Scaling Laws for the Internet Over Urban Regions

V.S. Anil Kumar Madhav V. Marathe
Virginia Tech

Ravi Sundaram
Northeastern U.

Mayur Thakur
U. of Missouri, Rolla

Sunil Thulasidasan
Los Alamos National Labs

SUMMARY We construct and analyze realistic maps of the Internet spanning urban regions (city-nets). Like transport network, electrical networks, etc., the Internet is one of the critical societal infrastructures. Its robustness is crucial not only for day-to-day operations but also during natural and human initiated crises events. The importance of maintaining a functional communication network spanning large urban regions was underscored during the recent hurricanes.

Using our methodology we construct city-nets for 12 of the 25 most populated urban regions in the US. Substantial effort is spent on ensuring the validity of our findings and the completeness of our data.

We comprehensively analyze the structure of the city-nets; the analysis reveal interesting scaling laws — the city-nets display a strikingly similar structure. The measured statistical properties are often different than the measured properties of both the ISP level networks and AS level networks. For example, all 12 city-nets studied share a common topological structure: A small subset of routers (*waist*) lie on the bulk of the routes into the city and within each city the routes rapidly branch out to a large number of hosts (*hip*).

Based on our structural characterization we show that city-nets are vulnerable to Distributed Denial of Service (DDoS) attacks. In contrast, we prove the counter-intuitive result that *any* large-scale attack on the Internet can be detected by monitoring only a *constant-sized* subset of routes.

DESCRIPTION Recently, there has been renewed interest in the study of complex networks. It is driven by a number of studies showing that structure of the network is important in understanding the overall behavior of complex systems; see [1, 4, 3, 8, 10] for recent results.

In this context, much attention has been given to studying the topology of the wide-area Internet at the autonomous system (AS) [12, 5] ISP [11], and router level [5]. The degree distribution measure has attained great popularity as a measure characterizing the structure of the Internet and topology generators [12] are often rated based on their ability to produce graphs with the requisite power-law distributions [4, 3]. These generated graphs are then utilized as the basis for simulations that evaluate the robustness of the Internet [1, 8, 9]. This paper is based on the premise that the above activity, though well-founded, is incomplete and inaccurate in some important respects.

Here, we undertake a systematic study of constructing and analyzing the Internet that serve large urban regions (city-nets). We see this as a step in the direction of producing improved models and topology generators for the Internet. To our knowledge, this is the first attempt to construct realistic maps of city-nets. There are at least two compelling reasons for the study.

First, metropolitan areas are hubs of human activity and the Internet plays a crucial role both in terms of economic impact as well as a means of information and communication in times of crises, man-made (New York) or natural (New Orleans). Thus a structural analysis of the cross section of the Internet (just like the AS, or ISP level cross-sections) that focuses on spatial nature of the Internet is essential — our results show that these networks display interesting similarities as well as differences as compared to AS and ISP level networks. Importantly, the comprehensive structural analysis reveals that the Internet restricted to urban areas has a form and structure that is common across cities and different from the larger Internet.

Second, most existing studies view the Internet as a graph, either directed or (more typically) undirected. In this paper we depart from the standard graph-based model of the Internet topology and instead represent city-nets as collections of paths (routes). This *path view* of the Internet is grounded in the fact that flows are governed by policy (BGP) and cannot be inferred from the underlying graph. Further, path-based measures are clearly superior to graph-based measures for studying the vulnerability of the Internet - the number of paths passing through a node (path-degree) is a better indicator than the number of links (graph degree) it is connected to.

Recently, there has been substantial interest in uncovering scaling laws in natural and engineered systems Inspired by the ”first-principles” approach of [7] we

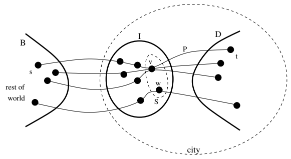

Figure 1: BID model for city-nets. Small solid circles denote individual IP addresses. Thin solid lines denote paths. The path labeled P starts at node s in B (outside the city) and ends at t (inside the city).

present and validate the hypotheses to account (at least in part) for this structure. Combined with the work of [7], this provides a basis for a new class of generative models for city-nets. Our approach consists of three main steps.

Step 1: Mapping City-Nets: We use traceroute-based techniques to infer IP paths between the outside world and the end hosts in city-nets of 12 of the 25 most populous cities in the United States. A path-based view of city-net map as a *BID model* (see Figure 1) is used in our analysis. To identify the geographic location of IPs we use *geographic data* from Akamai, Digital Envoy, and Quova. We collect traceroute information from several vantage points to a carefully selected representative subset of IPs within the city. We combine this data with Skitter [2]. The BID models for 12 city-nets use a total of over 2 million individual traceroutes.

Step 2: Structural Properties of City-Nets: Next, we study classical structural parameters (such as the degree distribution) of the city-nets in our study and propose new path-based measures such as depth of end hosts in a city-net, pathdegree distribution, and waist and hip of a city-net. As illustrations of general findings, we show that the indegree and outdegree distributions have the following properties: (i) they obey power laws across city-nets (with exponents in a tight range), (ii) the exponent of the power law is *different* than the global Internet [4] as well as the ISP networks [11]. The path degree distribution are also similar across cities and obey a power law. We find that over 80% of all paths into a city use a small number (between 8 and 87, median of 35) of routers. We term this the *waist* of the city. Inside the city-net, the traffic disaggregates and there is a massive disaggregation right near the end-hosts. We call this the *hip* of the city. We explain the waist/hip structure of city-nets in two different ways. First we show that in the BID model, if the indegree and outdegree distributions are power laws with exponents that we have measured, then the ratio of I-D edges to B-I edges is $\Theta(\log N)$, where N is the number of I nodes. Second, in the spirit of [7]: we conjecture that the waist/hip structure is a direct result of economic factors and (partially) validate the *winner-takes-all hypothesis* by discovering that a small number of ISPs (between 6 and 16, median of 8) together carry more than 99% of the paths into the city-nets. We formulate the *apartment hypothesis* which states that end hosts are organized into large blocks served by a common last hop router (similar to the organization of physical access to apartments in an apartment complex).

Step 3: Robustness and Monitoring: We discover that, the city-net is an order-of-magnitude more vulnerable than the global Internet i.e., a much smaller fraction of nodes need to be removed to disrupt a given fraction (say 80%) of the incoming paths (or traffic, assuming our uniform block traffic weights.) We find a high correlation between the vulnerability of a city and its population size. In light of our findings that a small number of targeted node and edge failures might result in a significant disruption of the traffic, we look at how such attacks can be detected. Inspired by [6] we employ the concepts of VC-dimension and ϵ-nets to obtain the surprising result that there exists a *constant-sized* set of paths that is guaranteed to be affected by any large-scale attack. Thus, by frequently monitoring this subset of paths one is guaranteed to detect any and all large-scale disruptions.

References

[1] R. Albert and A. Barabasi. Statistical mechanics of complex networks, *Rev. Mod. Phys.* 74, pp. 47-97, (2002).

[2] K. Claffy, T. Monk and D. McRobb. Internet Tomography. *Nature* 1999.

[3] Q. Chen, H. Chang, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. The origin of power-laws in internet topologies revisited. In *INFOCOM-02*, 2002.

[4] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. *Computer Communication Review*, 29(4):251–262, 1999.

[5] R. Govindan and H. Tangmunarunkit. Heuristics for Internet map discovery. In *INFOCOM-00*, 2000.

[6] J. Kleinberg. Detecting a network failure. In *FOCS 00*, pages 231–239, 2000.

[7] L. Li, D. Alderson, W. Willinger, and J. Doyle. A first-principles approach to understanding the Internet's router-level topology. In *SIGCOMM-04*, 3–14, 2004.

[8] M. Newman. The structure and function of complex networks, SIAM Review 45:167-256, 2003.

[9] C. Palmer, G. Siganos, M. Faloutsos, C. Faloutsos, and P. Gibbons. The connectivity and fault-tolerance of the Internet topology. In *NRDM-01*, 2001.

[10] R. Pastor-Satorras, A. V azquez, and A. Vespignani. Dynamical and correlation properties of the internet. *Phys. Rev. Lett.* 87, 258701 (2001).

[11] N. Spring, R. Mahajan, D. Wetherall, and T. Anderson. Measuring ISP topologies with Rocketfuel. *IEEE/ACM Transactions on Networking*, 12(1):2–16, 2004.

[12] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: degree-based vs. structural. In *SIGCOMM-02*, 2002.