A Rapid Method for Detecting Geographically Disconnected Areas after Disasters

Ling-Jyh Chen[1], Chia-Wei Li[1], Yu-Te Huang[1], and Chi-Sheng Shih[2]

[1]Institute of Information Science, Academia Sinica

{cclljj, dreamcwli, ytls}@iis.sinica.edu.tw

[2]Department of Computer Science and Information Engineering, National Taiwan University

cshih@csie.ntu.edu.tw

Abstract—In this paper, we present a novel solution called Internet Footprint Investigation (IFI) for the rapid detection of network outages after a natural or man-made disaster. IFI is comprised of two components: 1) the Active Network Probing (ANP) module, which proactively probes the network infrastructure to detect geographic areas that may be disconnected; and 2) the Reactive Footprint Search (RFS) module, a reactive mechanism that improves the accuracy of the ANP results by incorporating the footprints of location-based social networks (LBSNs) established after a disaster occurs. Using Typhoon Morakot, which struck Taiwan in August 2009, as a case study, we implement the IFI system and evaluate its feasibility in a real-world scenario. We observe that the accuracy of existing IP geolocation services is unsatisfactory, and posit that localized IP geolocation services should be deployed and maintained all the times. Moreover, we demonstrate how existing LBSNs can be used to search for disaster victims in areas reported by ANP, and identify so-called "critical areas," which have no Internet activity, for priority inspection. The proposed IFI solution is simple and effective, and it can be deployed worldwide.

I. INTRODUCTION

A disaster is a natural or man-made phenomenon that causes loss of life as well the destruction of property and the natural environment. Disaster response strategies have received a substantial amount of attention from different disciplines because, in recent years, the incidence of disasters has increased dramatically in terms of their frequency, scale/intensity, and consequent damage. Notable disasters in 2010 included the *Haiti earthquake* (magnitude 7.0 and 222,570 fatalities), the *Chile earthquake* (magnitude 8.8 and 521 fatalities), and the *Yushu earthquake* in China (magnitude 6.9 and 2,698 deaths). The first half of 2011 witnessed the *Rio de Janeiro floods and mudslides* (903 fatalities and USD1.2 billion damage), and the *Tohoku earthquake and tsunami* in Japan (at least 15,365 deaths, USD300 billion damage, and the release of radiation from the damaged Fukushima nuclear plant).

Appropriate disaster response strategies to such extreme hazards are essential. To be effective, a strategy must meet four criteria: 1) the response must be *rapid* because the situation is a matter of life and death (i.e., the so-called '72-hour golden rescue period'); 2) it must be *methodical* so that resources (e.g., first responders, equipment, and emergency services) can be allocated appropriately to locations without duplication or shortages; 3) it must be *continually updated* by exploiting the latest technologies; and 4) it requires close cooperation at the personal, community, government, and international levels.

Among the wide variety of technologies utilized in disaster response operations, we believe that data communication is the key to fulfilling the above criteria because it facilitates the dissemination and exchange of information, which is crucial for disaster victims, rescue teams, and government authorities [13, 15]. Network outages, which result in the isolation of disaster victims, may give some indication of the level of damage in the affected areas. Since the scale of a disaster must be assessed with the highest priority, a rapid network outage detection mechanism that can report outages accurately in a timely manner is absolutely essential.

In this study, we present a novel solution, called *Internet Footprint Investigation* (IFI), for the rapid detection of network outages in the event of a natural or man-made disaster. IFI is comprised of two components: the Active Network Probing (ANP) module and the Reactive Footprint Searching (RFS) module. ANP, which relies on accurate IP geolocation services, proactively probes '*network landmarks*' that may have been damaged by the disaster to identify the geographical areas that have been disconnected. Meanwhile, RFS exploits emerging location-based social networks (LBSNs) to search for Internet footprints in the areas reported by ANP, and updates the list of disconnected areas accordingly. By combining ANP and RFS, IFI can identify geographical areas affected by network outages promptly after a disaster. The information is crucial because it provides disaster response teams with a preliminary assessment of the damage, and it facilitates the rapid deployment of a provisional network infrastructure and the allocation of rescue resources.

Taking Typhoon Morakot, which struck Taiwan in 2009, as a case study, we implement the proposed IFI approach and evaluate its feasibility in a real-world scenario. We find that the accuracy of existing IP geolocation services is unsatisfactory, and we posit that localized and accurate IP geolocation services should be deployed so that they are available immediately in the event of a disaster. To this end, we present a top-down implementation of the ANP approach that combines geolocation knowledge and network diagnostic tools to measure the network topology and improve the accuracy of IP geolocation services. We also demonstrate how
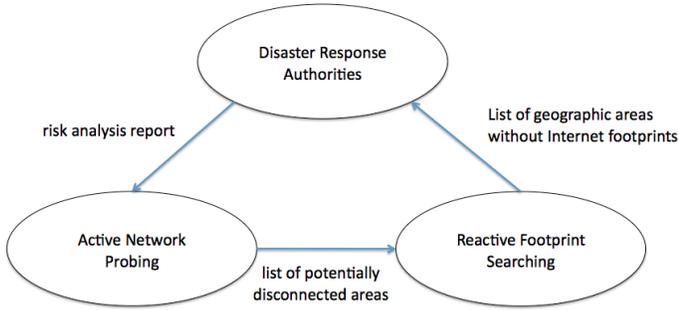
Fig. 1. The architecture of the Internet Footprint Investigation (IFI) approach

existing LBSNs (e.g., Facebook Places [1], Foursquare [2], GeoTagging Flickr [3], and GeoTagging Twitter [9]) can be used to search for disaster victims in the areas reported by ANP, and for identifying 'critical areas' that have lost Internet connectivity because of the disaster.

The contribution of this work is four-fold: 1) we propose a simple and effective approach called IFI, which can detect network outages after a disaster in a timely manner; 2) we present a top-down mechanism (i.e., ANP) to improve the accuracy of existing IP geolocation services; 3) we develop a bottom-up mechanism (i.e., RFS) that uses emerging LBSNs to identify geographically disconnected areas after a disaster; and 4) we evaluate the feasibility of the proposed solution using a real-world scenario.

The remainder of the paper is organized as follows. Section II introduces the proposed *IFI* approach. In Section III, we describe the implementation and evaluation of *IFI*, and analyze the results. In Section IV, we discuss a number of issues related to IFI and consider possible solutions. Section V contains some concluding remarks.

II. THE PROPOSED APPROACH: INTERNET FOOTPRINT INVESTIGATION (IFI)

In this section, we introduce the *Internet Footprint Investigation* (IFI) approach for rapid discovery of geographically disconnected areas after a disaster. As mentioned earlier, IFI is comprised of two key components, namely, the Active Network Probing (ANP) module and the Reactive Footprint Search (RFS) module, which we describe in the following subsection. Figure 1 shows the architecture of the IFI approach.

A. Active Network Probing (ANP)

Active Network Probing (ANP) is a *top-down* mechanism that proactively probes 'network landmarks' that may have been damaged by a disaster. It then identifies geographically disconnected areas based on the IP geolocation mapping results. There are two technical challenges in ANP implementation:

- **IP Geolocation Service:**
  The IP Geolocation Service maps IP addresses to geographic coordinates. Although several IP geolocation services are available on the Internet (e.g., IP2Location[1] and

Quova[2]), for three reasons, off-the-shelf IP geolocation services are not suitable for ANP:

1) The accuracy of existing IP geolocation services is not acceptable for disaster response purposes. Later in this study, we show that existing services can only achieve reasonable accuracy in areas with a good network infrastructure. Their accuracy is not good in underdeveloped areas or remote regions, which are more challenging for disaster response teams after being seriously damaged.

2) Existing IP geolocation services only support one-way mapping (i.e., from IP addresses to geographic coordinates). Hence, they are of limited use in disaster response situations, where most queries are in the opposite direction (e.g., searching for the IP addresses of all network terminals within a certain distance of a given location).

3) Existing IP geolocation services are commercial applications that charge subscribers on a per server per year basis. Consequently, they are too expensive for most disaster relief organizations, which operate on a non-profit basis.

To improve the disaster response capability, we propose establishing *local* IP geolocation services, which should be deployed and maintained at all times so that they are available immediately in the event of a disaster. In contrast to commercial global IP geolocation services, *localized services* are designed specifically for disaster response purposes. Based on the design principle, a local IP geolocation service can be implemented easily by combining public Points Of Interest (POI) information (e.g., using Google Places API [8]), Internet naming and addressing models (e.g., using DNS and Whois lookups [11]), and localized heuristics (e.g., using common sense to map domain names to POIs and vice versa). Note that all the required information is freely available on the Internet or from experts with local knowledge. Localized services improve ANP because they are simple and accurate. Moreover, the cost is lower than using existing global IP geolocation services.

- **Network Topology Discovery:**
  Knowledge of the network topology is essential for network diagnosis, recovery, and Quality of Service (QoS) assurance. Conventional network topology discovery methods send out a number of '*traceroute*' commands from a specific set of terminals (i.e., terminals owned or accessible by the user/commander) to all other terminals in the network. Then, the methods use different algorithms to parse the collected end-to-end network path information and construct a graph of the network topology [12]. We consider that existing approaches cannot be applied directly to disaster response scenarios for the following reasons.

  1) Existing network topology discovery approaches

focus on the topology of the core network (i.e., the autonomous system level). They expend less effort on the discovery of edge devices, which are widespread geographically and crucial for detecting disconnected areas after a disaster.

2) Conventional approaches probe a network by using *traceroute*-like tools, which are based on the Internet Control Message Protocol (ICMP). However, because of the increasing security concerns about ICMP attacks [14, 19], ICMP packets are blocked by many network routers and hosts, resulting in the failure of conventional network topology discovery approaches.

3) The advances in load-balancing and cloud computing techniques have facilitated the development of a dynamic DNS-based service discovery approach, which maps a domain name to a set of IP addresses dynamically [21]. As a result, conventional network topology discovery approaches cannot report the topology correctly.

To resolve the problem, we propose that *a 'representative' network topology should be maintained at all times so that it is available immediately in the event of a disaster. To be effective, the topology must be mashed up with IP geolocation information.* Unlike conventional approaches, the proposed solution does not need to cover all devices in the network. Instead, to facilitate disaster responses, the solution must be 'representative' and include 'network landmarks' of geographic diversity. The landmarks are stationary end devices that are positioned approximately evenly in designated areas[3]. After the landmarks have been identified, a network connectivity graph can be constructed by using existing network topology discovery approaches, as well as localized heuristics wherever necessary (e.g., the ICMP blocking problem and the dynamic DNS problem can be resolved/avoided by consulting experts with local knowledge).

By combining the above IP geolocation services and network topology discovery services, the ANP scheme can respond to disasters promptly and detect geographical areas that may be disconnected. Specifically, ANP 1) prioritizes the areas to be searched based on the hazard identification, risk analysis, and impact analysis conducted by the disaster management authority; 2) uses the reverse IP geolocation service to find network landmarks in the specified search area; and 3) probes the identified network landmarks and reports geographically disconnected areas.

B. Reactive Footprint Searching (RFS)

In contrast to ANP, the *Reactive Footprint Searching* (RFS) mechanism is a *bottom-up* approach. It is designed to 1) receive lists of geographically disconnected areas provided by the ANP mechanism; 2) search for Internet footprints in the

[3]We use K-12 schools as network landmarks because their geographic distribution reflects the population density in different areas.

areas on the list; and 3) report the areas in which no Internet footprints were detected after the disaster. Specifically, RFS searches for Internet footprints in the designated geographic areas by exploiting the "*social check-in*" and "*geotagging*" services of emerging Location-Based Social Networks (LBSNs), which allow each user to leave an Internet footprint and share his/her location with friends [17].

ANP is an active scheme that tries to detect geographic areas that are disconnected after a disaster. The RFS scheme, on the other hand, is a passive approach that improves the accuracy of the ANP results by combining them with LBSNs to filter out areas that show signs of Internet activity after a disaster. Thus, by combining ANP and RFS, the Internet Footprint Investigation (IFI) scheme can identify geographically disconnected areas after a disaster in a rapid and accurate manner.

III. IMPLEMENTATION AND EVALUATION

In this section, we describe the implementation of the proof-of-concept IFI system, and evaluate the key components using the devastation caused by Typhoon Morakot as a case study. Morakot, the deadliest typhoon to impact Taiwan in recorded history, began forming to the east of Taiwan on August 2, 2009, and made landfall in the center of the island late on August 7. It produced a record amount of rainfall in the south of the country (peaking at 2,777 mm at Alishan in Chiayi County), and resulted in numerous landslides and flash floods. Approximately 750 people lost their lives in the disaster. The most catastrophic event was a mudslide that engulfed the entire town of Xiaolin, killing 474 inhabitants.

Morakot severely challenged and shocked the government and the disaster response community, as it revealed that *conventional disaster response systems cannot react efficiently to extreme disasters of this kind*. The reason is that such systems rely to a large extent on *situation reports* (i.e., firsthand observations) sent from the frontline to the disaster recovery center by various means, such as phone calls, faxes, emails, and messages on web-based systems. Although the strategy has worked well in the past, it has a fundamental flaw in that it cannot determine if areas are safe or severely damaged if there are no situation reports for the areas.

A. ANP Implementation

In this study, we evaluate the accuracy of existing IP geolocation services, and consider how to improve the services by incorporating local knowledge. We chose K-12 schools as network landmarks because, generally, they reflect the population density in different areas. To prototype the ANP system, we obtained a list of K-12 schools (including their names, postal addresses, and network domain names) in the greater Kaohsiung area (which was severely damaged by Morakot) from the Ministry of Education's website. Then, we used the Google Maps API [6] to convert the postal address of each K-12 school to its GPS coordinates (i.e., the latitude and the longitude), and took the geocoded results as the ground truth for the evaluation.
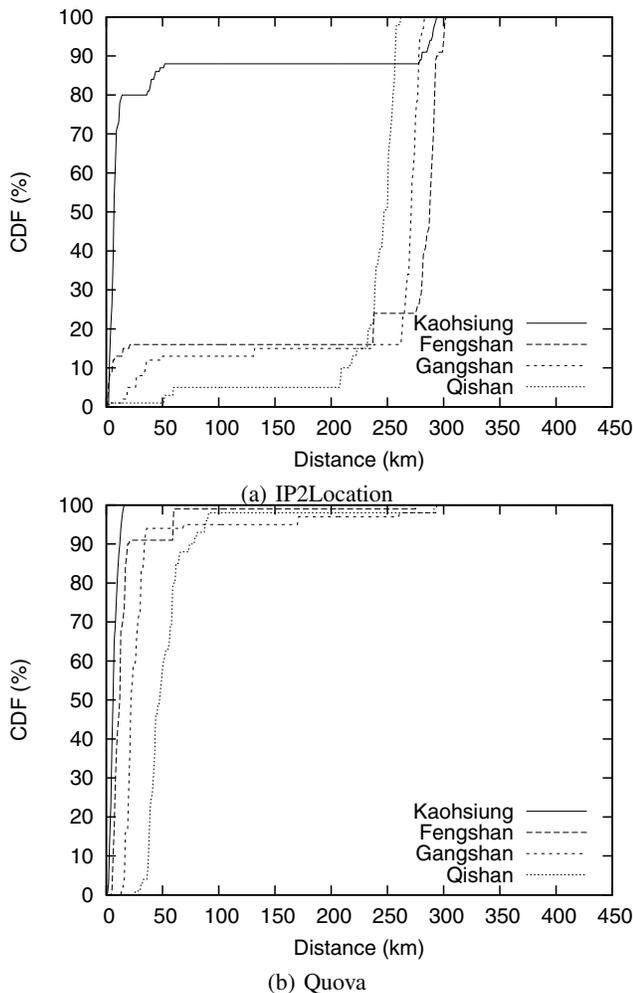
Fig. 2. Evaluation of the accuracy achieved by two off-the-shelf IP geolocation services.

Next, we used the DNS lookup tool, *nslookup*, to convert the domain names of the K-12 schools to their IP addresses (e.g., the domain name of *Ling Ko Elementary School* is *lkp.ks.edu.tw*, and its IP address is *163.16.174.123*). We evaluated the accuracy of two off-the-shelf IP geolocation services, namely, IP2Location and Quova, by calculating the average difference between the IP geolocation results and the ground truth for all K-12 schools.

Figure 2 shows the location distance results for the K-12 schools in the *old Kaohsiung area*, *Fengshan district*, *Gangshan district*, and *Qishan district*, which represent an urban area, a satellite town, a suburb, and a rural area respectively. From the results, we make two observations:

1) *The accuracy of the two IP geolocation services is positively correlated to the urbanization level of the geographic area.* For instance, under the Quova service, 80% of the mapping results are within 9 km in the urban area, 17 km in the satellite town, 31 km in the suburb, and 60 km in the rural area respectively. Intuitively, the higher the level of urbanization, the better will be the Internet infrastructure.

2) *In the evaluation, the Quova service outperformed the IP2Location service in terms of accuracy; however, neither service is suitable for disaster response applications.* Specifically, under the Quova service, 10% of the schools (i.e., network landmarks) in the satellite town have a distance error greater than 20km. The ratio increases to 73% in the suburb, and 100% in the rural area. The results indicate that existing IP geolocation services cannot locate geographically disconnected areas accurately; hence, they are not suitable for the disaster response strategy considered in this paper.

We believe that existing services are accurate for most application scenarios, but they cannot locate "long-tailed" cases, which usually occur in less urbanized areas because they are more vulnerable when a disaster strikes. Thus, a specially tailored IP geolocation service that combines existing services (to provide the baseline knowledge) and local expertise (to verify and correct the location mapping results) is essential.

B. RFS Implementation

In addition to ANP, we prototype the RFS system on the cloud platform hosted by Academia Sinica (URL: http://nrl.iis.sinica.edu.tw/RFS/). The implementation is a mashup of Google Maps [5] and the results of inquiries made via the application programming interfaces (APIs) of modern LBSNs. Specifically, in the prototype system, we have incorporated Facebook Places, Foursquare, GeoTagged Flickr, and GeoTagged Twitter to search for Internet footprints in three LBSNs (i.e., LBSN check-ins and location-based microblogging). The system is highly extensible and can support other LBSNs as long as appropriate APIs are provided (e.g., Google Latitude [4] and Gowalla [7]). Figure 3 shows a screen snapshot of the prototyped system.

To search for LBSN footprints in the RFS system, the user must first pinpoint the center of the search range by clicking the cursor on the map. It is also necessary to specify the search range in kilometers, the types of LBSNs to be searched, and the start time and end time of the search for footprints. Then, based on the user's input, the system mashes up all the matched LBSN footprints on the map, and labels them with the icons of the corresponding LBSN logos.

IV. DISCUSSION

In the following sub-sections, we consider some research and practical issues related to the proposed IFI approach and suggest possible solutions.

A. The selection of representative network landmarks

The proposed ANP scheme selects a set of *network landmarks* that are representative of geographic diversity. We suggest using K-12 schools as network landmarks because, generally, they reflect the population density in different areas. However, in our analysis, we found that some K-12 schools in rural areas have very limited computer resources (including equipment and human resources) to set up their own Internet servers, and they tend to rely on third parties (e.g., companies
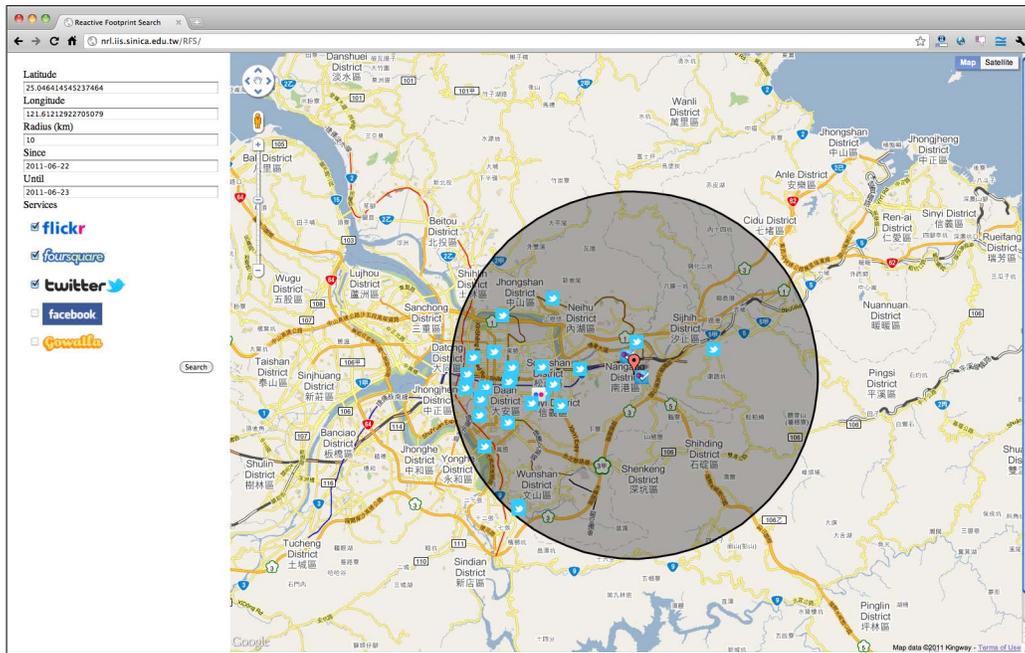
Fig. 3. A screen snapshot of the RFS system implemented in the prototype system.

or schools in urban areas) to host their Internet services (including web and mail servers). As a result, there are no stationary Internet servers in some rural schools, and the IP geolocation service has no means of mapping their IP addresses to the correct GPS coordinates.

There are two possible solutions to the problem: 1) use local knowledge to find other stationary Internet servers in those areas, such as bank ATMs, convenience stores, and police stations; and 2) recruit people in those areas and ask them to ensure their computers are always reachable via the Internet [10]. The first solution is less intrusive, but it is impractical because accessing non-public Internet servers raises security concerns. The second solution is more promising because it is based on a crowdsourcing approach [16]; and it should be practical, as long as good incentives are provided (e.g., monetary rewards, fun/entertainment, and prestige).

B. Finding a list of disconnected areas in the shortest possible time via ANP

After a disaster, it is crucial that disconnected network landmarks are identified as soon as possible. One straightforward solution is to conduct an *exhaustive search* and probe all network landmarks one by one in an arbitrary order. However, this approach is unacceptable because it may result in a very large latency in the worst case (i.e., when the disconnected landmarks are only identified at the end of the search process).

There are two possible solutions to the problem: 1) obtain a risk analysis report from the disaster management authority, and prioritize the probing of network landmarks (i.e., probe the network landmarks in the most vulnerable areas first); and 2) consider the network topology as a graph, and formulate it as an optimization problem to minimize the number of

probes required to cover the whole graph. Note that the risk analysis report includes hazard identification, risk analysis, and impact analysis. Utilizing it requires close collaboration with the disaster management authority, which may be challenging in practice.

C. The limitations of each LBSN API

We have proposed the use of emerging LBSNs to harvest Internet footprints with GPS coordinates, and prototyped a proof-of-concept RFS using Facebook Places, Flickr, Foursquare, and Twitter. There are many free LBSN APIs available on the Internet, but we found that each one has certain limitations. Some APIs, such as Foursquare, do not support *direct* queries that search for LBSN check-ins within a specified spatio-temporal range (for example, *How many check-ins were made within 5 km of City Hall in the last 10 minutes?*). Meanwhile, for security and privacy reasons, some APIs (e.g., Facebook Places and Google Latitude) only report LBSN check-ins made by the requester's friends.

To overcome these limitations and facilitate RFS, it is necessary to design special operations for each LBSN API. For instance, in the first case, we use *two-step* queries, instead of direct queries, to obtain a list of POIs in the specified area, and then search all users' check-ins associated with each POI iteratively. In the second case, we query the number of users that check with a particular POI every minute. The changes in the numbers of users can reflect the arrival and departure of users at each POI. Because of these operations, RFS can obtain information about Internet footprints from the LBSNs after a disaster occurs.

D. The importance of Internet footprints

Although we only consider users' Internet footprints in this study, it is essential that we differentiate the semantics of each footprint's content, which may be a short message, an image, or a video clip. How to gain an accurate understanding of such content in a timely manner is still an open question because it involves the use of natural language processing, data mining, pattern recognition, and crowdsourcing techniques [18, 20]. We defer consideration of this issue to a future work.

V. CONCLUSION

We have proposed an approach called Internet Footprint Investigation (IFI) for the rapid detection of geographically disconnected areas after a natural or man-made disaster. IFI combines a top-down approach (i.e., ANP), which detects the activeness of network hosts by proactively probing the network, and a bottom-up approach (i.e., RFS), which exploits modern LBSNs to search for Internet footprints. We used a proof-of-concept prototype to evaluate the practicability of the proposed system. The results show that existing IP geolocation services are not suitable for disaster response applications, so it is necessary to improve their accuracy by incorporating local knowledge. Moreover, the results demonstrate that visualizing LBSN footprints on a map could help rescuers identify geographically disconnected areas, and speed up disaster response operations. The proposed system is simple and effective, and it can be deployed worldwide.

VI. ACKNOWLEDGEMENT

REFERENCES

[1] Facebook Places. http://www.facebook.com/places/.

[2] Foursquare. http://foursquare.com/.

[3] GeoTagging Flickr. http://www.flickr.com/groups/geotagging/.

[4] Google Latitude. https://www.google.com/latitude/.

[5] Google Maps. http://maps.google.com/.

[6] Google Maps API Family. http://code.google.com/apis/maps/index.html.

[7] Gowalla. http://gowalla.com/.

[8] The Google Places API. http://code.google.com/apis/maps/documentation/places/.

[9] Twitter. http://twitter.com/.

[10] D. P. Anderson. Public Computing: Reconnecting People to Science. In *Conference on Shared Knowledge and the Web*, 2004.

[11] L. Daigle. WHOIS Protocol Specification. IETF RFC 3912, September 2004.

[12] B. Donnet and T. Friedman. Internet topology discovery: a survey. *IEEE Communications Survey and Tutorials*, 9(4):56–69, 2007.

[13] S. Erjongmanee, C. Ji, J. Stokely, and N. Hightower. Large-Scale Inference of Network-Service Disruption upon Natural Disasters. *Lecture Notes in Computer Science*, 5840:134–153, 2010.

[14] F. Gont. ICMP attacks against TCP. Technical report, draft-gont-tcpm-icmp-attacks-05.txt, IETF Internet draft, October 2005.

[15] D. J. Houck, E. Kim, G. P. O'Reilly, D. D. Picklesimer, and H. Uzunalioglu. A network survivability model for critical national infrastructures. *Bell Lab Technical Journal*, 8(4):153–172, 2004.

[16] J. Howe. The Rise of Crowdsourcing. *WIRED Magazine*, 14(6), June 2006.

[17] M. Kirkpatrick. Why We Check In: The Reasons People Use Location-Based Social Networks. ReadWriteWeb.com, June 28 2010. Retrieved May 20, 2011.

[18] B. D. Longueville, R. S. Smith, and G. Luraschi. "OMG, from here, I can see the flames!": a use case of mining location based social networks to acquire spatio-temporal data on forest fires. In *ACM International Workshop on Location Based Social Networks*, 2009.

[19] V. Thing, M. Sloman, and N. Dulay. A Survey of Bots Used for Distributed Denial of Service Attacks. In *IFIP International Information Security Conference*, 2007.

[20] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen. Microblogging during two natural hazards events: what twitter may contribute to situational awareness. In *ACM International Conference on Human Factors in Computing Systems*, 2010.

[21] P. Wendell, J. W. Jiang, M. J. Freedman, and J. Rexford. DONAR: Decentralized Server Selection for Cloud Services. In *ACM SIGCOMM*, 2010.