MapKIT: Mapping Key Internet Terrain

PIs: Alberto Dainotti, Georgia Tech (formerly CAIDA, UC San Diego)
Paul Barford, University of Wisconsin Madison
CNS-1705024 , CNS-1703592
http://mapkit.cc.gatech.edu

Challenge

To apply a military analogy to Internet research, the science of cybersecurity has focused heavily on weapons and tactics, but has largely neglected "terrain". **Strategic points in the macroscopic Internet topology constitute key terrain in the cyberspace battlefield**. Hackers, terrorists or states can disrupt, intercept or manipulate the Internet traffic of entire countries or regions by targeting structural weaknesses of the Internet topology. Despite much recent interest and a large body of research on cyber-attack vectors and mechanisms, we lack rigorous tools to reason about how the macroscopic Internet topology of a country or a region exposes its critical communication infrastructure to compromise through targeted attacks.

Collecting and interpreting data about the Internet connectivity, configurations and associated vulnerabilities is challenging. Due to the massive scale and broadly distributed nature of Internet infrastructure and the scarcity of publicly available data, we must resort to complex measurement and inference methodologies that require significant effort in design, implementation, and validation. **Through a novel multi-layer mapping effort, the MapKIT project aims at identifying important components of the Internet topology of a country/region**—Autonomous Systems (ASes), Internet Exchange Points (IXPs), PoPs, colocation facilities, and physical cable systems—**which represent the "key terrain" in cyberspace.**

Scientific Impact

A better understanding and modeling of key weakness of the Internet infrastructure in countries that expose them to targeted attacks by state actors and organized groups.

Data and models also relevant to resiliency of critical infrastructure (e.g., preparedness to natural disasters).

Finally, our results will enable polictical scientists to better reason about "opportunity and willingness" to engage into large-scale cyber conflict.

Country-Level Transit Influence (CTI) Metric

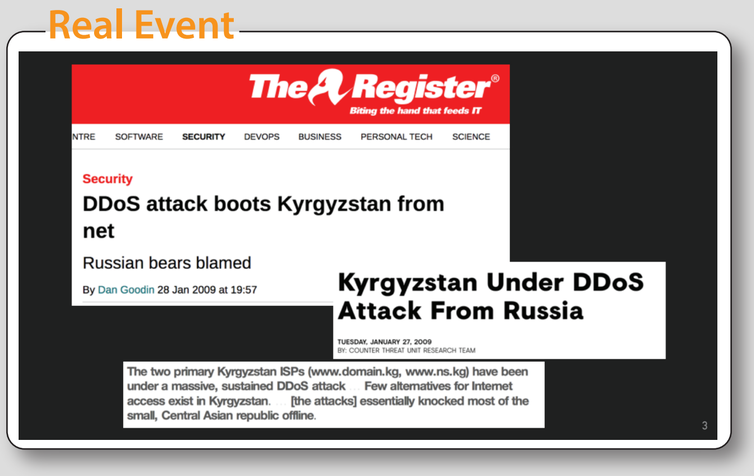
Identifying the most influential transit providers in each country that may have potential to observe or selectively tamper with a significant fraction of the country's Internet traffic. Intuitively, CTI captures the fraction of each country's announced addresses that the transit AS serves.

Country-Level Transit Influence (CTI). The *transit influence* metric is the fraction of address in announced, preferred paths (excl. backup links and less preferrred paths) towards orgin prefixes in a given country where a particular transit AS is present. In a given country, the transit influence $CTI(AS_t,C) \in [0,1]$ for each transit *AS* on country C:

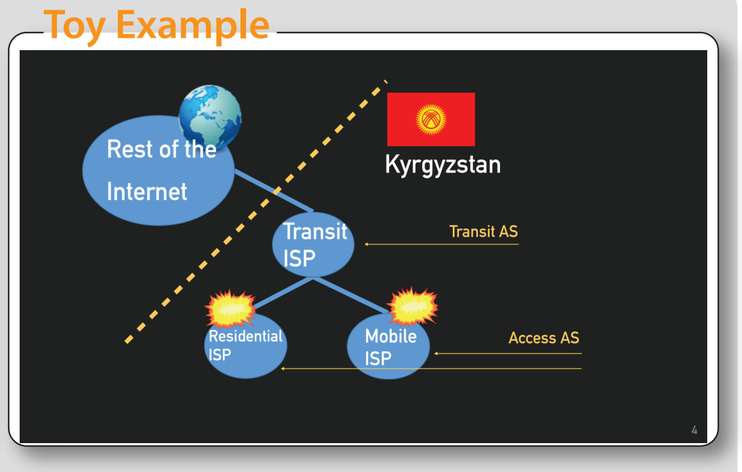
$$CTI(AS_t,C) = \frac{1}{A(C)} \sum_{p \in C} \sum_{m \in M} \frac{A(AS_t, p) \cdot w(m)}{d(AS_t, p) \cdot G}$$

where $A(AS, p)$ is the number of addresses in prefix p in country C where *AS* is present in the preferred AS path, $w(m)$ is monitor m's weight among the set of BGP monitors M and G is the number of ASes hosting BGP monitors, $A(C)$ is the total number of addresses originated by any origin *AS* that are geolocated to the country, and $d(AS, p)$ is the number of AS-level hops between *AS* and prefix p.

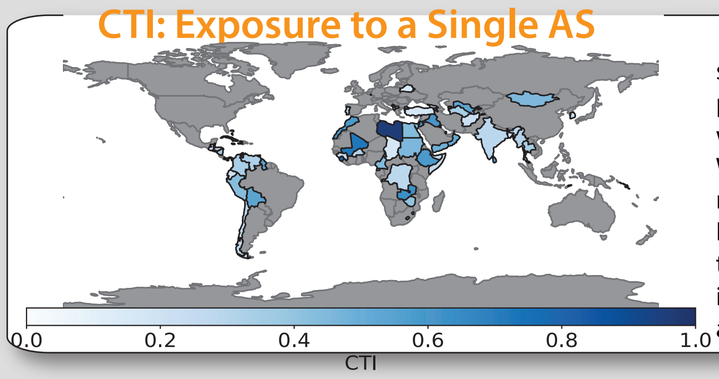
Real Event



Toy Example



CTI: Exposure to a Single AS

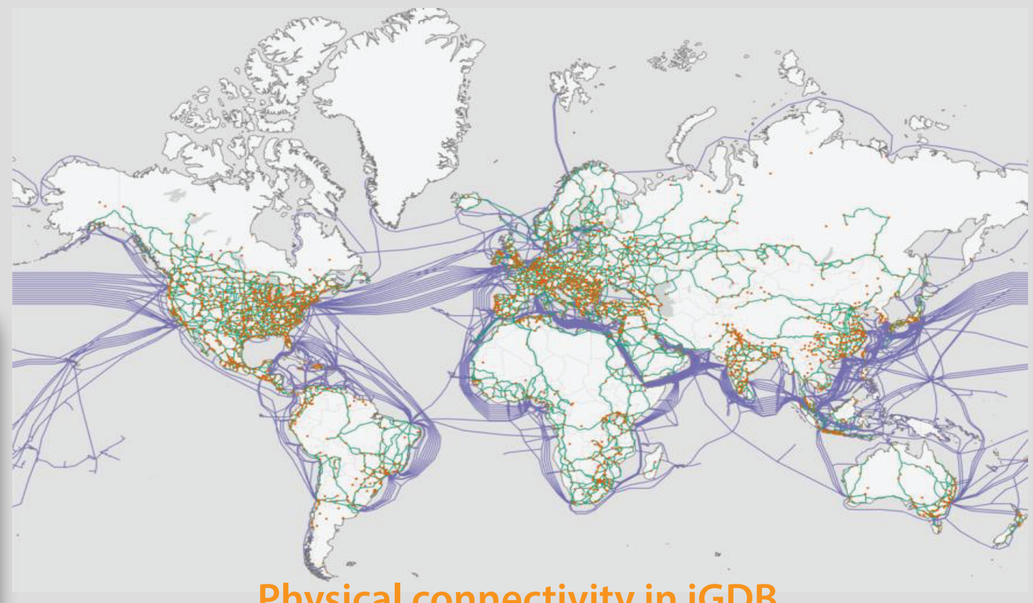


We explore the robustness of each country's AS topology to a specific scenario: how much of the country's address space is exposed if an attacker successfully compromises— e.g., infiltrates or wiretaps—a single, highly influential AS.
We infer the most influential transit networks in 75 countries that rely primarily on transit for international connectivity (1 billion Intenet users; 26% of the world). We find that many of these countries have topologies exposing them to observation or tampering: in the median case, the most influential transit network manages traffic towards 35% of the nation's IP addresses.

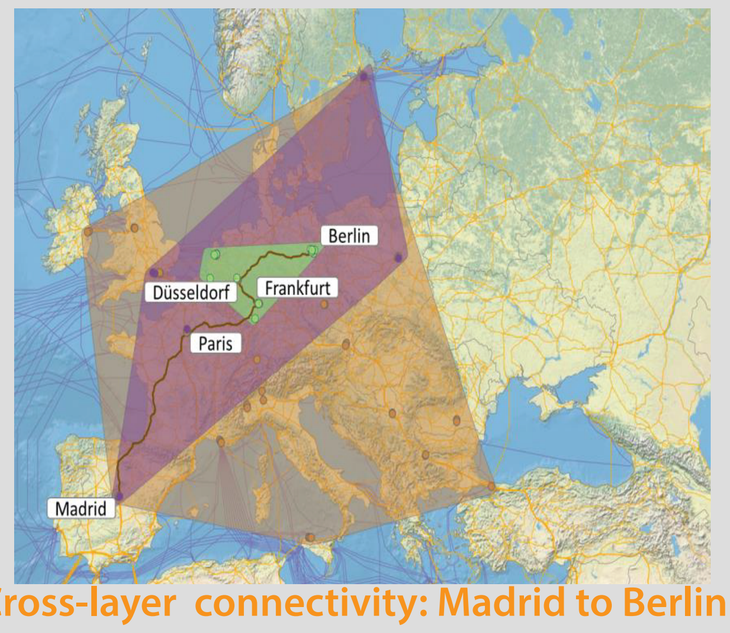
iGDB
A Database Connecting the Physical and Logical Layers of the Internet

Maps of physical and logical Internet connectivity that are informed by and consistent with each other can expand scope and improve accuracy in analysis of performance, robustness and security.

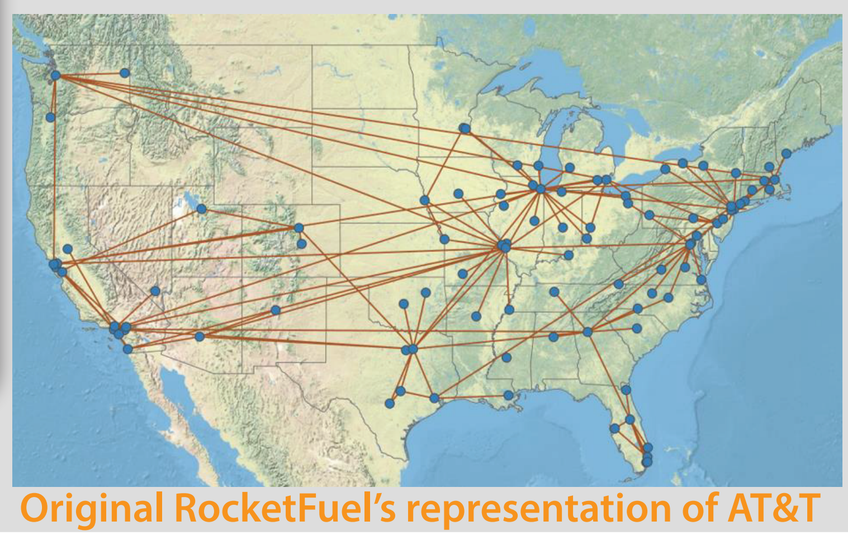
The objective of our work is to create an open repository of maps of physical and logical Internet connectivity by bringing together disparate datasets using geography as unifying feature. A key organizing principle of this repository is cross-layer consistency, which assures that nodes and links at each layer are related in a way that is consistent with standard Internet organization.



Physical connectivity in iGDB



Cross-layer connectivity: Madrid to Berlin



Original RocketFuel's representation of AT&T

iGDB's representation of AT&T (AS7018)

Team: Alberto Dainotti, Paul Barford, Zachary Bischof, Alex Gamero-Garrido, Scott Anderson, Alex C. Snoeren, Bradley Huffaker, Shuai Huao, Esteban Carisimo, Loqman Salamatian

As the Internet has become a critical infrastructure on which all other critical infrastructures depend, safety and prosperity of our society as well as international relations depend on cybersecurity. Yet the exposure of a country's Internet macroscopic infrastructure to targeted attacks with potential massive impact (e.g., in the context of cyberwarfare) is unclear. This project contributes to bridge this crucial gap.

Education and Outreach

Curriculum Development
- 3 Postdoctoral researchers
- 7 PhD Students
- 7 Master Students
- 7 Undergraduate Students

Awards
- Microsoft Dissertation Award
- IMC 2019 Distinguished Paper Award
- PAM 2022 Best Dataset Paper Award

Publications
- 26 Papers (IMC, PAM, CoNEXT, FOCI, CCR, KDD, ANRW, ...)
- 3 PhD Theses
- 32 Presentations
- 7 Code repositories