Predicting Internet Network Distance with Coordinates-Based Approaches

10/29 2003

By wusf

Outline

- Introduction
- Triangulated Heuristic
- Global Network Position (GNP)
- IDMaps, Triangulated Heuristic, and GNP Comparison
- Experimental Results
- Summary

Introduction

- The critical problem is to devise techniques that can predict network distance accurately, scalably, and in a timely fashion
- Three techniques:
  1. IDMaps
  2. Triangulated Heuristic
  3. Global Network Position (GNP)

Introduction

- IDMaps：

  IDMaps is an infrastructural service in which special HOPS servers maintain a virtual topology map of the Internet consisting of end hosts and special hosts called Tracers

Example：

Host：A,B　　Tracer：T1,T2

T1 is the tracer that is nearest A

T2 is the tracer that is nearest B

Then compute the value

d(A,T1) +d(B,T2)+d(T1,T2)

Introduction

- We propose to use coordinates-based mechanisms in a P2P architecture to predict Internet network distance
- The main idea is to ask end hosts to maintain coordinates

  →characterize their locations in the Internet such that network distances can be predicted by evaluating a distance function over hosts' coordinates

Introduction

- Coordinates-based approaches fit well with the peer-to-peer architecture

  When an end host discovers the identities of other end hosts in a peer-to-peer application, their pre-computed coordinates can be piggybacked

→ network distances can essentially be computed instantaneously by the end host

Introduction

Triangulated Heuristic：

based on relative coordinates that are simply the distances from a host to some special network nodes

→ To reduce the computation overhead of shortest-path searches in interdomain graphs

Introduction

Global Network Positioning：

Based on absolute coordinates computed from modeling the Internet as a geometric space

1. To model the Internet as a geometric space
2. Characterize the position of any host in the Internet by a point in this space

Triangulated Heuristic

- Select N nodes in a network to be **base nodes**
- Then, a node H is assigned coordinates which are simply given by the N-tuple of distances between H and the N base nodes

$$\text{i.e. } (d_{\mathcal{HB}_1}, d_{\mathcal{HB}_2}, .., d_{\mathcal{HB}_N})$$

Triangulated Heuristic

1. Given two nodes H1 and H2
2. Assuming the triangular inequality holds

- The triangulated heuristic states that the distance between H1 and H2 is bounded below by L and above by U which

$$L = \max_{i \in \{1,2,..,N\}} (|d_{\mathcal{H}_1 \mathcal{B}_i} - d_{\mathcal{H}_2 \mathcal{B}_i}|)$$

$$U = \min_{i \in \{1,2,..,N\}} (d_{\mathcal{H}_1 \mathcal{B}_i} + d_{\mathcal{H}_2 \mathcal{B}_i})$$

Global Network Position (GNP)

- propose a two-part architecture

1.**Landmarks** ：

Compute their own coordinates in a chosen geometric space

→serve as a frame of reference and are disseminated to any host who wants to participate

2. **Ordinary host**：

Equipped with the Landmarks' coordinates

→any end host can compute its own coordinates relative to those of the Landmarks

GNP：Landmark Operations

- Suppose there are N Landmarks, L1 to LN.
- The Landmarks

1. Measure the inter-Landmark round-trip times using ICMP ping messages
2. Take the minimum of several measurements for each path to produce the bottom half of the N x N distance matrix

GNP：Landmark Operations

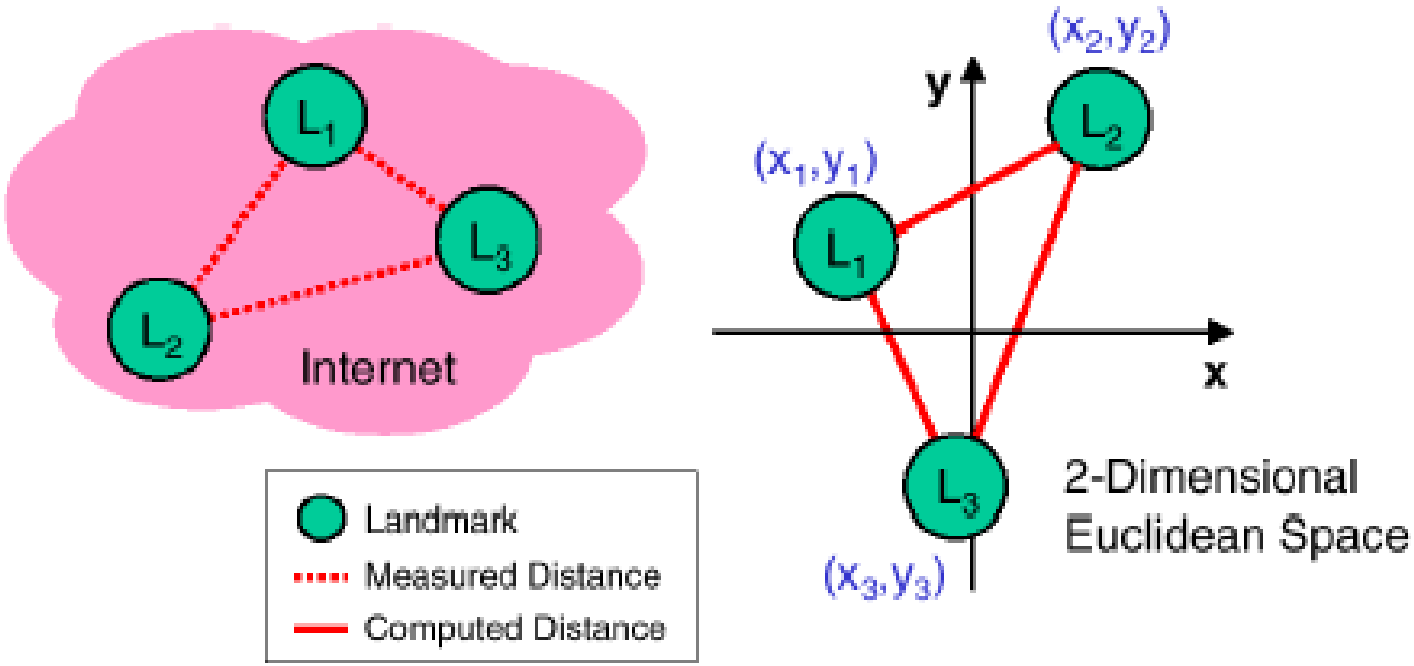


Fig. 2. Part 1: Landmark operations

GNP：Landmark Operations

$$\mathcal{E}(d_{\mathcal{H}_1\mathcal{H}_2}, \hat{d}^{\mathcal{S}}_{\mathcal{H}_1\mathcal{H}_2}) = (d_{\mathcal{H}_1\mathcal{H}_2} - \hat{d}^{\mathcal{S}}_{\mathcal{H}_1\mathcal{H}_2})^2$$

$$f_{obj1}(c^{\mathcal{S}}_{\mathcal{L}_1}, .., c^{\mathcal{S}}_{\mathcal{L}_N}) = \sum_{\mathcal{L}_i, \mathcal{L}_j \in \{\mathcal{L}_1, .., \mathcal{L}_N\} \mid i > j} \mathcal{E}(d_{\mathcal{L}_i\mathcal{L}_j}, \hat{d}^{\mathcal{S}}_{\mathcal{L}_i\mathcal{L}_j})$$

GNP：Ordinary Host Operations

- Ordinary hosts are required to actively participate

→Using the coordinates of the Landmarks in the geometric space S, each ordinary host now derives its own coordinates

- An ordinary host H measures its round-trip times to the N Landmarks using ICMP ping messages

GNP：Ordinary Host Operations

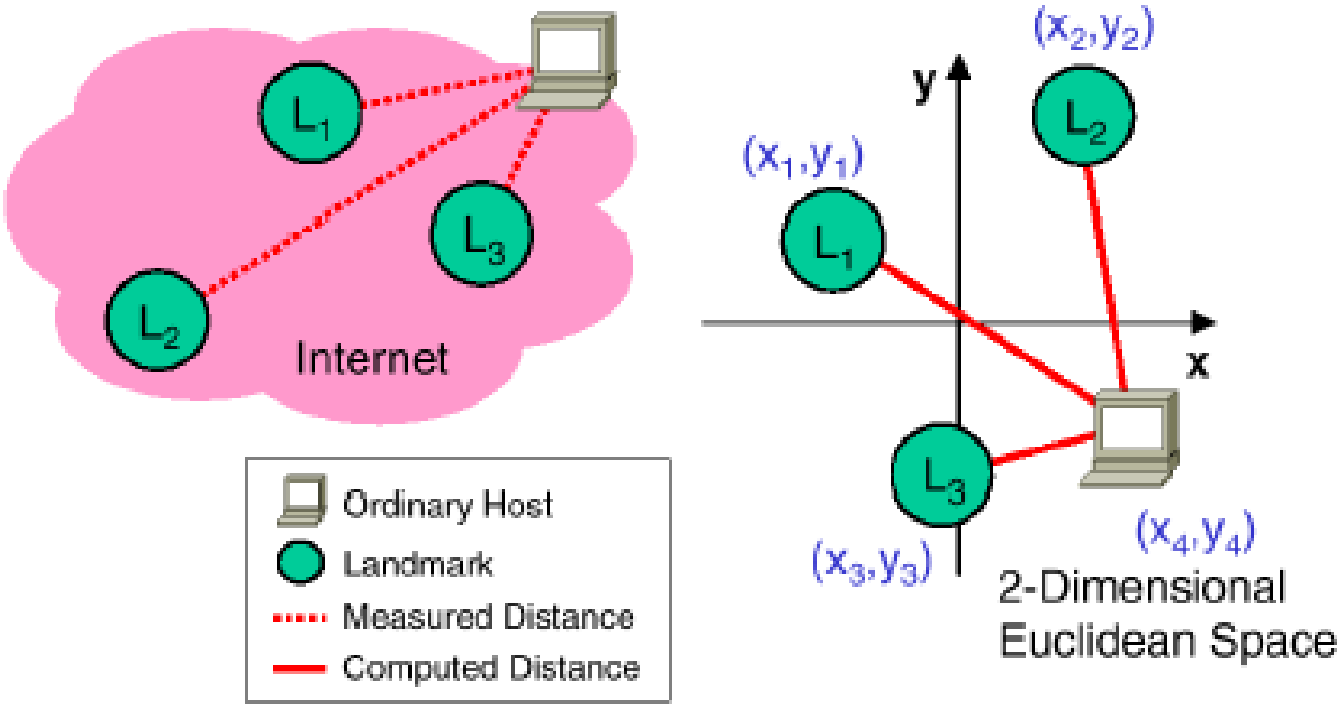


Fig. 3. Part 2: Ordinary host operations

GNP：Ordinary Host Operations

$$f_{obj2}(c_{\mathcal{H}}^{S}) = \sum_{\mathcal{L}_i \in \{\mathcal{L}_1,..,\mathcal{L}_N\}} \mathcal{E}(d_{\mathcal{H}\mathcal{L}_i}, \hat{d}_{\mathcal{H}\mathcal{L}_i}^{S})$$

IDMaps, Triangulated Heuristic, and GNP Comparison

About IDMaps：

- IDMaps is an infrastructural service in which hosts called Tracers are deployed to measure the distances between themselves
- Each Tracer is responsible for measuring the distances between itself and the set of IP addresses

IDMaps, Triangulated Heuristic, and GNP Comparison

- The coordinates-based approaches have higher scalability
- The main difference between Triangulated Heuristic an GNP

→Triangulated Heuristic :triangulated inequality

GNP：special function

Experimental Results

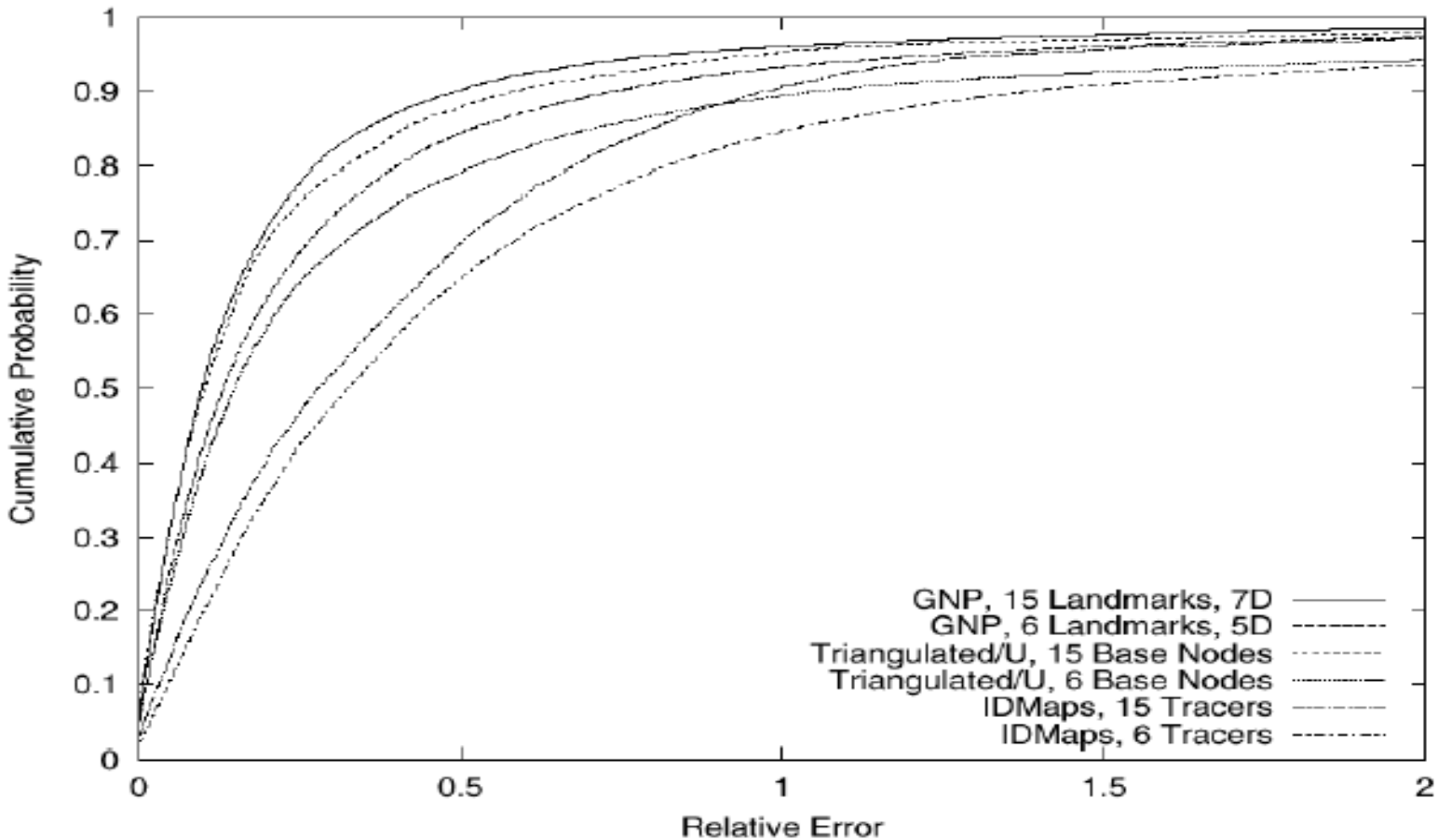
- Directional relative error：

$$\frac{predicted\ distance - measured\ distance}{\min(measured\ distance, predicted\ distance)}$$

- Rank accuracy：

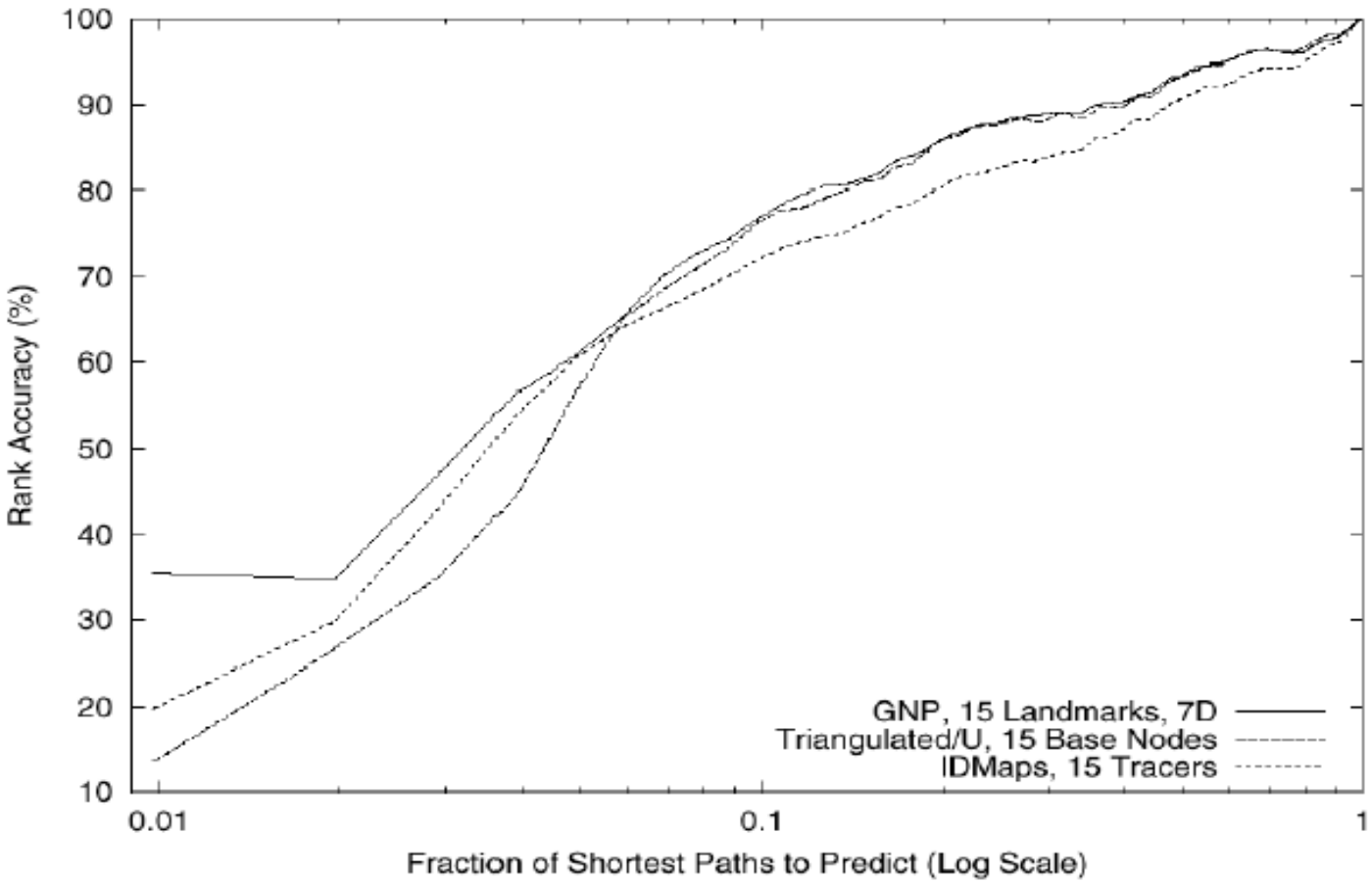
  The percentage of paths correctly selected

Experimental Results



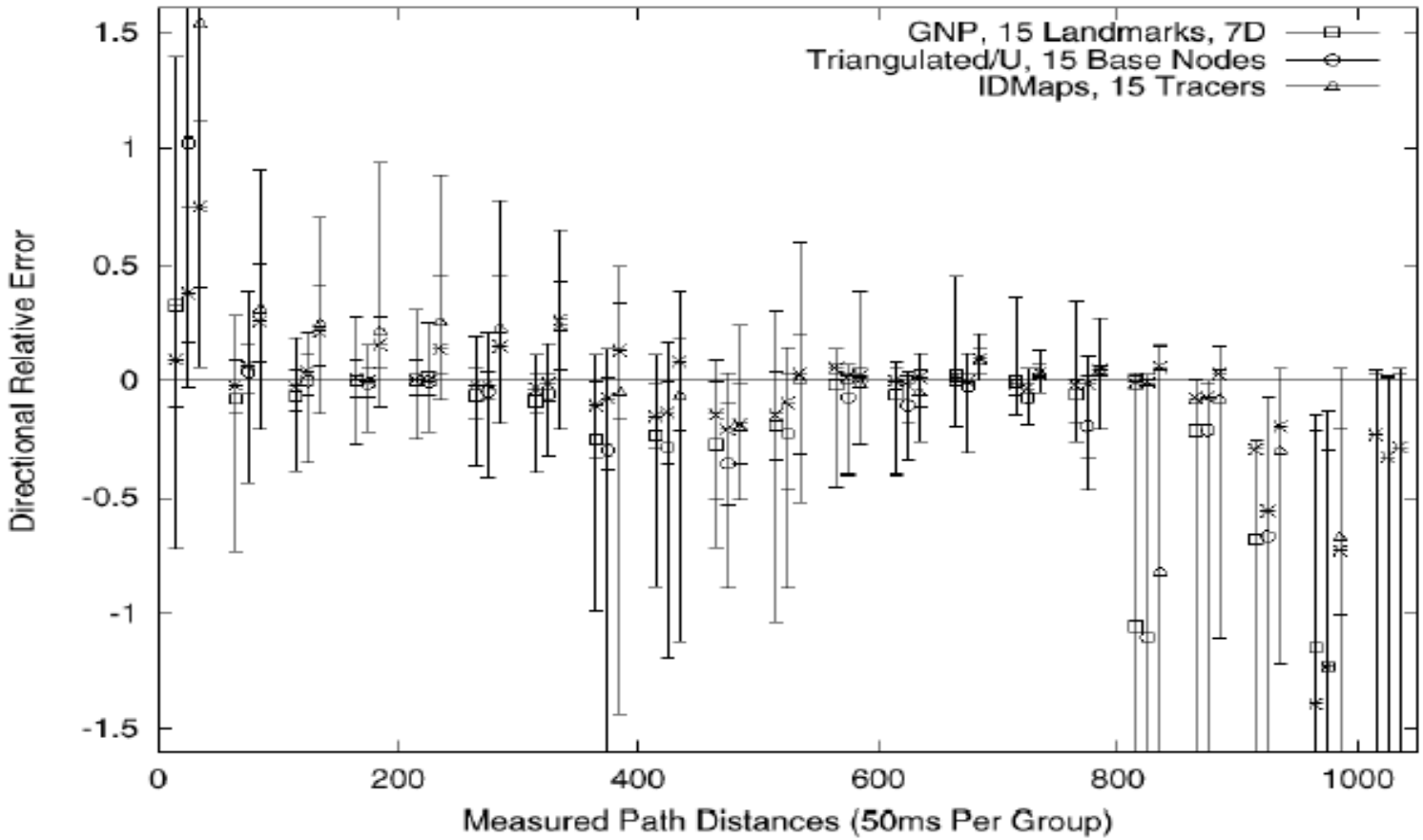
Relative error comparison (Global)

Experimental Results



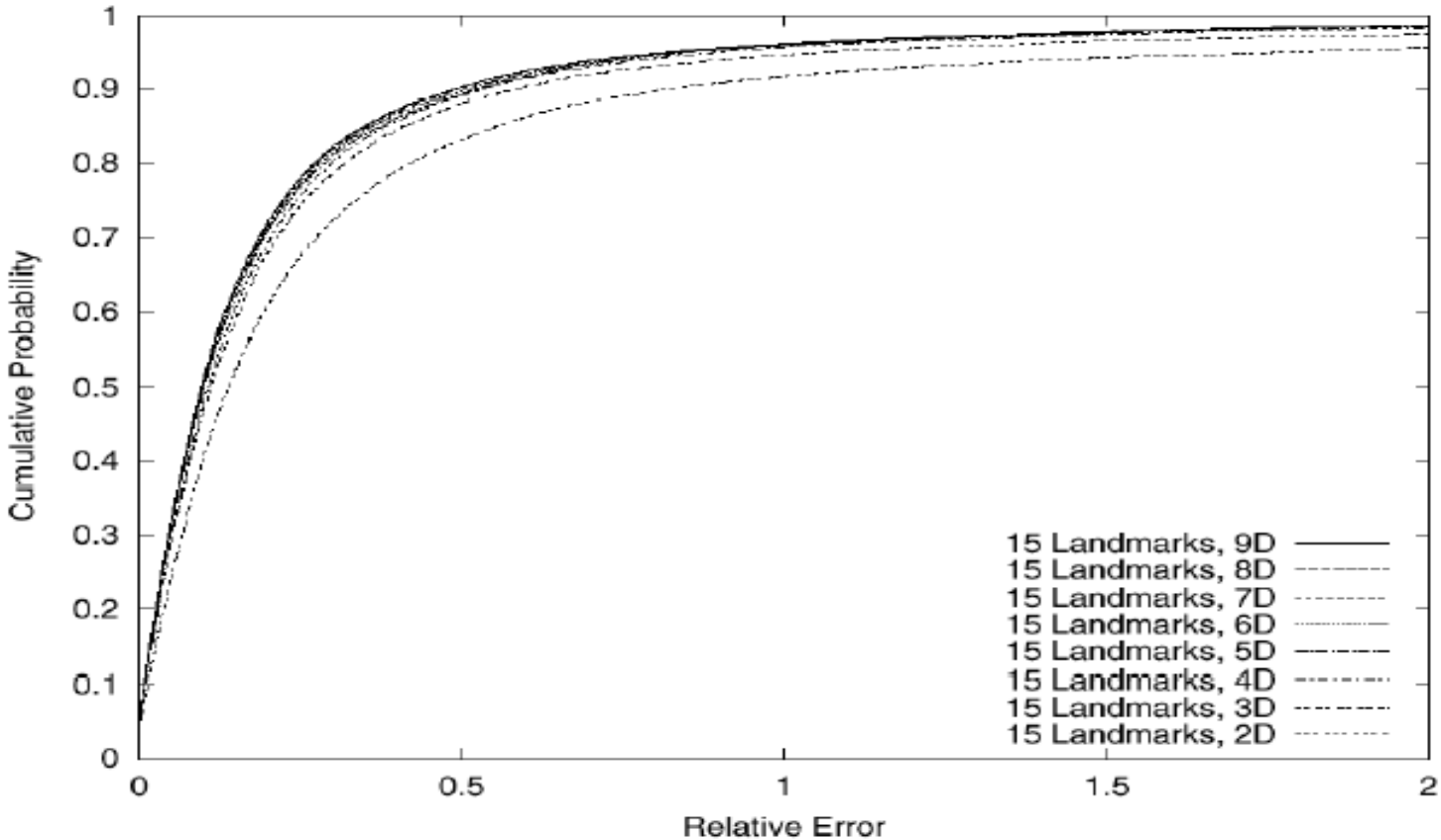
Rank accuracy comparison (Global)

Experimental Results



Directional relative error comparison (Global)

Experimental Results



. Convergence of GNP performance

Summary

- We have studied a new class of solutions to the Internet distance prediction problem ：

1.Previously proposed triangulated heuristic
2.New approach called GNP

→GNP is most accurate and robust