



**CENTER FOR APPLIED
CYBERSECURITY RESEARCH**

INDIANA UNIVERSITY
Pervasive Technology Institute

Geography of Internet2 Netflow

David A. J. Ripley

Indiana University, Center for Applied Cybersecurity Research.

Tony H. Grubescic

Indiana University, Department of Geography.

Timothy C. Matisziw

*University of Missouri, Department of Geography/Department of Civil and
Environmental Engineering.*

- *geography*
- **1. a.** The science which has for its object the description of the earth's surface, treating of its form and physical features, its natural and political divisions, the climate, productions, population, etc., of the various countries.

[Oxford English Dictionary, 2nd Edition, 1989.]



CENTER FOR APPLIED
CYBERSECURITY RESEARCH

INDIANA UNIVERSITY
Pervasive Technology Institute

- Historically – spatial and temporal movement of goods, money, people and *information*.
- The Internet doesn't move goods or people directly, but it can influence their movement.
- The Internet transports information.
 - Information \approx Money

- Access and Awareness
 - What kinds of data are available?
 - How are they acquired, from whom?
 - What are their strengths and limitations?
- Unfamiliar territory
 - Opportunities!

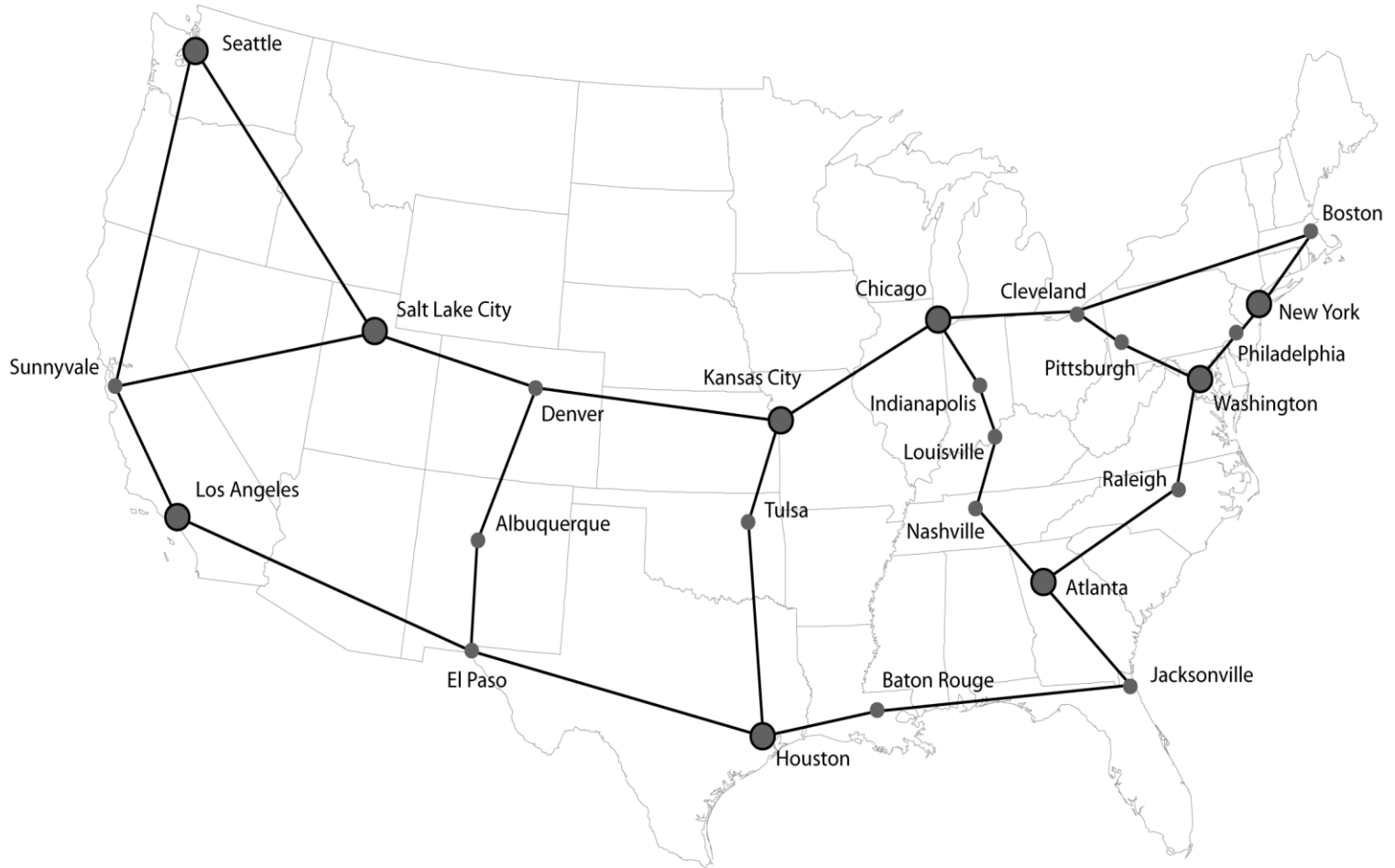
- Interest in Internet traffic volumes
 - Backbone traffic levels not interesting enough
 - IP-level a bit too specific
 - Policy/access difficulties
 - Interested in relative connectedness (and volume of traffic) between locations.
 - For certain values of “location.”

- Internet2 Network Netflow records:
 - Sourced from 9 core routers
 - 100:1 Sampling
 - /21 Anonymization required
 - ASN is not present/reliable
 - Includes CPS
 - So it's not just academic traffic.
 - Typically approx 2 billion records per day.



CENTER FOR APPLIED CYBERSECURITY RESEARCH

INDIANA UNIVERSITY
Pervasive Technology Institute



- Needed some way to identify source and destination locations
 - The policy-mandated /21 anonymization presents a problem.
 - As does the lack of ASN
- How much of a problem is the anonymization?



CENTER FOR APPLIED
CYBERSECURITY RESEARCH

INDIANA UNIVERSITY
Pervasive Technology Institute

- IP Anonymization – how much of a problem is it?
 - Surprisingly, not quite as much as we thought.
 - We’re not interested in specific hosts.
- 12,000 prefixes announced on I2 Network;
 - 36% are 21 bits or shorter
 - Assuming constant utilization*, that’s 98% of addresses unambiguously identified to the level of an AS.



**CENTER FOR APPLIED
CYBERSECURITY RESEARCH**

INDIANA UNIVERSITY
Pervasive Technology Institute

- Admittedly, there are some assumptions that definitely need to be tested.
- But still not as bad as we initially thought it might be.

- ASN is a reasonable identifier for our purposes.
 - It may be a very poor way to determine location.
 - Local vs. national organizations; IU vs. Comcast.
 - But it gives us a location for the flow of information/money/whatever
 - Which may not have anything to do with the location of the end systems.

- Scripted WHOIS lookups
 - Registering entity's name and postal address.
- Geocoding
 - Most GIS packages will do this, with greater precision than we were interested in.
 - Centroid of ZIP code areas.

- 24 hour period
- ~2 billion flow records
- 175,000 unique sIPs, 232,000 unique dIPs
- 29,709 ZIP code areas.
- Aggregate traffic hourly by sZIP/dZIP pair.

- Capturing Connectivity
 - *Degree of node* - simple measure of connectivity
 - number of connections from one node to another.
 - Weight with traffic volume?
 - Calculate for both ingress and egress
 - Although there are issues here with NetFlow



CENTER FOR APPLIED CYBERSECURITY RESEARCH

INDIANA UNIVERSITY
Pervasive Technology Institute





CENTER FOR APPLIED CYBERSECURITY RESEARCH

INDIANA UNIVERSITY
Pervasive Technology Institute

Table 3: Top 25 ZIP Code Areas for Averaged Ingress and Egress Flows

ZIP Code Area	City	State	Avg. Flow (Bytes)
98052	Redmond	WA	167253553.9
94043	Mountain View	CA	93077533.8
85281	Tempe	AZ	74199354.0
08054	Mount Laurel	NJ	67033097.2
20171	Herndon	VA	40396656.5
94089	Sunnyvale	CA	38924803.7
02142	Cambridge	MA	22123045.8
20910	Silver Spring	MD	19639697.8
30319	Atlanta	GA	18881723.6
08002	Cherry Hill	NJ	18804546.6
94129	San Francisco	CA	18488762.3
60510	Batavia	IL	15503621.5
95125	San Jose	CA	15380908.5
08542	Princeton	NJ	13317164.7
90025	Los Angeles	CA	12992821.1
53706	Madison	WI	11362748.7
78758	Austin	TX	11270133.0
94301	Palo Alto	CA	10583457.6
94720	Berkeley	CA	9175479.5
65201	Columbia	MO	9161357.3
02139	Cambridge	MA	8800400.8
98105	Seattle	WA	8490026.4
08540	Princeton	NJ	8344733.5
77002	Houston	TX	8334969.0
77843	College Station	TX	8202075.0



CENTER FOR APPLIED CYBERSECURITY RESEARCH

INDIANA UNIVERSITY
Pervasive Technology Institute

Top 25 ZIP Code Areas for Averaged Degree of Node Measure

ZIP Code Area	City	State	Average Connections
97403	Eugene	OR	3016.5
47907	West Lafayette	IN	2981.5
50011	Ames	IA	2922.0
52242	Iowa City	IA	2894.0
47408	Bloomington	IN	2798.5
65201	Columbia	MO	2718.5
02138	Cambridge	MA	2697.0
61801	Urbana	IL	2633.5
06269	Storrs-Mansfield	CT	2597.5
68588	Lincoln	NE	2579.0
48824	East Lansing	MI	2549.5
43201	Columbus	OH	2524.5
02215	Boston	MA	2518.0
02139	Cambridge	MA	2516.0
97331	Corvallis	OR	2424.0
06105	Hartford	CT	2385.5
43212	Columbus	OH	2366.0
50131	Johnston	IA	2364.0
70803	Baton Rouge	LA	2287.5
46556	Notre Dame	IN	2281.5
46202	Indianapolis	IN	2246.5
45221	Cincinnati	OH	2239.5
40475	Richmond	KY	2220.5
48104	Ann Arbor	MI	2213.5
74078	Stillwater	OK	2208.0



CENTER FOR APPLIED CYBERSECURITY RESEARCH

INDIANA UNIVERSITY
Pervasive Technology Institute

Top 10 ZIP Code Areas for the Origin-based Degree of Node Measure

ZIP Code Area	City	State	Unique Connections
97403	Eugene	OR	4073
02139	Cambridge	MA	4065
02138	Cambridge	MA	3966
65201	Columbia	MO	3845
47907	West Lafayette	IN	3838
68588	Lincoln	NE	3792
61801	Urbana	IL	3703
47408	Bloomington	IN	3700
50011	Ames	IA	3697
52242	Iowa City	IA	3696

Top 10 ZIP Code Areas for the Destination-based Degree of Node Measure

ZIP Code Area	City	State	Unique Connections
50011	Ames	IA	2147
47907	West Lafayette	IN	2125
48824	East Lansing	MI	2109
52242	Iowa City	IA	2092
97403	Eugene	OR	1960
55455	Minneapolis	MN	1955
06269	Storrs-Mansfield	CT	1952
47408	Bloomington	IN	1897
48104	Ann Arbor	MI	1889
06105	Hardtford	CT	1772

- Repeatability methodology
 - Approximate endpoints by... location/corporate entity.
 - Not ideal, but maybe good enough
 - Working with what's available
- Novel research community & direction
 - Additional efforts in the pipeline



**CENTER FOR APPLIED
CYBERSECURITY RESEARCH**

INDIANA UNIVERSITY
Pervasive Technology Institute

- Questions, comments?
- daripley@iu.edu