Approximating Centralization and Hop Density in Internet Traffic with Traceroute

Andrew S. Gardner ‘14 – (Sponsor: Professor Cynthia Conti)

CLARK UNIVERSITY 1887

Abstract

The Internet is comprised of a wide range of autonomously operated node clusters. Although it generally adheres to a decentralized design, the structure of these interconnected networks still encompasses a certain degree of centrality. This can be evidenced through the application of social network analysis to uncover densely populated nodes and transit routes. Here, we seek to adapt the traceroute command, which allows one to approximate the physical network paths connecting any two specified Internet nodes, to measure the density of use among particular nodes. In order to achieve these means, traceroute testing was performed on a random sample of 50 popular websites from 13 distinct testing locations. Results show that physical infrastructures owned by powerful network actors were more often utilized for transmitting information flows. By exploring these concepts, this study aims to develop a hop density theory of autonomous systems. Individual nodes that attract a large number of hops are highly dense and can therefore be regarded as having more power within the larger network.

Introduction

The Internet is a collection of geographically dispersed networks that frequently communicate with each other. Each node can be perceived to represent one point of connection, such as a computer server, that exists within a larger cluster of locally unified machines. All of the individual nodes that make up these local clusters, which are formally called autonomous systems, are centrally controlled by the same administrative body [3]. It is common for Internet traffic to travel through many different autonomous systems before it finally reaches its intended location. The decentralized structure of the Internet ensures that information doesn’t flow in a completely consistent fashion. Physical network paths that data packets traverse when navigating to a particular website are likely to vary upon each request of this information. Despite these democratic ideals, the application of social network analysis suggests that a certain degree of centrality can still be evidenced within the information flow of this study.

Theory

Social network analysis is employed to capture patterns of information exchange between the linked actors in a network. While it has been established that each autonomous system on the Internet is comprised of individual nodes, the communicative relationships that form between autonomous systems can be regarded as ties. Within this study, the frequency of contact between particular autonomous system nodes will be examined as an important indicator of tie strength [2]. By empirically analyzing common information routes where data flow, we can better understand where power is structurally concentrated on the Internet.

	Alexa Global Rank	Website		Alexa Global Rank	Website
1	2214	bbandt.com	**26**	570452	arrangeyourpc.com
2	106747	uwhealth.org	**27**	577351	ozq.net
3	123703	usagoals.me	**28**	585576	cr8corp.com
4	141196	dailybrainteaser.blogspot.in	**29**	619733	you-fit.ru
5	183822	spicyscripts.com	**30**	630207	superbillets.com
6	188840	vintageadbrowser.com	**31**	664039	pa-educator.net
7	213255	idm.net.lb	**32**	684347	iwaya-ski.com
8	220454	anime-chu-2.com	**33**	694611	trendrender.com
9	229310	gesund-heilfasten.de	**34**	711020	belgica-turismo.es
10	268103	senha-yuzu.jp	**35**	727066	trade-in-platform.com
11	271662	vip.co.id	**36**	755320	read19.com
12	293956	myitstudy.com	**37**	781461	lunchactually.com
13	313326	insightsonindia.com	**38**	788039	fullmovierulz.com
14	318489	gbemembers.com	**39**	790894	greentech-germany.com
15	321542	gfcnieuws.com	**40**	793585	amule.org
16	328341	tradeinchecker.com	**41**	915521	floriancrouzat.net
17	342639	sbbsnet.se	**42**	947047	lufkindailynews.com
18	398763	freewheelbicyclestore.com.au	**43**	948780	bestbizleads.com
19	403446	hochzeitenundflitterwochen.de	**44**	948971	gonoreya.ru
20	452902	kartenmeister.com	**45**	954552	socialmart.ru
21	461477	mieadham86.blogspot.com	**46**	968711	smartinfosys.com
22	483691	downloadur.org	**47**	976791	howto-growtallernaturally.com
23	499884	itsmaca.com	**48**	980592	lapa.co.za
24	523977	stella-berlin.de	**49**	997815	togozik.com
25	564978	dazebaonews.it	**50**	997978	vprognoze.net

Fig 1: Random Sample of 50 Websites.

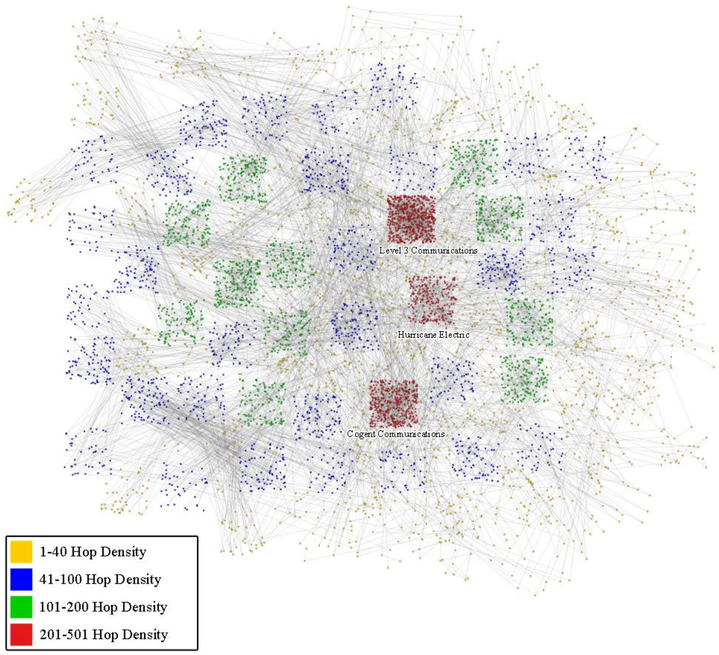

Fig 3: Social Network Analysis.

Methods

This study used traceroute software to empirically identify common trends of decentralized information flow over the Internet. Traceroute is a utility program that can specify which autonomous systems that Internet traffic traversed while en route to a particular website [1]. Each traceroute test identifies a physical path of individual autonomous system nodes, which are formally called hops, that fall between two specific network locations.

A random sample of 50 websites was extracted from the Alexa Global Top Sites dataset, which hierarchically ranked the 1,000,000 most heavily trafficked websites during October of 2013. The websites chosen for inclusion in this study, which are exemplified in Fig. 1, were designated as the target locations that would be traced to. In order to determine how physical geography alters the flow of transnational Internet traffic, 13 separate testing locations were designated as points of origin where testing procedures were initiated from. These points of origin were located at; Clark University (U.S.A.), The Institute of High Energy Physics (China), TruTeq Wireless (South Africa), Universidad Estatal Paulista (Brazil), Host Europe (Germany), KC Internet (New Zealand), Nexlinx (Pakistan), Telstra Internet (Australia), RUSnet (Russia), University of Southern California (U.S.A.), RHnet (Iceland), Sdv Plurimédia (France), and HiNet/ Chunghwa Telecom (Taiwan).

Traceroute testing was conducted to approximate the network paths between these 13 points of origin and our sample of 50 target websites. Testing procedures at each starting location, with the exception of Clark University, were performed remotely via the Internet on publicly accessible traceroute servers. Finally, the resulting data concerning each node was processed in a geolocation database (MaxMind GeoIP2) in order to determine which autonomous system it belonged to.

Results

Throughout the course of this research, 509 individual traceroute tests were executed, and 6391 hops were subsequently identified. Fig. 2 lists the frequency of hops that are attributed to 178 different autonomous systems. Organizations with a large frequency of hops can be regarded as highly dense because they have been traversed the most within this study. The three autonomous systems with the highest hop density were: Level 3 Communications (501 hops), Cogent Communications (354 hops), and Hurricane Electric (214 hops).

Discussion

Social network analysis was employed to determine if areas of centrality can still be evidenced within the Internet’s decentralized network structure. As is demonstrated in Fig. 3, the autonomous systems with the highest hop density, which are denoted in red, encompass a great deal of centrality in relation to other entities in the sample. These findings have implications for the principle of network neutrality, which contends that no single organization should have the ability to restrict how society accesses information over the Internet. The results from this study seem to indicate that certain networks are clearly being favored over other networks.

Organization	Frequency
Level 3 Communications	501
Cogent Communications	354
Hurricane Electric	214
CHTD, Chunghwa Telecom Co., Ltd.	184
Google	157
1&1 Internet AG	147
Telstra Internet	145
Hetzner Online AG	144
Sdv Plurimedia	139
Pacnet Services (Japan) Corp.	132
NORDUnet	124
Clark University	105
Transworld Associates (Pvt.) Ltd.	104
China Science And Technology Network	103
Abovenet Communications	98
Hosteurope GmbH	97
Tinet SpA	96
National LambdaRail, LLC	95
Neo Telecoms S.A.S.	89
Posix Systems (Pty) Ltd	89
CENIC	83
Institute of High Energy Physics	80
Host Europe GmbH	75
Deutsche Telekom AG	71
St. Petersburg State Technical University	71
Init Seven AG	68
Charter Communications	65
Softlayer Corporate C	65
Ace Data Centers, Inc. (ADC-96)	65
Link Telecom (NZ) Ltd	60
TeliaSonera International Carrier	60
TeliaSonera AB	59
AT&T Worldnet Services	53
Amsterdam Internet Exchange B.V.	47
California State University Network	46
Telia International Carrier	46
University of Wisconsin Madison	46
DE-CIX Frankfurt IXP	44
Suddenlink Communications	44
Telus Communications	44
Florida International University	43
SURIS/RHnet Iceland University Research Network	42
Universidade Estadual Paulista	42
Autonomous System Number for Nexlinx	41
Russian Federal University Network	41
KC Computer Service	40
NTT America	40
University of Southern California	40
FundaCAo De Amparo A Pesquisa Do Estado SAo Paulo	39
nLayer Communications Internal/Backbone	38
NPO RUSNet Ltd.	38
Unified Layer	37
Telstra Global	36
Zirato Rechenzentrum, Berlin	35
euNetworks Managed Services GmbH	33
Telecom New Zealand Ltd	33
VimpelCom	33
Federal University Computer Network	32
GMO Internet,Inc.	32
Private Joint Stock Company datagroup	32
Belgacom Skynet	31
Internet Initiative Japan Inc.	31
Harvard University	30
Amazon.com	29
TELECOM ITALIA SPARKLE S.p.A.	29
Worcester Polytechnic Institute	29
FloridaNet	28
Media Temple	28
Equinix	25
Keyweb AG	25
MV Communications	25
Reach Networks HK Ltd Network Blocks	25
CyrusOne LLC	24
Net2EZ	24
NForce Entertainment B.V.	24
CloudFlare	23
London Internet Exchange (LINX)	23
Node4 Limited	23
ZAO ELTEL	22
IncoNet Data Management sal	22
SingTel Internet Exchange	21
Associacao Rede Nacional de Ensino e Pesquisa	20
Kddi Corporation	20
Limelight Networks	20
Telekomunikasi Indonesia International, Pte.ltd	20
IU America	18
NZGate Aggregate Networks	18
ELInet IP Backbone Network	17
Filanco.ru	17
Netway Communications Ltd	17
Node4 UK Hosting	17
Telstra Global Internet Services Network Blocks	17
HiNet Service Center in U.S.A	16
AT&T Services	15
K-Opticom Corporation	15
Gkg.net	14
Terrenap Data Centers	14
Plus.line AG	13
Sprint PCS	13
TripartZ B.V	13
Beyond The Network America	12
NetIRD	12
OSN Online Service Nuernberg GmbH	12
State Institute of Information Technologies and	12
Reget Ltd	11
Hosting Telesystems network	11
Vodien Internet Solutions Pte Ltd	11
Wischnet	11
Belgacom International Carrier Services SA	10
IncoNet-Data Management s.a.l.	10
NFrance Conseil	10
Pugmarks - Pugmarks	10
SAKURA Internet Inc.	10
Filanco, ltd.	9
Sprint	9
Telkom SA Limited	9
Value Core	9
TATA Communications (Canada) Ltd.	8
AGT Advanced Communications	7
AT&T Internet Services	7
Cable & Wireless Telecommunication Services GmbH	7
Comcast Cable	7
Customer PtP BKL	7
General Telecommunication Organization	7
Le Web	7
NTT America - Global IP Network	7
nLayer Communications	6
PT Telkom Indonesia	6
Savvis	6
Sprint/United Information Service	6
TATA Communications	6
Tinet	6
Cable & Wireless UK P.U.C.	5
Cyrus One	5
DSCI Corporation	5
Internet2	5
One.com A/S	5
RETN Limited	5
ReTN's Backbone	5
Singapore Telecommunications Ltd	5
StPetersburg Internet Exchange	5
Almond Oil Process, LLC.	4
Cable & Wireless Americas Operations	4
China Next Generation Internet Beijing IX	4
Cybericon	4
Network for Backbone p2p	4
Verizon Business	4
XO Communications	4
Comcast Business Communications	3
EDPNET	3
France IX Services SASU	3
NTT Communications	3
NTT Europe - DE	3
Nucleo de Informacao e Coordenacao do Ponto BR	3
Belgacom International Carrier Services	2
Branch Banking and Trust Company	2
Global Crossing	2
IU Internet	2
Japan Internet Exchange Co.	2
Monster-Host.com	2
National LambdaRail	2
Sedo GmbH	2
Verio NL amsterdam facility	2
Ace Data Centers	1
ASP Privet online	1
BHARTI Airtel	1
Bharti Broadband	1
Cable & Wireless	1
Danish Network for Research and Education	1
Equinix Asia Pacific Pte Ltd	1
Global-Gateway Infrastructure	1
IP STATIC	1
NTT Europe - London	1
Russian Institute for Public Networks	1
ServerBeach	1
Unet Pipex	1
Verizon Nederland B.V.	1
VzB Services BV	1

Fig 2: Hop Frequency.

References

[1] Feng, Guangchao Charles, and Steve Zhongshi Guo. 2012. "Tracing the route of China’s Internet censorship: An empirical study." Telematics and Informatics (0). doi: http://dx.doi.org/10.1016/j.tele.2012.09.002.

[2] Haythornthwaite, Caroline. 1996. "Social network analysis: An approach and technique for the study of information exchange." Library & Information Science Research 18 (4):323-342. doi: http://dx.doi.org/10.1016/S0740-8188(96)90003-1.

[3] Stewart, C., and J. I. Khan. 2012. "A Large Scale Evolutionary Analysis of the Internet Autonomous System Network." Information Technology: New Generations (ITNG), 2012 Ninth International Conference on, 16-18 April 2012.