Detection and analysis of large-scale Internet infrastructure outages

http://www.caida.org/funding/dals-satc/

**summary**
Despite the Internet’s status as a critical infrastructure of our society, there is little scientific instrumentation dedicated to monitoring global Internet behavior. In particular, we have no rigorous framework for measurement, analysis, or quantifying the impact of network outages, filtering, or other abnormal connectivity dynamics on a global scale.

We have developed and demonstrated a methodology that can identify not only which networks have been affected by an outage, but also which techniques have been used to effect a deliberate disruption (e.g., control plane vs. data plane intervention). We have also developed metrics to quantitatively gauge the geographic and topological extent of impact of geophysical disasters on Internet infrastructure, and techniques to investigate the chronological dynamics of the outage and restoration. Our approach relies on:

- the extraction of signal from a pervasive and continuous source of malware-induced background radiation in Internet traffic (IBR);
- combining multiple types of data (active probing, passive IBR measurement, BGP routing data, and address geolocation and registry databases) to assess the scope and progression of the outage.

This project will result in an experimental operational deployment to validate and extend an empirically grounded methodology for detection and analysis of large-scale Internet outages. In addition to improving our understanding of how measurements yield insights into network behavior, and strengthening our ability to model large-scale complex networks, use of such a system will also illuminate infrastructure vulnerabilities that derive from architectural, topological, or economic constraints, suggesting how to mitigate or eliminate these weaknesses in future Internet architecture and measurement research.

A deployed platform will detect and monitor connectivity disruption and censorship events on a planetary scale thus enabling situational awareness of the nature and causes of network outages to national decision-makers who must determine the type and extent of proper response.

**outages caused by natural disasters**

![Map of geolocated networks affected by Tohoku earthquake and tsunami](image)

The diagram to the left illustrates the main components of the proposed system. The system will fuse pre-existing data sources collected by other Internet measurement infrastructures, creating a rich archive for subsequent processing and visualization. Processing modules will include data sanitization, classification, geolocation, AS mapping, and data aggregation. Other modules will extract and correlate statistics from these aggregated data, and store output in efficient data structures that will feed data visualization and change-point detection algorithms. A web platform will provide data visualization and a user interface through which the user can locate an event on a geographical map and track metrics, e.g., background radiation in Internet traffic (IBR); the extraction of signal from a pervasive and continuous source of malware-induced background radiation in Internet traffic (IBR); and address geolocation and registry databases to assess the scope and progression of the outage.

**architecture overview and timeline**

The diagram to the left illustrates the main components of the proposed system. The system will fuse pre-existing data sources collected by other Internet measurement infrastructures, creating a rich archive for subsequent processing and visualization. Processing modules will include data sanitization, classification, geolocation, AS mapping, and data aggregation. Other modules will extract and correlate statistics from these aggregated data, and store output in efficient data structures that will feed data visualization and change-point detection algorithms. A web platform will provide data visualization and a user interface through which the user can locate an event on a geographical map and track metrics, e.g., impact, calculated over time.

**Task 1:** investigating and defining strategies and methodologies for how to combine multiple heterogeneous data sources to detect and characterize outage events (Years 1, 2, and 3);

**Task 2:** defining (and refining) the system requirements for continuous monitoring and (near) real-time analysis of outages as they occur (will start in the second half of Year 1);

**Task 3:** testing and experimental deployment of such a system (Years 2 and 3).