the horizontal axis represents the asymmetric coefficient between network nodes, and the vertical axis represents the probability of the cumulative distribution. In this figure, we find that the probability that the asymmetric coefficient of the available bandwidth between the nodes is greater than 2 is about 55%, which indicates that the available Internet bandwidth space has an asymmetric characteristic.

When performing spatial embedding and available bandwidth prediction, we should consider the available bandwidth between nodes asymmetrically, and therefore as expressed by separate representations of incoming bandwidth and outgoing bandwidth.

4 MD-AVB

The above analysis makes clear the multi-manifold characteristic of the available Internet bandwidth space,
so the traditional prediction mechanism based on
global embedding cannot obtain a sufficiently precise
prediction. Therefore, in this paper, we propose a
multi-manifold based available bandwidth prediction
algorithm denoted as MD-A VB, which is made up of the
following two main parts: (1) a multi-domain partition
algorithm, and (2) an available Internet bandwidth
prediction algorithm for each domain.

4.1 Multi-domain partition algorithm

The multi-domain partition algorithm can divide the
available Internet bandwidth space into multiple lower-
dimensional subspace domains iteratively. In this way,
the differences in the performance of each subspace
of the network can be considered, and the accuracy of
prediction can be improved.

With the available Internet bandwidth space denoted
as $D$, the multi-domain partition algorithm consists of
the following steps:

Step 1: Select the first node in the space $D$, then
select $L$ nodes containing this node randomly as the
alternative subspace domain $D_i$.

Step 2: Apply the PCA algorithm to $D_i$ to determine
whether it is valid and get the vector space $U_k$ of the
selected domain $D_i$.

Step 3: Determine whether the node outside of the
alternative subspace $D_i$ can join $D_i$. If it meets the
requirement, this node can join the subspace domain;
if not, set this as the first node and repeat Step 1.

Step 4: Finally, select all the subspace domains
iteratively with all the nodes in different subspace
domains. Ensure that any two network nodes have a
subspace domain that contains them both.

This algorithm is illustrated in Fig. 5. When selecting
the valid subspace domains, we firstly select one node
as the first node, and select $L$ nodes containing this
node randomly as the landmarks. In order to ensure
the low-dimensionality characteristic of each subspace
domain $D_i$, we use the PCA algorithm to decompose its
covariance matrix $C$:

$$C = \frac{1}{T} \sum_{i=1}^{L} (d_i - \mu)(d_i - \mu)^T$$

$$CU = U\Lambda$$

where $\mu = \frac{1}{T} \sum_{i=1}^{L} d_i$ denotes the mean vector of the
candidate subspace domain, $U$ denotes the eigenvector
matrix, and $\Lambda$ represents the matrix whose diagonal
elements are eigenvalue $\lambda_i$, which is denoted as the
eigenvalue matrix. We then choose the corresponding
eigenvectors of the first $k$ largest eigenvalues to form
a $k$-dimensional space $U_k$, which denotes the vector
space of the selected domain $D_i$. This subspace
domain will be accepted when it satisfies the following
conditions:

$$S(k) > \delta, \ 0 < \delta < 1$$

$$S_i = \frac{\lambda_i}{\sum_{i=1}^{P} \lambda_i} < \eta, \ \forall i > k, \lambda_i \in \Lambda$$

This means that the cumulative distribution of the
first $k$ dimensions should be larger than the predefined
threshold $\delta$, and the individual contribution of the
remaining vectors should be smaller than the given
threshold $\eta$.

When a new node needs to be added to a candidate
subspace domain, we should examine whether it can
be embedded in $D_i$. We examine whether the relevant
parameters exceed the given threshold $\delta$ and $\eta$ after
this node is embedded in a subspace domain. If
the space domain can continue to maintain the low-
dimensionality characteristic, this node is permitted to
join this domain, otherwise it will find another subspace
domain. Finally, all of the nodes will be embedded into
different subspace domains and any two network nodes
are in at least one common domain.

The steps of the multi-domain partition algorithm are
summarized in Algorithm 1.
The coordinates of the landmark nodes are calculated with the following formulae:

\[ X_i = (L^i_m Y)(Y^T Y)^{-1} \]

\[ Y_i = (L^i_{out} X)(X^T X)^{-1} \]

Once all nodes are embedded in domains, it is estimated that the available bandwidth between any two nodes \( H_i \) and \( H_j \) in the network can be calculated directly using their coordinates \((X_i, Y_i)\) and \((X_j, Y_j)\) in the current subspace domain, as follows:

\[ \text{avb}(H_i, H_j) = X_i \cdot Y_j \]  \hspace{1cm} (11)

\[ \text{avb}(H_j, H_i) = X_j \cdot Y_i \]  \hspace{1cm} (12)

The steps of the available bandwidth prediction algorithm for each domain are summarized in Algorithm 2.

### 5 Performance Evaluation

MD-AVB aims at improving prediction accuracy by dividing the available space into many low-dimensional subspace domains, then embedding their nodes to each domain. In this section, the 400×400 available bandwidth matrix extracted from the dataset HP \( S^3 \) is used to experimentally evaluate three aspects of the prediction performance of MD-AVB: (1) the dimensionality of the sub-datasets, (2) the related parameters selection of each of the subspace domains, and (3) the embedding errors of the algorithm.

#### 5.1 Dimensionality of the sub-datasets

Firstly, the multi-domain partition algorithm is performed to divide the dataset HP \( S^3 \) into multiple low-dimensional subspace domains iteratively. In order to verify that the divided subspace domains do fit this characteristic, we analyze the dimensionality of these domains and accumulate the results.

Figure 6 shows the cumulative result of the dimensional analysis of the divided sub-datasets, representing different subspace domains. The horizontal axis represents the dimension and the vertical axis represents the contribution rate after the normalization of the dimension.

### Algorithm 2 Available bandwidth prediction algorithm

**Input:** \( B \) //the matrix of available bandwidth

**Output:** \( (X, Y) \) //the coordinate of the Internet node

1. for each subspace domain, initialize \( B \);
2. \( B_L = \text{RandomSelectLandmark}(B) \);
3. \( (X_L, Y_L) = \text{SVD}(B_L) \);
4. for each node in \((B - B_L)\)
5. \( X_i = (L^i_m Y)(Y^T Y)^{-1} \);
6. \( Y_i = (L^i_{out} X)(X^T X)^{-1} \);
7. \( (X, Y) \leftarrow (X_i, Y_i) \);
8. End;
9. Return;
In Fig. 6, the contribution rate of the first dimension is higher than 80% and after the fourth dimension any further contribution is negligible, indicating that the divided available Internet bandwidth subspace domains have obvious low-dimensional characteristics. This experimental result provides a guarantee for the subsequent algorithm, which applies embedding theory to predict the available bandwidth between nodes for each domain.

### 5.2 Related parameters selection for each subspace domain

After the available Internet bandwidth space is divided, the embedding algorithm based on singular value decomposition for each domain is used to predict the available bandwidth. There are two key parameters, being the embedded dimensions dim and the number of landmark nodes \( L \). For this experiment, the embedded dimensions ranging from 2 to 15 were selected for analysis. The number of landmark nodes is generally more than twice the number of embedded dimensions\(^{[15]}\), so we chose 2, 3, and 4 times the number of the embedded dimensions as the number of landmark nodes. The experiment for each combination of \((\text{dim}, L)\) was repeated 30 times with the landmark nodes randomly selected.

Table 2 shows the experimental results with the embedded dimension ranging from 2 to 12. It expresses the cumulative contribution rate of the algorithm under the condition of Unidirectional Relative Error (URE) being set to 0.9, and thereby compares the prediction accuracy of different combinations of \((\text{dim}, L)\). It can be inferred from the results that the greater the number of embedded dimensions and landmark nodes, the higher the accuracy. Furthermore, for the same dimension, the accuracy of the algorithm is greater when the number of landmark nodes is larger and the algorithm is able to obtain a better convergence. However, it was also found that once the number of landmark nodes increases to a certain level, the increase in accuracy is not very obvious, and overhead is increased.

Assuming that the the embedded dimension is fixed to 4 and the number of landmark nodes is fixed to 12, the cumulative distribution of this combination is shown in Fig. 7, where the horizontal axis represents URE and the vertical axis represents the cumulative distribution. Each line in the figure represents the result

---

**Table 2 Cumulative distribution of each combination of \((\text{dim}, L)\) (URE=0.9).**

| Dimension (dim) | Landmark nodes \((L)\) | Cumulative distribution |
|-----------------|-----------------------|------------------------|
| 2               | 4                     | 0.702                  |
| 2               | 6                     | 0.758                  |
| 2               | 8                     | 0.763                  |
| 3               | 6                     | 0.741                  |
| 3               | 9                     | 0.767                  |
| 3               | 12                    | 0.785                  |
| 4               | 8                     | 0.749                  |
| 4               | 12                    | 0.792                  |
| 4               | 16                    | 0.803                  |
| 5               | 10                    | 0.760                  |
| 5               | 15                    | 0.798                  |
| 5               | 20                    | 0.812                  |
| 6               | 12                    | 0.774                  |
| 6               | 18                    | 0.801                  |
| 6               | 24                    | 0.823                  |
| 9               | 18                    | 0.769                  |
| 9               | 27                    | 0.826                  |
| 9               | 36                    | 0.837                  |
| 12              | 24                    | 0.790                  |
| 12              | 36                    | 0.818                  |
| 12              | 48                    | 0.833                  |

---

**Fig. 7 Cumulative distribution of the combination of \((4, 12)\).**
of the operation of each subspace domain for the combination (4, 12).

5.3 Embedding errors of MD-AVB

The comparison of embedded precision between the whole dataset HP $S^3$ and each of the sub-datasets is illustrated in Fig. 8. We chose to experiment with the parameter (dim, L) set as in the previous section at (4, 12). As shown in this figure, the experiment was run on several selected divided sub-datasets. The red curve represents the embedded URE cumulative contribution rate of the entire HP $S^3$ dataset, while the blue curves represent the sub-datasets. It can be seen from Fig. 8 that the precision is significantly improved after dividing the available Internet bandwidth space.

The prediction precision on the HP $S^3$ dataset when selecting a random available bandwidth matrix dataset was then compared with MD-AVB. In Fig. 9, the results are shown with the PathGuru[3] algorithm also added for comparison. Results show that the accuracy of the MD-AVB algorithm is superior to the alternative techniques, with the UREs of 80% of the prediction results being lower than 0.5.

The results provide evidence that the available Internet bandwidth space can be divided into subspace domains with low-dimensional characteristics to achieve superior prediction accuracy. Furthermore, the experiment shows that choosing appropriate parameters can further improve the success of the algorithm.

6 Conclusion

Existing solutions to predict the available Internet bandwidth using network coordinate systems are based on space embedding theory, which cannot achieve a high level of accuracy because the available bandwidth is a non-metric space.

In this paper, a new mechanism was presented to predict the end-to-end available Internet bandwidth with a small number of measurements. Based on the results of Internet measurement data, we verified the multi-manifold and asymmetry features of the available Internet bandwidth space. Drawing on these features, we proposed a multi-manifold based available bandwidth prediction algorithm named MD-AVB. In MD-AVB, the available bandwidth space is divided into multiple lower-dimensional domains iteratively. An embedding algorithm based on singular value decomposition for each domain is then used to predict the available bandwidth. The experimental evaluation of the algorithm shows that it can achieve superior accuracy. In future work, we will study how to deploy MD-AVB to an actual network.

Acknowledgment

This work was supported by the National Key Research and Development Program of China (No. 2016YFB0801302).

References

[1] N. N. Hu and P. Steenkiste, Exploiting internet route sharing for large scale available bandwidth estimation, in Proc. 5th ACM SIGCOMM Conf. Internet Measurement, Berkeley, CA, USA, 2005. p. 16.

[2] A. Shiriram, Efficient techniques for end-to-end bandwidth estimation: Performance evaluations and scalable deployment, https://cdr.lib.unc.edu/concern/dissertations/vn40xs86h, 2009.

[3] C. Y. Xing, M. Chen, and L. Yang, Predicting available

Fig. 8 Comparison of embedded precision between the whole dataset and sub datasets, where the red curve represents the embedded URE cumulative contribution rate of the entire HP $S^3$ dataset, and the blue curves represent the sub-datasets.

Fig. 9 Comparison for prediction error.
bandwidth of internet path with ultra metric space-based approaches, in Proc. GLOBECOM 2009—2009 IEEE Global Telecommunications Conf., Honolulu, HI, USA, 2009.

[4] P. Key, L. Massoulié, and D. C. Tomozei, Non-metric coordinates for predicting network proximity, in Proc. IEEE INFOCOM 2008—The 27th Conf. on Computer Communications, Phoenix, AZ, USA, 2008, pp. 1840–1848.

[5] T. S. E. Ng and H. Zhang, Predicting internet network distance with coordinates-based approaches, in Proc. 21st Ann. Joint Conf. of the IEEE Computer and Communications Societies, New York, NY, USA, 2002.

[6] F. Dabek, R. Cox, F. Kaashoek, and R. Morris, Vivaldi: A decentralized network coordinate system, ACM SIGCOMM Comp. Commun. Rev., vol. 34, no. 4, pp. 15–26, 2004.

[7] V. Ramasubramanian, D. Malkhi, F. Kuhn, M. Balakrishnan, A. Gupta, and A. Akella, On the treeeness of internet latency and bandwidth, in Proc. 11th Int. Joint Conf. on Measurement and Modeling of Computer Systems, Seattle, WA, USA, 2009.

[8] S. Song, P. J. Keleher, and A. Sussman, Searching for bandwidth-constrained clusters, in Proc. 31st Int. Conf. on Distributed Computing Systems, Minneapolis, MN, USA, 2011.

[9] C. Y. Xing, G. M. Zhang, B. Xu, C. Hu, and M. Chen, A hierarchical internet path available bandwidth inference mechanism, J. Int. Technol., vol. 18, no. 1, pp. 135–146, 2017.

[10] P. Yalagandula, P. Sharma, S. Banerjee, S. Basu, and S. Lee, S3: A scalable sensing service for monitoring large networked systems, in Proc. of 2006 SIGCOMM Workshop on Internet network Management, Pisa, Italy, 2006.

[11] Y. Wang, Y. Jiang, Y. Wu, and Z. H. Zhou, Multi-manifold clustering, in Proc. 11th Pacific Rim Int. Conf. on Trends in Artificial Intelligence, Daegu, Korea, 2010.

[12] I. T. Jolliffe, Principal Component Analysis. Springer, 2011.

[13] J. B. Tenenbaum, V. De Silva, and J. C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science, vol. 290, no. 5500, pp. 2319–2323, 2000.

[14] A. Ghodsi, Dimensionality reduction a short tutorial, https://www.math.uwaterloo.ca/~aghodsi/b/courses/06stat890/reading/tutorial_stat890.pdf, 2006.

[15] Z. F. Wang, M. Chen, C. Y. Xing, and G. M. Zhang, T-aware: An AS topology based available bandwidth prediction algorithm, (in Chinese), J. Beijing Univ. Posts Telecommun., vol. 34, no. 3, pp. 66–70, 2011.

Pei Zhang received the PhD degree from Beijing University of Posts and Telecommunications in 2012. After graduation, he joined Tsinghua University for postdoctoral research. Since 2014, he has joined Beijing University of Posts and Telecommunications and now he is a lecturer. His current interests are network security and network measurement and management.

Changqing An received the BS and MS degrees from Tsinghua University in 1992 and 1996, respectively. She is currently an associate professor of the Institute for Network Sciences and Cyberspace of Tsinghua University. Her research interests include network architecture, network protocols, network measurement, and network performance modeling.

Zhanfeng Wang received the PhD degree from PLA University of Science and Technology, Nanjing, China, in 2012. He is currently a postdoctor in Computer Science and Engineering, Southeast University, Nanjing, China. His research interests include network architecture, network performance modeling, network measurement, IPv6 network, and network security.

Fengyuan Ma received the BS degree from Beijing Jiaotong University. She is now a master student at Beijing University of Posts and Telecommunications. She has conducted research on network performance space modeling, and her current interest is front-end development and data mining about DNS traffic feature pattern mining.