行政院國家科學委員會專題研究計畫　成果報告

網際網路的連結結構:分析與模式建構(2/2)

計畫類別：個別型計畫
計畫編號：NSC93-2213-E-002-026-
執行期間：93年08月01日至94年07月31日
執行單位：國立臺灣大學電機工程學系暨研究所

計畫主持人：黃寶儀

計畫參與人員：徐榮德、蔡明倉、胡書瑜、林宗翰

報告類型：完整報告
報告附件：出席國際會議研究心得報告及發表論文
處理方式：本計畫可公開查詢

中　華　民　國 94年12月10日

行政院國家科學委員會專題研究計畫 期末完整報告
網際網路的連結結構：分析與模式建構 (2/2)
計畫編號：NSC 93-2213-E-002-026
執行期間：93 年 8 月 1 日 至 94 年 7 月 31 日 (第二年)
執行單位：國立台灣大學電機工程學系暨研究所
計畫主持人：黃寶儀
計畫參與人：徐榮德、蔡明倉、胡書瑜

一、中英文摘要

本研究的主要目的在探討網際網路連結架構(Internet topology)在 Autonomous System(AS)層面的各種特徵與模式的建構。我們的做法有別於既存的“統計爲主直覺爲輔”(statistics and intuition)方法，以圖學(spectral graph theory)爲出發點，將網際網路的連結予以嚴謹的理論分析。工作主要分三項：1)圖學上與網際網路相關的量測值(metrics)研究、2)本團隊提出的 Normalized Laplacian Spectrum(nls)用於鑑別網際網路連結架構與其他圖學良測值的比較、及 3)網際網路連結架構的模型建構與產生器的編寫。第一年的研究進度略超過預期、各項工具程式的編寫已大致底定、並且已經開始分析比較的研究。期間、我們也發現了目前運算設備上的瓶頸。由於網際網路的成長迅速，節點與節線數量大幅增加。經過最佳化之後，運算所需的記憶體與 CPU cycle 量極高。針對這個問題，我們也藉期中報告的機會一併提出在電腦設備上小幅增加預算。

關鍵詞: 網際網路連結結構, 圖學, 連結結構產生器

Our goal is to investigate the various characteristics and to establish a high-fidelity model of the Internet topology at the autonomous system (AS) level. Our approach is distinctive of the existing methods aided by statistics and experts' intuition. In this project, we propose to analyze the Internet topology using rigorous graph theory.
The project execution is constructed in 3 parts: 1) the study of the Internet graphs using graph-theoretical metrics, 2) the effectiveness and comparison of the use of Normalized Laplacian Spectrum (nls) to distinguish the Internet topology, and 3) the modeling of the Internet topology and topology generation. We have accomplished more than we anticipate for the first year. As of the time this mid-term report is fired, we are about to finish all implementation of the graph theoretical analysis tools and have begun the analysis and the modeling of the Internet topology. During the year, we also found a bottleneck in computation. Due to the excessive growth of the Internet, the size of the topology in our analysis has exceeded the capacity of the computing equipment we have. After optimizing our tool implementation, the

extensive use of memory and CPU cycles still poses a serious challenge. Therefore, we would like to take advantage of the opportunity to plead for a raise on the computing equipment budget to accommodate the unexpected computation resource problem.

Keywords: Internet Topology, Graph Theory, Topology Generator

二、計畫目標與規劃

We have identified weaknesses in existing Internet topology models and generators that demonstrate the need for further research. Given the main problem of the existing Internet topology analysis is associated with the bias view of the Internet topology from the statistics and common wisdom, we set out in the project to take a more robust graph theoretical approach. ***Our goal*** is four-fold:

1) Using the graph theoretical metrics to determine the relevant characteristics of the AS graphs and to compare the actual AS graphs to the synthesized graphs from the existing topology generators
2) Computing the normalized Laplacian spectrum (*nls*) of AS graphs, i.e., the eigenvalues of the normalized Laplacian matrix and verifying the suitability of nls as the Internet topology fingerprint
3) Defining an graph-theory-inspired model of the Internet topology and implementing the model as a graph generator for Internet-like topologies
4) Using the graph theoretical metrics developed in 1) and 2) to verify whether the model obtained in 3) is better than the existing models

In this 2-year project, we propose an in-depth study of the Internet topology on the autonomous system (AS) level, referred to as the AS graph. The ***project execution*** is dissected into three parts. We concentrate, for the year one, on the *metric implementation* and the *first cut of the Internet topology model*. For the year two, we anticipate to complete the metric implementation, the graph theoretical analysis, and the *topology model and generator*. Listed below are specific items to accomplish and those <u>underlined</u> are items completed.

1) Graph theoretical metrics (100% completed)
   Single value metrics: <u>number of nodes, number of links, average degree, clustering coefficient, cardinality of matching</u>

Distribution metrics: rank-degree, degree-frequency, expansion, distortion, resilience, link value

2) Spectral graph theoretical metrics (100% completed)

   Single value metrics: ratio of pendant nodes, ratio of quasi-pendant nodes, ratio of isolated-inner nodes, ratio of quasi-pendant links, average degree of the quasi-pendant component

   Distribution metrics: normalized Laplacian spectrum

3) For the graph theory inspired model: we further phase the process in 4 stages. (100% completed)

   The 1st stage: classifying the pendant, isolated-inner, and residua-inner nodes into the leaf (outskirt) component of the Internet topology and the quasi-pendant nodes into the core component

   The 2nd stage: the generation of the core and leaf clouds and interconnecting the two clouds

   The 3rd stage: the generation of the intra-core links

   The 4th stage: the generation of the intra-leaf links

4) Verification of our own Internet topology model (100%)

   This requires completion of the above 3 tasks

三、分析與討論

Internet AS Graph Analysis

As the Internet evolves and concerns many aspects of our lives, a better understanding of its topological structure is increasingly critical in that it provides not only the insights to protocol designs, but also the confidences in network simulations. There has been a significant amount of work analyzing the Internet topology at the AS-level which gives rise to a number of topology models. Each of these models synthesizes graphs that match to a certain graph metric at the time of analysis. For instance, inet3.0 [16] and TANG [1] are Internet AS-level graph generators based on the power-law relationships [5]. Prior work [10] has shown that the graphs generated by the power-law-based generators better match the properties of the Internet and they capture implicitly the structure of the Internet AS-level topology.

While we agree that the power-law relationships are important qualitative properties, we do not think the power-law relationships tell us all about the AS-level topology. In fact, a recent study shows evidences that the power-law relationships are easily observed in the presence of bias [7]. In addition, there are two pitfalls in graph generation using the power-law-based models. The first is that the graphs generated do not contain structural semantics. In other words, we will not be able to distinguish whether a node is a stub AS or a transit AS which makes it difficult to assign traffic sources and sinks in network simulations. Secondly, the generated graphs are inherently large to reproduce the power-law properties. These graphs are often too large to be practical to simulate with an average network simulator [2].

To advance the state of Internet topology modeling, we think a more plausible approach should be to take a step back and to examine this fundamental question - are there Internet topology invariants? We share the same viewpoint towards high-confidence Internet simulations as advocated in [6]. For an immense moving target such as the Internet, a robust topology model needs to build upon properties that will continue to hold into the future. Motivated to address this question, we examine how the Internet AS-level topology changes from 1998 to 2004 using a variety graph metrics. Our methodology is to 1) apply the various metric computations on the monthly Internet AS-level topology data, 2) examine the evolving trends with respect to the metrics, and 3) observe if the AS-level topology shows a clear converging trend in any of the metrics.

The metrics in consider as identity are classified into two categories, the single-value and the distribution-based metrics. The single-value metrics include the average degree, average path length, and clustering coefficient [13]. The distribution-based metrics include the rank-degree and degree-frequency power-law relationships in [5], the expansion, resilience, distortion and link value used to compare Internet graphs in [10], and the normalized Laplacian spectrum (nls) proposed in [12] as a candidate Internet graph fingerprint. We find that the AS-level topology does not show any clear trend in metrics such as average path length, clustering coefficient, expansion, distortion, and resilience. We do observe, however, either a clean increasing or a decreasing trend in the average degree, power-law properties, and link value. Remarkably, the AS-level topology shows a conversing trend in the normalized Laplacian spectrum (nls). The nls of the AS-level topology has become steady and remained similar from January 2001, which suggests that the nls is a promising topology invariant.

Furthermore, from the theory of nls [4], we find that the conversing trend indicates the ratio of the leaf and core ASs on the Internet is coming to a steady state. This phenomenon correlates well to the rise and burst of the Internet bubble. Before the bubble bursts, the Internet thrives. The backbone and provider ISPs expend optimistically, and hence there are relatively more core ASs(or provider) on the Internet. After the bubble bursts, the ISPs expand conservatively and react very much based on the customer demand. This balance in supply and demand supports indirectly that the leaf-core ratio is a plausible invariant.

In this work, our contribution is two-fold. First of all, we identify nls as a promising topology invariant. This hints on a new generation of nls-based topology models from which we can generate small-scale Internet-like graphs without the constrain of the power-law. Second of all, we discover a surprising structural property. That is, the ratio of the leaf to core ASs on the Internet is converging to a steady state. The topology models can be improved by taking this structural property into consideration. As a result, we will be able to generate graphs with the structural semantics. We are confident of our analysis, because this structural property echoes the rise and burst of the Internet bubble. This suggests that the nls might be a good Internet economic indicator as well.

Internet Topology Modeling

Internet continues to grow and concerns aspects of our lives. As a result, it is increasingly significant to model the Internet topology for insights to protocol design and for confidence in network simulations. There are two classes of models for the Internet topology, structural and degree-based. The structural models are designed to model the router-level topology. Due to the difficulty of obtaining the router-level topology data, much of the recent advances concern the modeling of AS-level topology. This gives rise to the degree-based models that focus on reproducing the power-law relationships [5] observed in the AS-level topology.

The degree-based models are shown [10] promising and capturing implicitly the structural properties of the AS-level topology, compared to the structural models.

While we agree that the power-law relationships are important qualitative properties, we do not think the power-law relationships are all there are to the AS-level topology. Some show evidences that the power-law relationships are easily obtained in the presence of bias [7], either in the physical mechanism or data sampling of the system. In addition, there are two more pitfalls in graph generations for network simulations using the degree-based models. The first is that the graphs generated do not contain structural semantics. In other words, we will not be able to distinguish whether a node is a stub AS or a transit AS. Second of all, the generated graphs are inherently large in order to reproduce the power-law properties. These graphs are too large to be practical to simulate with the average network simulator, ns-2 [2].

To advance the state of Internet topology modeling, we think a more plausible approach should be to 1) explore the other fundamental invariants [6] in AS-level topology, not to be limited to the power-law relationships, 2) model explicitly the structural properties, and 3) opt for models that generate practically sized graphs. Moving towards a model that satisfies the above requirements, i.e. generating Internet-like graphs that enables high-confidence, feasible-scale wide-area network simulations, we present in this paper 1) a newly discovered topology invariant that give rise to a surprising structural property in the topology, 2) a hybrid Internet topology model (Hinet) that captures this structural invariant, and 3) a thorough evaluation of the Hinet model on its fidelity generating Internet-like graphs, practicality generating small-scale graphs and adaptability adjusting to extended topology data sets.

Our contribution is two-fold. 1) We discover a brand new topology invariant. Unlike the power-law statistical invariants, this is the first to be found that concerns the explicit structure of Internet. 2) Based on the understanding of this structural invariant, we propose a high-fidelity, practical, and adaptive model that encompasses both statistical and structural properties.

四、成果自評

We have completed the proposed work on implementing the graph evaluation metrics, the analysis, and a model with the generation tool. The analysis work, "**On the Search of Internet Topology Invariants**", and the modeling work, **Hinet: A Hybrid Model for Internet Topology**, are written as two technical reports and are currently under going revision. The two reports will be submitted for conference and then journal publication. The related graph metric computation software developed are being documented and to put available on the Web for cross reference.

五、參考文獻

[1] S. Bar, M. Gonen, and A. Wool. An incremental super-linear preferential internet topology model. In the 5th annual Passive & Active Measurement Workshop, pages 53–62, April 2004.

[2] L. Breslau, D. Estrin, K. Fall, S. Floyd, J. Heidemann, A. Helmy, P. Huang, S. McCanne, K. Varadhan, Y. Xu, and H. Yu. Advances in network simulation. The VINT Project IEEE Computer, 33(5):59–67, May

[3] 2000.

[4] H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Towards capturing representative as-level internet topologies. Technical Report UM-CSE-TR-454-02, University of Michigan Computer Science, 2002.

[5] F. R. K. Chung. Spectral Graph Theory. American Mathematical Society, 1997.

[6] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In Proc. SIGCOMM, pages 301–313, August 1999.

[7] S. Floyd and V. Paxson. Difficulties in simulating the internet. IEEE/ACM Transactions on Networking, 9(4):392–403, August 2001.

[8] Lakhina, J. W. Byers, M. Crovella, and P. Xie. Sampling biases in ip topology measurements. In Proceedings of INFOCOM, 2003.

[9] U. of Oregon Route Views Project. http://www.routeviews.org.

[10] D. Spielman. Spectral Graph Theory and its Applications Applied Mathematics 500A. MIT.

[11] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: Degree-based vs. structural. In Proc. of ACM SIGCOMM, 2002.

[12] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger.

Characterizing network hierarchy. Tech Report 03-782 University of Southern California, 2003.

[13] D. Vukadinovic, P. Huang, and T. Erlebach. A spectral analysis of the internet topology. Technical report. In the Innovative Internet Computing Systems, June 2002.

[14] D. J. Watts and S. H. Strogatz. Collective dynamics of ’small-world’ networks. volume 42, pages 393–440,

[15] 1998.

[16] E. P. Wigner. Characteristic vectors of bordered matrices with infinite dimensions. Ann. Math, 62:548–564, 1955.

[17] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. Ann. Math., 67:325–328, 1958.

[18] J. Winick and S. Jamin. Inet3.0: Internet topology generator tech report um-cse-tr-456-02, 2002.