archives-ouvertes.fr

Le modèle Inframétrique pour Internet

Pierre Fraigniaud, Emmanuelle Lebhar, Laurent Viennot

▶ To cite this version:

Pierre Fraigniaud, Emmanuelle Lebhar, Laurent Viennot. Le modèle Inframétrique pour Internet. David Simplot-Ryl and Sebastien Tixeuil. 10ème Rencontres Francophones sur les Aspects Algorithmiques des Télécommunications (AlgoTel'08), 2008, Saint-Malo, France. pp.73-76, 2008. <inria-00374460>

HAL Id: inria-00374460
https://hal.inria.fr/inria-00374460

Submitted on 8 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le modèle Inframétrique pour Internet

Pierre Fraigniaud[1,3], Emmanuelle Lebhar[1,3] and Laurent Viennot[2,3]

[1] *CNRS and University Paris 7, France.* [2] *INRIA and University Paris 7, France.* [3] *The authors are members of the INRIA Projet-Team "GANG" between INRIA Rocquencourt and LIAFA. Additional supports from the COST Action 295 "DYNAMO", CRC "MARDI", and from the ANR projects "ALPAGE" and "ALADDIN".*

De nombreux algorithmes ont été récemment con cus pour l'Internet sous l'hypothèse que la mesure du temps d'aller retour (RTT) est une distance. De plus, nombre de ces algorithmes (construction de réseau logique, conception de schéma de routage compact, construction de couverture ayant peu d'arêtes) reposent sur l'hypothèse que la métrique d'Internet a une croissance de boule ou une dimension doublante bornée. Cet article étudie la validité de ces hypothèses et propose un modèle formel et analysable qui correspond aux observations expérimentales.
La version complète de cet article est [2] : `www.liafa.jussieu.fr/∼elebhar/exposes/inframetric.pdf`.

Keywords: Topologie d'Internet, métrique doublante, routage compact, RTT.

1 Introduction

The quest for a better understanding of the Internet structure at the router level as well as at the AS level has yield a tremendous amount of work over the last decade, initiated by the pioneering contributions of Faloutsos et al. [1] identifying power laws in the Internet. A large amount of algorithms has also recently been designed for the Internet, including overlay network construction, routing scheme design, sparse spanner construction, closest server selection, etc. The design of these algorithms assumes that the distance defined by the round-trip delay (RTT)[†] is a metric, and hence, in particular, that the triangle inequality $RTT(u,v) \leq RTT(u,w) + RTT(w,v)$ is satisfied for any triple u,v,w. Moreover, the performance analysis of many of these algorithms relies on the assumption that RTT has bounded ball growth, i.e., for any $r > 0$, the size of any ball of radius r can be bounded by a constant times the size of the ball of radius $r/2$ centered at the same node .Important contributions have relaxed this assumption to the bounded doubling dimension hypothesis, i.e., for any $r \geq 0$, any ball of radius r can be covered by a constant number of balls of radius $r/2$. Metrics of bounded doubling dimension have recently received a considerable attention because they provide a richer framework for the design and analysis of algorithms.

The bounded ball growth assumption can be well motivated intuitively and is consistent with the transit-stub model [9]. Although it can be shown that the RTT delays are poorly correlated to metrics as physical distances, the formal verification of the bounded ball growth assumption has been statistically established in average by [1, 10]. For instance, [1] shows that the RTT distance over all pairs follows a power law, i.e., the number of pairs $P(h)$ at RTT distance h satisfies $P(h) \propto h^c$ for some constant c. As a consequence, $P(2h) \propto 2^c \cdot P(h)$. Nevertheless, the averaging over all pairs may hide the ball growth misbehavior for a large number of centers, and the assumption $P(2h)/P(h) \leq O(1)$ is not strong enough to enable algorithms designed under the bounded ball growth assumption to perform efficiently in a framework in which the bound only holds in average. On the other hand, previous work tends to indicate that the basic metric assumption is questionable.

In this paper, we experimentally revisit the validity of the metric and geometrical assumptions made for the RTT latency distances. Significant contributions have been made (e.g. [3, 6]) to better understand the RTT distance, but their approach relies on extrapolating the missing data by using the triangle inequality, which is not necessarily satisfied by the RTT distance.

We propose a tractable analytical model for the RTT distances in the Internet, matching our experimental observations. We demonstrate the tractability of our model by showing how it can be used to design and

[†] The RTT between nodes u and v is the time taken to send a packet from u to v and to receive an acknowledgment back from v to u.

analyze sophisticated algorithms. Our results are therefore complementary to the results in [10]. Indeed, [10] was among the first contributions considering experimentally the ball growth of the Internet, and modeling the violation of the triangle inequality. The objective was however quite different from ours. It developed a fine tuned and compact statistical tool box for the purpose of simulation and emulation of protocols. This tool box provides an artificial synthesis of a realistic delay space. In contrast, we develop an analytical model for the purpose of design and analysis of algorithms. Consequently, we focus on worst case analysis of the experimental data, rather than on the global statistical distribution of the parameters.

2 The Inframetric Model

Inframetrics. In this paper, we call *distance* any function aiming at capturing a notion of proximity between elements of a finite set V. Clearly the RTT latency falls into this category. Recall that a nonnegative (distance) function $d : V \times V \to \mathbb{R}$ is a *metric* if it satisfies :

- $d(u,v) = 0$ if and only if $u = v$;
- $d(u,v) = d(v,u)$ (symmetry property) ;
- $d(u,v) \leq d(u,w) + d(w,v)$ (triangle inequality).

As we mentioned in the introduction, and as our experiments presented later in the paper will demonstrate, the RTT latency merely satisfies the two first properties, but significantly violates the triangle inequality. We thus introduce a relaxed version of this latter property, yielding the notion of *inframetric*.

Definition 1 *A distance $d : V^2 \to \mathbb{R}$ is a* ρ-inframetric *for $\rho \geq 1$ if it satisfies the two first axioms of metrics and the following relaxed triangle inequality : for any triple u,v,w in V, $d(u,v) \leq \rho \max\{d(u,w), d(w,v)\}$.*

Definition 2 *Let $0 < \delta \leq 1$. A triangle u,v,w is* δ-skewed *if $d(w,v) \leq \delta\, d(u,v)$. An inframetric d is* (ρ_s,δ)-skewed *for $\rho_s > 0$ if for any δ-skewed triangle u,v,w, we have $d(u,w) \leq \rho_s\, d(u,v)$.*

The notion of (ρ_s,δ)-skewness for a ρ-inframetric d is particularly interesting if ρ_s is significantly smaller than ρ. In particular, a detour via w when routing from u to v in a δ-skewed triangle u,v,w results in a stretch factor $\rho_s + \delta$ instead of $\rho + \delta$.

Ball Growth and Doubling Dimension. We generalize the notions of ball growth and doubling dimension to inframetrics. Given a distance function d on a set V, $B_u(r)$ denotes the ball of radius $r \geq 0$ centered at $u \in V$, i.e., $B_u(r) = \{v \in V \mid d(u,v) \leq r\}$. The standard definitions of ball growth and doubling dimension compare balls of radius $2r$ with balls of radius r. These notions can naturally be extended to ρ-inframetrics by comparing balls of radius ρr with balls of radius r. A ρ-inframetric d has *growth* $\gamma \geq 1$ if, for any $r \geq 0$ and $u \in V$, $|B_u(\rho r)| \leq \gamma\, |B_u(r)|$. Metrics of bounded growth are special cases of metrics of bounded doubling dimension. A ρ-inframetric is γ-doubling if, for any $r \geq 0$ and $u \in V$, $B_u(\rho r) \subseteq \cup_{i \in I} B_{v_i}(r)$, for some $v_i \in V$, $i \in I$, $|I| \leq \gamma$. We have the following lemma (proof omitted).

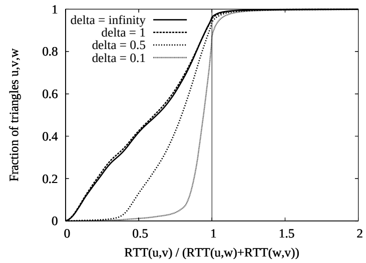
Lemma 1 *Let d be an inframetric. If d is of growth γ then it is γ'-doubling with $\gamma' \leq \gamma^4$.*

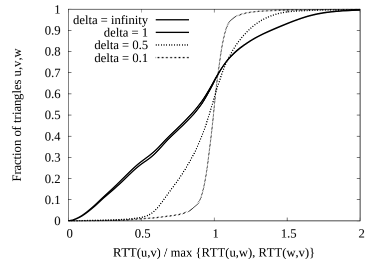
Regarding our measures on Internet, the aforementioned definitions of ball growth and doubling dimension have two drawbacks. First, we observe a limited growth of the ball sizes only for radii above a certain threshold. Second, the definitions $|B_u(\rho r)| \leq \gamma\, |B_u(r)|$ or $B_u(\rho r) \subseteq \cup_{i \in I} B_{v_i}(r)$, $|I| \leq \gamma$, still yields relatively large γ's even for reasonably large balls. These two problems are handled by the following definition.

Definition 3 *A ρ-inframetric is (α,β)-doubling with threshold τ if, for any $u \in V$, any $r \geq \tau$, and any $R \geq \rho r$, $B_u(R) \subseteq \cup_{i \in I} B_{v_i}(r)$, for some $v_i \in V$, $i \in I$, $|I| \leq \beta \alpha^{\log_\rho R/r}$.*

3 Internet Latencies as an Inframetric

Triangle inequality. We use here PlanetLab measures. In order to test the triangle inequality, we compute the ratios $RTT(u,v)/(RTT(u,w) + RTT(w,v))$ for all triangles for which the measures are available. We study separately the triangles depending on their skewness : the parameter $\delta > 0$ for which they satisfy $RTT(w,v) < \delta\, RTT(u,w)$. The ratios depending on the skewness are illustrated in the next figure. We can see that approximately 5 % of the triangles do not satisfy the triangle inequality. We also verified experimentally that for most δ-skewed triangles, the skewed triangle inequality is satisfied with ρ_s ranging from 1.2 to 1.8 for $0 \leq \delta \leq 0.7$.

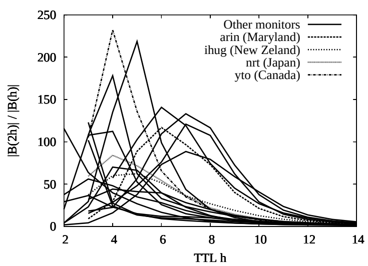


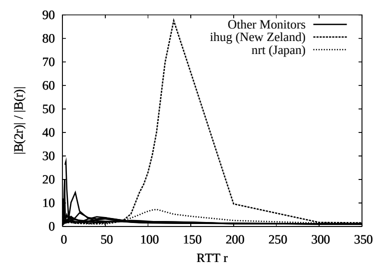


4 Ball Growth and Doubling Dimension of Internet Latencies

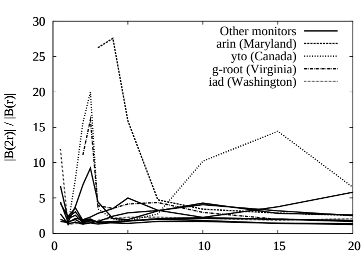
Caida Skitter measurements The Skitter project of Caida consists in few monitors regularly probing a fixed very large set of IP addresses. Each probe consists in a `traceroute` query. We extract from this data a RTT measurement from each monitor to each IP address probed as well as a hop distance measurement.

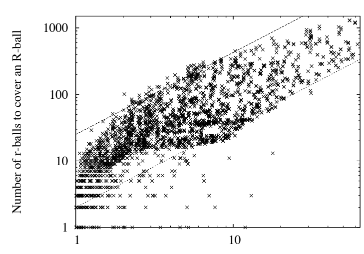
Bounded growth The left figure below plots the growth of balls defined by hop counts as a function of the TTL hop count. There is clearly no low bound on this growth. The right hand side figure plots the growth of balls defined by RTTs as a function of the RTT radius ; it is smaller than 8 except for small radii and for two monitors. The highest peak is due to `ihug` monitor which is an isolated monitor from New-Zealand.





The left hand side figure below zooms in smaller RTTs. The highest peak is due to `arin` monitor which is located in Maryland. It is probably due to high density of the network in that region. In the same region, `iad` and `g-root` present rather high ratios for small RTTs. Notably, `yto` monitor (Ottawa) has two peaks (at 2.5 and 15ms). The second one could be explained by a threshold *r* where high speed international connectivity occurs. The ball growth appears to be bounded by 15 for radii greater than 7 ms for all monitors but one and it is mostly lower than 5. To avoid artifacts, we ignore balls of less than 20 nodes.





Doubling property. Sampling nodes in a bounded growth space does not necessarily results in a bounded growth space, but always in a doubling metric. Including an isolated island of nodes (e.g. New-Zealand) can induce a high ball growth for these nodes. Considering the growths observed on Skitter data, doubling metrics seem a good candidate for modeling the geometry of latencies space. To test the doubling property on Internet latencies, we need a large set of nodes with all to all measurements ; we thus used two matrices (P2PSim [4] and Meridian [8]) obtained with the King method [5].

We test the (α, β)-doubling property as follows : repeatedly select two random radii r, R with $r < R$ and run a heuristically optimized greedy algorithm to cover a randomly chosen ball B of radius R with balls of radius r. The number $U_{r,R}(B)$ of balls of radius r necessary to cover B is an estimation of $\beta\alpha^{\log_\rho \frac{R}{r}}$. The right hand side of the last figure illustrates $U_{r,R}$ as a function of $\frac{R}{r}$ in Meridian matrix (log-log scale). Precisely, for each ball B of radius R and radius $r < R$, we plot a point (x,y) where $x = \frac{R}{r}$ and $y = U_{r,R}(B)$ is the number of balls of radius r found to cover B. Most of the points lie between $2 \cdot (2.35)^{\log_2 \frac{R}{r}}$ and $25 \cdot (2.35)^{\log_2 \frac{R}{r}}$. The highest values of $U_{r,R}(B)$ have been observed for all radii ranges from 1ms to 200ms. High $\frac{R}{r}$ ratios can only be observed with small r. Observation on short radii cannot be conclusive since King matrices are made between DNS and may miss high local density situations. Bounds were similar in P2PSim matrix.

Based on the observation of the largest complete latency matrices available, we can argue that Internet latencies satisfy the (α,β)-doubling property for a small value of α. The data sets investigated suggest $2 \leq \alpha \leq 3$ and $\beta \leq 30$. Note that points in a D dimensional grid with L_1 norm form a $(1, 2^D)$-doubling metric. We should thus compare α to 2 for a one-dimensional space and 4 for a bi-dimensional space.

5 Compact routing in doubling inframetrics

In this section, we consider an n-node ρ-inframetric space (V,d) which is (α,β)-doubling with threshold τ. We also assume that the inframetric is (ρ_s,δ)-skewed with $\delta \leq 1$ (this assumption is relaxed in the last corollary). Doubling metrics properties in computer science algorithmic problems are often analyzed trough a central decomposition tool : the *r-nets*. We say that a subset S of V is an r-net if for any $u,v \in S$, $d(u,v) > r$, and, for any $w \in V$, there exists $u \in S$ such that $d(u,w) \leq r$.

Lemma 2 *For any $r > \rho\tau$, any ball of radius $t \geq \rho r$ contains at most $\beta\alpha^{1+\log_\rho(t/r)}$ nodes of an r-net in V.*

Let Δ be the aspect ratio of d. We say that a routing scheme has stretch (a,b) if the path length of the routing is at most $aD+b$ between any two nodes at distance D. Each node has an $O(\log n)$ bits ID. The following theorem demonstrates the existence of a low stretch compact routing scheme.

Theorem 1 *For any $0 < \varepsilon < \delta/\rho$, there exists a compact routing scheme in (V,d) with stretch $(\frac{\rho_s}{1-\rho\varepsilon}, \tau\rho^2 \log n)$, table size $\beta\alpha^{3+\log_\rho(\frac{1}{\varepsilon})} \log_\rho(\frac{\Delta}{\tau}) \log n$ bits per node, node label size $O(\log_\rho(\frac{\Delta}{\tau}) \log n)$ bits.*

Proof sketch To prove this theorem, we show how Slivkins routing scheme [7] can be extended to our setting. The crucial point consists in using $\rho^i\tau$-nets to build a hierarchy of nets with which to design routing tables. The analysis of the routing scheme designed then yields the result using Inframetric properties.

Corollary 1 *For any positive $\varepsilon < 1/\rho$, there exists a compact routing scheme in (V,d) with stretch $\frac{\rho}{1-\rho\varepsilon}, \tau\rho^2 \log n)$, table size $\beta\alpha^{3+\log_\rho(\frac{1}{\varepsilon})} \log_\rho(\frac{\Delta}{\tau}) \log n$ bits per node, node label size $O(\log_\rho(\frac{\Delta}{\tau}) \log n)$ bits.*

Références

[1] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *SIGCOMM*, pages 251–262, 1999.

[2] P. Fraigniaud, E. Lebhar, and L. Viennot. The inframetric model for the internet. In *INFOCOM*, 2008. To appear.

[3] P. Francis, S. Jamin, C. Jin, Y. Jin, D. Raz, Y. Shavitt, and L. Zhang. IDMaps : a global internet host distance estimation service. *IEEE/ACM Transaction on Networking*, 9(5) :525–540, 2001.

[4] T. M. Gil, F. Kaashoek, J. Li, R. Morris, and J. Stribling. P2psim : a simulator for peer-to-peeer protocols.

[5] P. K. Gummadi, S. Saroiu, and S. D. Gribble. King : estimating latency between arbitrary internet end hosts. *Computer Communication Review*, 32(3) :11, 2002.

[6] J. Kleinberg, A. Slivkins, and T. Wexler. Triangulation and embedding using small sets of beacons. In *FOCS*, pages 444–453, 2004.

[7] A. Slivkins. Distance estimation and object location via rings of neighbors. In *PODC*, pages 41–50, 2005.

[8] B. Wong, A. Slivkins, and E. Gün Sirer. Meridian : a lightweight network location service without virtual coordinates. In *SIGCOMM*, pages 85–96, 2005.

[9] E. W. Zegura, K. L. Calvert, and S. Bhattacharjee. How to model an internetwork. In *INFOCOM*, volume 2, pages 594–602, 1996.

[10] B. Zhang, T. S. E. Ng, A. Nandi, R. Riedi, P. Druschel, and G. Wang. Measurement based analysis, modeling, and synthesis of the internet delay space. In *6th Internet Measurement Conference (IMC)*, pages 85–98, 2006.