Inference of Network-Service Disruption upon Natural Disasters

Supaporn Erjongmanee[1], Chuanyi Ji[1], Jere Stokely[2], and Neale Hightower[2]

[1] Georgia Institute of Technology, Atlanta, GA 30332
[2] AT&T Labs, 575 Morosgo Drive, Atlanta, GA 30324
gtg730d@mail.gatech.edu, jic@ece.gatech.edu,
jere.stokely@att.com, nealeh1@bellsouth.net

Abstract. Large-scale natural disasters cause external disturbances to networking infrastructure that lead to large-scale network-service disruption. To understand the impact of natural disasters to networks, it is important to localize and analyze network-service disruption after natural disasters occur.

This work studies an inference of network-service disruption caused by the real natural disaster, Hurricane Katrina. We perform inference using large-scale Internet measurements and human inputs. We use clustering and feature extraction to reduce data dimensionality of sensory measurements and apply semi-supervised learning to jointly use sensory measurements and human inputs for inference.

Our inference shows that after Katrina, approximately 26% of subnets were inferred as unreachable. We find that 57% of unreachable subnets were small subnets at the edges of networks, and 45% of these unreachabilities occurred after the landfall. The majority (73%) of unreachable subnets lasted longer than four weeks showing that Katrina caused extreme damage on networks and a slow recovery.

Network-service disruption is inevitable after large-scale natural disasters occur. Thus, it is crucial to have effective inference techniques for more understanding of network responses and vulnerabilities to natural disasters.

1 Introduction

The Internet is composed of a large number of heterogeneous sub-networks (subnets). Large-scale natural disasters cause damage on networking infrastructure where subnets can become unreachable, resulting in large-scale network-service disruption. Network-service disruption at subnets directly supporting emergency units, e.g., hospitals and government agencies, can cause further damage to the society.

Before a natural disaster occurs, we cannot exactly specify when or where network-service disruption will take place or how long or how large service disruption will be. Hence, to effectively monitor and locate network-service disruption after a natural disaster is crucial. We also need to understand how networks

respond to natural disasters. Thus, an inference of network-service disruption after natural disasters is an important problem to study.

In a general setting, there are devices (hardware or software), e.g., border routers, in the Internet that perform data collection. They can be regarded as "sensors," and collected measurements can be considered as sensory measurements. Sensory measurements are regularly collected on a daily basis from large-scale networks, e.g., with tens of thousands of subnets. Therefore, they are spatially and temporally large-scale. Moreover, most of the measurements do not provide the complete status (failed or operational) of networks. Hence, we consider sensory measurements as unlabeled data.

Besides sensory measurements, human inputs are also available after natural disasters. A human input corresponds to a "network-911-call" from a disaster responder to report network outages. Normally, a small number of human inputs are available. Human inputs are made at a particular time instance but afterward and often delayed. However, these human inputs provide valuable information as they report the underlying status of networks. Thus, human inputs can be considered as labeled data.

The current state of the art in inferring network-service disruption relies solely on sensory measurements. This work introduces a novel use of sensory measurements and human inputs for inference. Specifically, this work uses offline sensory measurements and human inputs from the real and large-scale natural disaster that is Hurricane Katrina. We use unsupervised learning, i.e., clustering and feature extraction, to reduce dimensionality of large-scale sensory measurements. Clustering reduces the spatial dimensionality by 81%, and sensory measurements are temporally extracted down to two features. To jointly use sensory measurements and human inputs for inference, we apply semi-supervised learning to derive the classifier to infer unreachable subnets.

We infer 25.87% of subnets as unreachable and find that the majority (49.21%) of unreachable subnets occurred after the landfall. Moreover, approximately 73% of subnet unreachabilities lasted longer than four weeks. This shows how Katrina critically caused network damage.

The main contributions of this work lie in two aspects. The first is an application of machine learning approaches to a novel networking problem, i.e., inference of network-service disruption upon natural disasters using sensory measurements and human inputs. The second is an analysis of network-service disruption caused by natural disasters.

The paper is organized as followed. Section 2 presents background. Section 3 provides problem formulation. Sections 4 and 5 respectively show the use of unsupervised and semi-supervised learning to sensory measurements and human inputs. Section 6 presents results. Section 7 discusses related works, and Section 8 concludes the paper.

2 Background

To provide the importance of network inference under large-scale external disturbances, we present examples of recent natural disasters that caused network damage. We also provide a review of network monitoring and discuss the heterogeneous data used in this work.

2.1 Natural Disasters and Network Damage

Intensive natural disasters can cause large-scale damage on networking infrastructure. We present four examples of network damage caused by recent natural disasters.

The first example is Taiwan Earthquakes in 2006. The earthquakes broke seven out of nine submarine cables that routed telecommunications services throughout Asia and caused communications loss in at least 14 countries. The impact spread out from Taiwan and China to more distant countries, e.g., India and Pakistan [1, 2].

In 2007, California Wildfires destroyed communications infrastructure and caused broadband, telephone, and Internet outages in local areas [3]. The next example is Hurricane Gustav in September 2008. There were reports of network outages after the hurricane [4].

The number of hurricanes in 2005 broke the record since 1969. There were seven major hurricanes of category three and higher [5]. The most severe hurricane was Hurricane Katrina that caused inconceivably large damage in Louisiana, Mississippi, and Alabama and became the costliest hurricane in U.S. history.

Hurricane Katrina caused large-scale disruption in telecommunication networks. After Katrina, three million telephone lines were out of service. Furthermore, more than 1000 wireless-sites and 38 9-1-1 call centers went down [6]. Network connectivity was critical but it was either unavailable or unstable as experienced by disaster responders [6, 7].

Despite the report of Katrina's effect on telecommunications systems, there were only a few public reports that showed Katrina's impact on communications at the Internet scale [8, 9]. More study is needed on detailed service disruption at subnet level, as subnets directly connect to organizations such as hospitals and government agencies that critically need network communications after disasters.

2.2 Network Monitoring

Network-service disruption can be characterized as unreachability of subnets. This service disruption has been mostly studied for day-to-day network operations [10]. The inference of large-scale network failures is studied only in a few cases, e.g., using simulation [11]. Questions arise pertaining to service disruption caused by a real large-scale natural disaster. How to remotely infer unreachable subnets? What measurements can be used?

Sensory measurements from Internet routing infrastructure can be used for remote monitoring and service disruption inference [12, 13]. The Internet consists

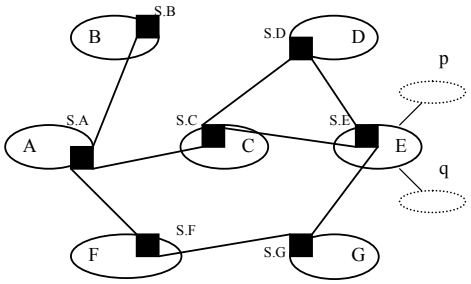


Fig. 1. Example of AS network.

of interconnected autonomous systems (AS), and the routing protocol among ASes is the Border Gateway Protocol (BGP) [14]. Each AS is served by at least one Internet service provider (ISP) and is composed of one or several subnets identified by prefixes (network addresses)[3]. In order to route traffic from one AS to a specific subnet, a BGP router at each AS collects streams of routing messages from peering BGP routers of its neighbor ASes. These messages are called BGP update messages and are regarded as raw Internet sensory measurements in this work. Figure 1 shows the example of AS network where X is an AS, S.X is the BGP router of AS X, and X $\in$ {A,B,...,G}. AS E has two subnets p and q. It also shows that the BGP router S.C collects Internet sensory measurements from peering BGP routers S.A, S.D, and S.E.

There are two types of BGP update messages: BGP withdrawal and BGP announcement. When a subnet becomes unreachable, all BGP routers that can no longer route Internet traffic to this subnet send BGP withdrawals to notify all of their peering routers the unreachability. When a subnet becomes reachable again, there would be new BGP announcements for this subnet. Note that besides network-service disruption, multiple withdrawals followed by new announcements can also be caused by other network events, e.g., a change of routes or routing policies. Hence, a burst of multiple withdrawals followed by new announcements is a symptom rather than a one-to-one mapping of network-service interruption [12, 13].

BGP update messages in this work are collected and stored by Oregon Route Views [15] and are publicly-available. In 2005, Oregon Route Views had about 35 geographically-distributed peering BGP routers. Oregon Route Views provides about 96 files of BGP update messages available per day, and the size of each file is approximately 8 megabytes.

2.3 Heterogeneous Data

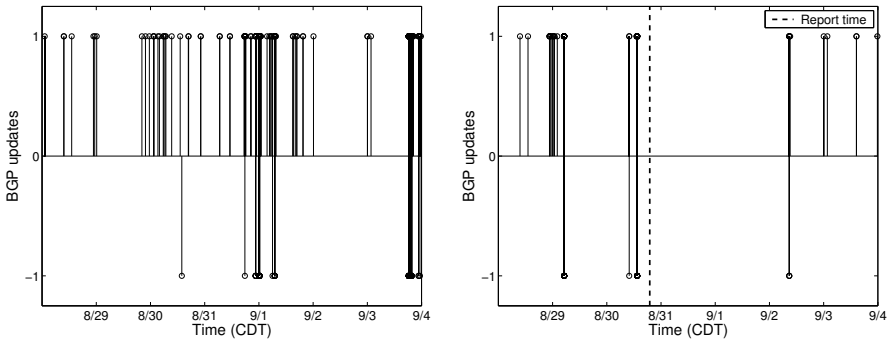
We obtain real sensory measurements and real human inputs from Hurricane Katrina. In particular, we choose BGP update messages to be our sensory measurements. BGP updates provide remote monitoring of service disruption when

[3] We shall use subnet and prefix interchangeably.

local measurements are not directly available because of an evacuation and limited accessibility to disaster area.

We identify geographic locations of subnets from Whois database [16] and select 1009 subnets from 48 ASes in the disaster area. This results in 1009 time-series sensory measurements, one per subnet. Figures 2(a) and 2(b) respectively show a time-series with unknown status and a human-input time-series.

We choose our study duration, i.e., the Katrina interval, to be between August 28 and September 4, 2005. Note that the mandatory evacuation was announced on August 28, 2005, one day prior to the Katrina landfall (August 29, 2005, 6:00 a.m., Central Daylight Time (CDT)), and most of network damage assessment, reported by our collaborating ISP, occurred within the first week after the landfall. In addition, we also select BGP update messages belonging to the same subnets but between August 1-28, 2005 for comparison; this study period is called the pre-Katrina interval.



(a) Time-series with unknown status. (b) Human input time-series.

Fig. 2. Examples of time-series with unknown status and human-input time-series. (1 = BGP announcement, -1 = BGP withdrawal.)

With 1009 subnets and eight-day duration, our sensory measurements are both spatially and temporally large-scale. As a burst of BGP messages is a symptom rather than a one-to-one mapping of service disruption, sensory measurements alone are insufficient to infer unreachability of all subnets.

Human inputs are reports of "this network is down". We collect total 37 human inputs from two sources. The first 28 human inputs are from the online message on NANOG mailing list posted by Todd Underwood from Renesys Corporation [9]. The other nine human inputs are network outage reports from our collaborating ISP. Human inputs provide valuable and mostly accurate information on network outage status because humans most likely to report outage only when they cannot get connected to networks. However, human inputs can be

delayed from the exact time that outage occurs. Thirty-seven human inputs are unlikely sufficient for inferring statuses for the other nearly 1000 subnets. Hence, sensory measurements and human inputs complement each other in inference of service disruption.

3 Problem Formulation

Consider an underlying network with n nodes, where a node corresponds to a subnet. Let $Z_i(t)$ be a binary state of a node i, $Z_i(t) = 1$ if a node i is outage (unreachable); $Z_i(t) = -1$ if a node i is normal (reachable); $1 \leq i \leq n$, and $t \in [0, T]$ is a time duration of interest. The state of a network is a collection of all n states, $Z(t) = \{Z_i(t)\}_{i=1}^{n}$, $t \in [0, T]$, and considered to be unknown. For our case, $n = 1009$, and $T = 8$ days (August 28-September 4, 2005). Service disruption is defined to be the same as unreachability of an individual subnet[4].

Let $X(t) \in R^n$ be an n-dimensional random vector that can be viewed as "response variables" corresponding to an underlying state $Z(t)$. Intuitively, $X(t)$ shows symptoms of $Z(t)$ and is related to both outage and normal states. A set D of m samples is assumed to be available on $X(t)$ and constitutes indirect observations on $Z(t)$. Hence, D is called unlabeled measurements. From Section 2.3, D corresponds to sensory measurements. In general, D is large-scale and insufficient for determining an underlying network state $Z(t)$ unless D is empowered by discriminative information.

Human inputs provide discriminative information. A set of k human inputs are assumed to be available for a fraction of nodes, i.e., $0 \leq k \leq n$. The simplest form of a human input is a symbol that takes binary values, 1 and -1, at time $t^{'}$. Let $t^{'}$ be the time that human reports the unreachability and t be the exact time that a network becomes unreachable. Generally, it is assumed that human reports unreachability of a subnet correctly[5], but a report can be delayed, i.e., $t^{'} > t$. Thus, a human input can be regarded as a direct but delayed observation on one specific nodal state $Z_i(t)$. A set of k human inputs is D_l, where k can be small, i.e., $0 \leq k \ll m$. In this work, we use 24 human inputs (65%) to be training data and the other 13 for validation. Hence, for our case, $k = 24$, $m = n - k = 985$.

Problem: Given a set of unlabeled sensory measurements, D, and a set of human inputs, D_l, how to infer $Z(t)$ for $t \in [0, T]$?

This is an inference problem where dichotomies between outage and normal states of subnets can be learned from sensory measurements and human inputs. Hence, we resort to machine learning approaches outlined below.

– We apply unsupervised learning algorithms that are clustering and feature extraction. Clustering is used to reduce the spatial dimension of time-series

[4] We shall use unreachability, outage, and service disruption interchangeably.

[5] Note that this is a natural assumption as human only reports when a network is outage.

sensory measurements. We then extract the temporal features from time-series measurements to a fewer observations in a low-dimensional feature space and use these features as unlabeled data.
– We apply semi-supervised learning algorithm. We first convert human inputs to human labels by assigning dichotomies to a small number of the temporal features in the low-dimensional feature space. After that, a large set of unlabeled data and a small set of labeled data are combined to infer the statuses of subnets.
– We provide an initial understanding and analyze network-service disruption upon Hurricane Katrina.

4 Unsupervised Learning

We now perform unsupervised learning to extract features from 1009 time-series sensory measurements belonging to our selected 1009 subnets. The first step is to cluster these time-series to reduce the spatial dimension. The second step is to extract temporal features from patterns in the time-series.

4.1 Spatial Clustering

Features can be extracted directly from time-series measurements of each individual subnet. However, 1009 subnets are large-scale, and subnets may have experienced correlated service disruption caused by the same disaster. Therefore, we first reduce the spatial dimension of time-series measurements by grouping similar time-series into clusters.

To measure the similarity of time-series from different subnets, we change the discrete time-series of BGP update messages for a subnet i to be the continuous waveform $r_i(t)$ such that: when BGP announcement arrives at time t, $r_i(t) = 1$; otherwise, for BGP withdrawal, $r_i(t) = -1$. Consider time t, suppose two consecutive BGP updates arrive at time t_1 and t_2, $r_i(t) = r_i(t_1)$ for $t_1 \leq t < t_2$. For a subnet i without BGP update arrival, $r_i(t) = 1$ for all $t \in [0, T]$.

The similarity between $r_i(t)$ and $r_j(t)$ of a subnet i and a subnet j is measured by the average distance $d(r_i(t), r_j(t))$, where $d(r_i(t), r_j(t)) = \frac{1}{T}\int_{t=0}^{T} |r_i(t) - r_j(t)|dt$ for $1 \leq i, j \leq n$. The set of similarity measures, $L = \{d(r_i(t), r_j(t))\}$, where $1 \leq i, j \leq n$, is used as the input for clustering.

We choose the average-linkage hierarchical clustering algorithm since a number of clusters does not need to be pre-chosen. After clustering, we further post-process to obtain a fewer clusters by merging any two clusters if the similarity between them is smaller than a parameter $\hat{T}$. The range of $\hat{T}$ values is varied and tested using the Davies-Bouldin index [17] to determine cluster compactness. The suggested values of $\hat{T}$ are between 45-90 minutes. This can also be interpreted such that two time-series are merged into the same cluster if their similarity measure is smaller than $\hat{T}$.

Clustering spatially reduces 1009 time-series to 191 clusters, resulting in 81% reduction. Although, the simple hierarchical clustering algorithm gives the reasonably good performance, other advanced clustering algorithms can be applied

Table 1. Subsets of subnets. (LA = Louisiana, MS = Mississippi, AL = Alabama)

Subset	1	2	3	4	5	6	7
Geographic location	LA	LA	LA	LA	LA	MS	AL
Number of subnets	166	53	49	115	180	232	214
Percent of reduction	84.3	56.6	67.4	81.8	76.7	81.5	90.7

Table 2. Example of geographic location and time-series pattern belonging to two subnets in the same cluster.

Subnet	Geographic Location	Initial t where $r(t) = -1$	Duration of $r(t) = -1$
1	Hammond, LA	8/30 18:53:42	2 hrs 53 mins
		9/3 23:39:09	17 mins
		9/4 00:25:10	10 mins
2	Hammond, LA	8/30 18:53:42	2 hrs 53 mins
		9/3 23:39:09	17 mins
		9/4 00:10:24	10 mins
		9/4 00:25:10	10 mins

to handle measurements with small similarity measures. The reduction percent are also obtained for smaller subsets by separating 1009 subnets into seven subsets based on the customers of seven local ISPs in the disaster area. Note that subsets 5-7 belong to our collaborative ISP. The reduction percent of each subset is shown in Table 1. In details, each cluster contains the subnets that have a correlation coefficient of $r_i(t)$'s between 0.9986-1.000. Table 2 shows the example of two subnets from the same cluster. This shows that subnets from the same cluster have a highly similar pattern of BGP updates, and the geographic locations belonging to these subnets are similar.

4.2 Temporal Feature Extraction

Because the resulting clusters have correlation coefficient almost one, we randomly choose one representative subnet per cluster and use this much smaller set of 191 representative subnets to extract temporal features of time-series.

As described in Section 2.2, a burst of multiple BGP withdrawals followed by new BGP announcements is a symptom of network-service disruption. Thus, there are two features of this symptom. The first is the number of withdrawal messages that peering BGP routers send in a given time-duration. The second is the length of an unreachable duration between the last withdrawal of a burst and the new announcements after a burst. Thus, a burst of withdrawals followed by new announcements and a succeeding unreachable duration form a BGP-burst pattern.

This duration can be used to infer whether a subnet is unreachable upon a disaster or not. For instance, a BGP-burst pattern with a short unreachable

duration can be caused by temporary service disruption, i.e., a change of routes or routing policies, and a subnet becomes reachable soon after. However, a BGP-burst pattern with a long unreachable duration is mostly caused by major service disruption. But questions arise: how many withdrawals are considered to be a burst, and how long is an unreachable duration of service disruption upon a large-scale disaster? Hence, we formally define features corresponding to a BGP-burst pattern.

Definition: Burst ratio S and unreachable duration T_{fail}

Let v be a time-duration in which a burst of BGP withdrawals is characterized. Let n_v be a number of accumulative BGP withdrawals belonging to a subnet that peering BGP routers send within v time-duration, and n_p be a number of peering BGP routers that could reach this subnet prior to the Katrina interval. Note that a peering BGP router can send more than one BGP withdrawal after a disruption.

The burst ratio is defined as $S = \frac{n_v}{n_p}$, and S measures percent of BGP withdrawals from peering BGP routers. The unreachable duration T_{fail} is defined as the time period between the last BGP withdrawal of a burst in v-duration and the first new BGP announcement after a burst. Therefore, S is the spatial variable indicating how many peering BGP routers fail to reach a subnet. T_{fail} is the temporal variable that characterizes an unreachable duration.

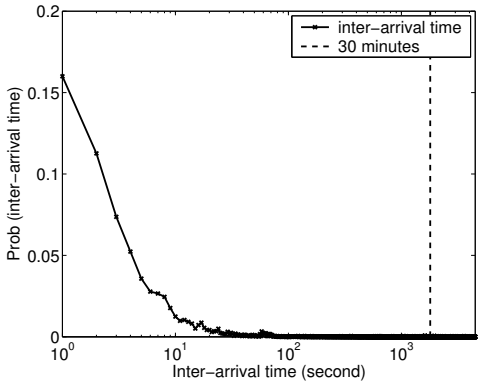
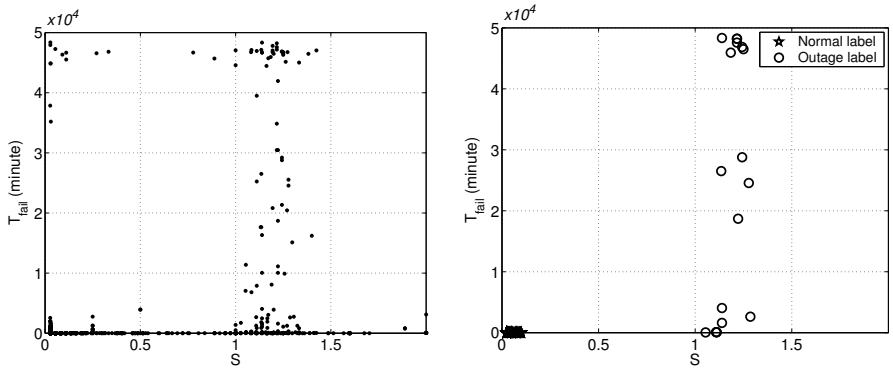


Fig. 3. Empirical distribution of BGP withdrawal inter-arrival time.

The parameter v is a time window such that if the inter-arrival time between two BGP withdrawals is larger than v minutes, these two withdrawals are not considered to be in the same burst. It is reported that, in day-to-day network operations, a burst generally lasts for 3 minutes [18] but can be up to 15 minutes [19]. However, there was no prior result on a burst caused by natural disasters. We derive the empirical distribution of BGP withdrawal inter-arrival time after Katrina as shown in Figure 3. We select $v = 30$ minutes that is large enough not to partition a burst. However, such a large v, a time window may



(a) Unlabeled data. (b) Labeled data.

Fig. 4. S and T_{fail} of unlabeled and labeled data.

include more than one burst. This shows a disadvantage of using a fixed-size time window to locate a burst. To be more precise in locating a burst, instead of monitoring only a number of BGP withdrawals, we can explicitly examine the content of every BGP withdrawal to check subnet reachabilities.

4.3 Feature Statistics

Statistics of S and T_{fail} belonging to time-series measurements are collected from the Katrina interval, and the result is shown in Figures 4(a). We also collect S and T_{fail} statistics from the pre-Katrina interval and find that there are less features with large T_{fail} values in the pre-Katrina than in the Katrina interval. This lack of large T_{fail} in the pre-Katrina interval results in the difficulty to select the appropriate Katrina unreachable duration. Section 5 shows how to use human inputs to derive the threshold of Katrina unreachable duration.

5 Semi-Supervised Learning

We extract 217 (S, T_{fail}) features from time-series measurements belonging to 191 representative subnets; these features can be used as unlabeled data. Note that subnets can have more than one (S, T_{fail}) feature while some subnets do not have (S, T_{fail}) features at all. However, can we use delayed human inputs to identify a BGP-burst pattern and to obtain labeled (S, T_{fail}) features? If so, sensory measurements and human inputs can be jointly used to infer service disruption.

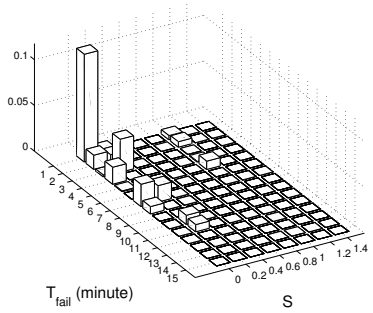


Fig. 5. Empirical probability distribution of S and T_{fail} from the pre-Katrina interval.

5.1 Labeling Human Inputs

When a human input is delayed, there can be more than one BGP-burst pattern in time-series of a subnet prior to a report time. For example, in Figure 2(b), there are three BGP-burst patterns before the report time. Also, among 24 human inputs, 11 human inputs have more than one BGP-burst pattern before a report time. This shows that it can be a complex process to correlate a delayed human report with a BGP-burst pattern. This work selects, for simplicity, the BGP-burst pattern immediately preceding a human report[6]. With 24 human inputs, we have 24 (S, T_{fail}) features that are labeled with "1" (outage).

To classify subnets into two dichotomies, outage and normal, we obtain (S, T_{fail}) features labeled as "-1" (normal) by using the pre-Katrina statistics. The assumption is made such that the majority of (S,T_{fail}) features in the pre-Katrina interval are normal. Figure 5 shows the empirical probability distribution of S and T_{fail} from the pre-Katrina interval. Small values, $S < 0.1$ and $T_{fail} < 3$ minutes, occurred with a large probability. This means that only a small (10%) percent of peering BGP routers send out BGP withdrawals pertaining to a subnet while the rest of peering BGP routers can still reach this subnet. Moreover, with small T_{fail}, this can be interpreted that a subnet quickly becomes reachable after a BGP burst. Hence, subnets with $S < 0.1$ and $T_{fail} < 3$ minutes are considered to be reachable. We extract 460 features of such values and then label these features as normal. Figure 4(b) shows (S, T_{fail}) features labeled as normal and outage.

In total, we have 217 unlabeled features, $\{(S_i, T_{fail_i})\}_{i=1}^{217}$, 24 features labeled as outage $\{(S_i, T_{fail_i}), 1\}_{i=1}^{24}$, 460 features labeled as normal, $\{(S_i, T_{fail_i}), -1\}_{i=1}^{460}$.

5.2 Learning Labeled and Unlabeled Data

Labeled and unlabeled data have been jointly used and studied in prior works as semi-supervised learning. Prior work showed that learning with a small number

[6] That is, humans are prompt in reporting a network outage.

of labeled data along with unlabeled data can reduce classification error from using only unlabeled data [20]. There are three major algorithms used in semi-supervised learning (see [21] and references in there), i.e., generative models, transductive support vector machine, and graph-based methods. The generative models and the graph-based methods require probabilistic models. Thus, these two algorithms are infeasible because the human inputs we obtained are too few to estimate prior probability of outages accurately. Hence, we use the transductive support vector machine (TSVM) by Joachims [22] that only relies on labeled and unlabeled data.

Our goal is to train the (S, T_{fail}) classifier to determine whether subnets are unreachable or not. To avoid over-fitting, we choose the simple semi-supervised learning that applies TSVM to S and to T_{fail} separately. The resulting two one-dimensional linear classifiers (one for S and the other for T_{fail}) are used together as the two-dimensional classifier to infer statuses of subnets.

Let x_i be labeled data and x_j^* be unlabeled data where $1 \leq i \leq k$, and $1 \leq j \leq m$, x_i or x_j^* is a generic variable in the algorithm that corresponds to either S or T_{fail}. Let y_i be the class label for x_i that is assigned according to Section 5.1, y_j^* be an unknown class label for x_j^* that is to be assigned by the classifier, and y_i, $y_j^* \in \{1, -1\}$. Let ξ_i be the so-called slack variable of x_i and ξ_j^* be the slack variable of x_j^*. The use of slack variables allows misclassified samples (see [23]).

Let w be the weight and b be the bias of a linear classifier to be obtained from minimizing

$$\frac{||w||^2}{2} + C\sum_{i=1}^{k}\xi_i + C_-^*\sum_{j:y_j^*=-1}\xi_j^* + C_+^*\sum_{j:y_j^*=+1}\xi_j^* \tag{1}$$

subject to

$$y_i(w{\cdot}x_i + b) \geq 1 - \xi_i, \tag{2}$$

$$y_j^*(w{\cdot}x_j^* + b) \geq 1 - \xi_j^*, \tag{3}$$

$$\xi_i \geq 0,\ \xi_j^* \geq 0 \tag{4}$$

where $\frac{2}{||w||}$ is the margin width of the classifier. $\sum_{i=1}^{k}\xi_i$ and $\sum_{j=1}^{m}\xi_j^*$ are bounds of classification error. C, C_-^* and C_+^* are tradeoff parameters between the margin width and the classification error (see [22] for details).

The outputs of the algorithm are w and b; $\frac{-b}{w}$ is a threshold for either S or T_{fail} to determine the class labels, $\{y_j^*\}_{j=1}^{m}$.

5.3 Experimental Setting and Validation

As unlabeled data is abundant, we separate the unlabeled features into 10 different subsets. Hence, 10 different classifiers are trained, and each training uses one separated subset of 21 (or 22) unlabeled features, all 24 features labeled as outage, and one subset of 30 randomly-chosen features labeled as normal. Other

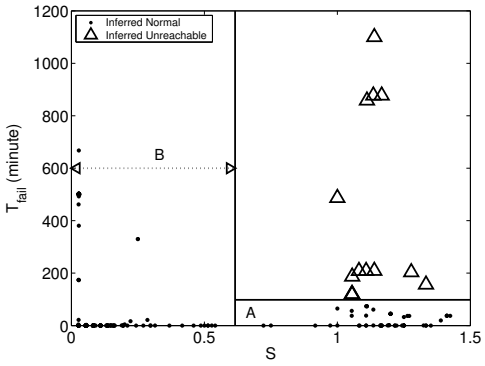


Fig. 6. Scatter plot of inferred S and T_{fail}. (Solid vertical line: $S = S^*$, Solid horizontal line: $T_{fail} = T^*_{fail}$.) Plot only shows values of T_{fail} up to 1200 minutes.

Table 3. Empirical probability distributions of T_{fail} from the pre-Katrina and the Katrina intervals. (For both intervals, probabilities of T_{fail} >75 minutes and $T_{fail} < T^*_{fail}$ are very small.)

T_{fail} (minutes)	0-15	15-25	25-35	35-45	45-55	55-65	65-75
Pre-Katrina	0.6291	0.0488	0.0438	**0.0006**	**0.0006**	0.0938	0.0000
Katrina	0.4238	0.0054	0.0018	**0.0301**	**0.0247**	0.0467	0.0015

parameters used in the TSVM algorithm are initialized such that $C = 0.1$, $C^* = 0.1$, and $num_+ = 0.5$ (these parameters are related to convergence of the TSVM algorithm, and see [22] for details on a choice of parameters).

Let S^* and T^*_{fail} be the thresholds such that if any subnet has features $S > S^*$ and $T_{fail} > T^*_{fail}$, this subnet is inferred as unreachable upon Katrina. Ten thresholds of S resulting from training 10 different classifiers are averaged to yield S^*. We follow the same process to find the value of T^*_{fail}. This results in $S^* = 0.6153$ and $T^*_{fail} = 1$ hour 38 minutes.

We use the rest of 13 human inputs for validation. The result shows that the features belonging to these 13 human inputs have $S > S^*$ and $T_{fail} > T^*_{fail}$ and thus are inferred as unreachable. The inferred unreachable statuses of these human inputs are consistent to the reports that these subnets were outages. Hence, the values of S^* and T^*_{fail} to infer unreachable subnets are valid.

6 Inferred Service Disruption

The thresholds learned are now used to infer service disruption caused by Katrina for the other 985 subnets.

6.1 Statistics of Subnet Statuses

The decision boundaries, $S = S^*$ and $T_{fail} = T^*_{fail}$, partition the feature space into two main regions shown in Figure 6:

- Outage region where $S{>}S^*$ and $T_{fail}{>}T^*_{fail}$ (upper right region in Figure 6). This region contains the inferred unreachable subnets.
- Normal region that has either $S{\leq}S^*$ or $T_{fail}{\leq}T^*_{fail}$. This region contains the inferred reachable subnets.

There are two sub-regions marked as regions A and B in Figure 6. These two sub-regions contain the features that are inferred as normal but show the interesting characteristics of network resilience and responses upon Katrina.

Region A is located where $S > S^*$ and $T{\leq}T^*_{fail}$. The subnets in this region experienced brief T_{fail} and resumed reachability soon after. Table 3 shows the empirical probability distribution of T_{fail} of both pre-Katrina and Katrina intervals. There are significantly more T_{fail} with moderate values, 35-55 minutes, in the Katrina interval while T_{fail} of such values were scarce during the pre-Katrina interval. This shows that Katrina caused networks to respond differently from day-to-day network operations.

Region B is located where $S{\leq}S^*$, and we study some corresponding subnets in this region and find that these subnets maintained reachabilities; hence, there might have been parts of the Internet that were not highly affected by Katrina.

We quantify the percent of subnets in these four regions. The results show that 25.87% of subnets were inferred as unreachable, and 57.09% of these unreachable subnets were at the network edges. There were approximately 42% of subnets from both regions A (12%) and B (30%). With subnets that maintained reachabilities or responded with brief disruption duration, this provides the signs of network resilience upon a large-scale disaster.

6.2 Spatial-Temporal Damage Maps

We now obtain the spatial damage map presented in Figure 7. The spatial map shows network-service disruption of different degree, based on the average disruption duration of the inferred unreachable subnets in each geographic location. The worst service disruption occurred near the coast of Louisiana. Nonetheless, our results show that not all subnets in the entire disaster area suffered from service disruption. This suggests that available network resources in the area could have been utilized if this information was shared among disaster responders.

We use T_{fail} to identify the initial time when service disruption started and the duration of service disruption. The temporal map in Figure 8 and Table 4 show that 49.21% of service disruption occurred after the landfall while only 5.12% occurred on August 28, 2005 (the mandatory evacuation day). There were also substantial service disruption (45.67%) occurred during six hours before the landfall.

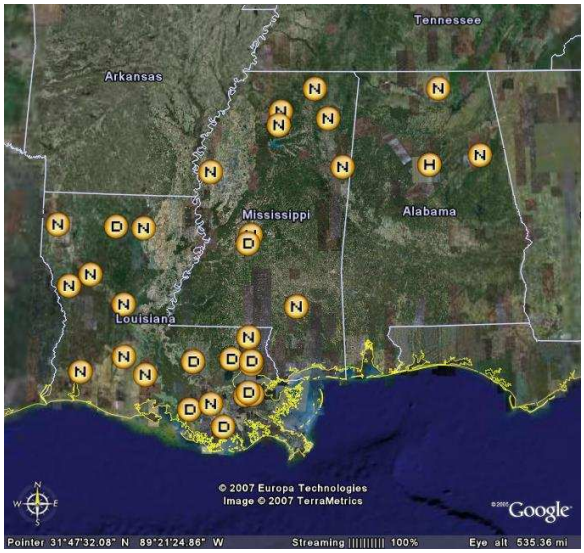


Fig. 7. Impact degree of network-service disruption. (N): $T_{fail} < T^*_{fail}$, (H): $T^*_{fail} < T_{fail} < 24$ hours, and (D): $T_{fail} \geq 24$ hours.

Table 4. Percent of unreachable subnets with different initial time.

Initial time	Before Aug. 29 12.00 a.m.	Between Aug. 29 12:00-6:00 a.m.	Between Aug. 29 6:00 a.m.-11:59 p.m.	Between Aug. 30-31	Between Sept. 1-4
Percent of Unreachable subnets	5.12	45.67	37.01	6.69	5.50

For the unreachable subnets that occurred during six hours before the landfall, our collaborative ISP expected that these subnets were likely to be intentionally withdrawn by network operators, not disrupted by Katrina. On the other hand, there was the report of network connectivity loss because of the lack of power supply on August 29, 2005 at 3:00 a.m. [24]. Thus, it is inconclusive what exactly caused subnet unreachabilities during six hours before the landfall.

Figure 9 shows the initial time and the duration of unreachable subnets located in different cities that were critically damaged by Katrina. These cities are near the coast of Louisiana. The results show that unreachable subnets located in the same city did not necessarily occur in the same time or last with approximately the same duration.

The percent of unreachable subnets with different unreachable duration is shown in Table 5. Approximately 73% of unreachable subnets lasted longer than

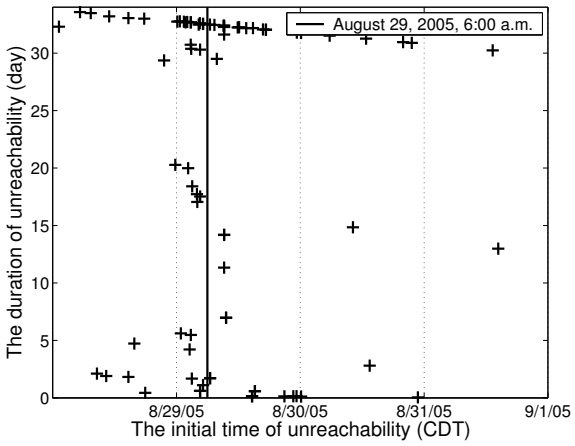


Fig. 8. Initial and duration time of inferred unreachable subnets. ("+" = an inferred unreachability.)

Table 5. Percent of of unreachable subnets with different unreachable duration.

Unreachable duration	Les than 1 day	1-3 days	3-7 days	1-2 weeks	2-4 weeks	Longer than 4 weeks
Percent of Unreachable subnets	7.09	6.30	2.76	3.54	7.48	72.83

four weeks. This means that these subnets did not recover even after four weeks and also shows how much Katrina severely caused extreme damage on networks.

Communications are critical after the occurrence of natural disasters. The application of this work can be used in the future to infer network-service disruption upon other disasters.

7 Related Work

There have been studies of BGP update messages related to the widely affected network-service disruption; the examples of these studies are the September 11 attack in 2001 [25], the Code-Red and the Nimda worm attacks in 2003 [26], and the Middle East cable break in 2008 [27]. In [8], Cowie et al. presented that some disrupted networks after Hurricane Katrina were not recovered after 10 days had passed. Among these studies, little has been done on detailed study of service disruption at subnet level using public available sensory measurements. Furthermore, human data has not been used in these prior works.

There have been studies of machine learning applications to BGP update messages. They were done either for day-to-day network operations or with dif-

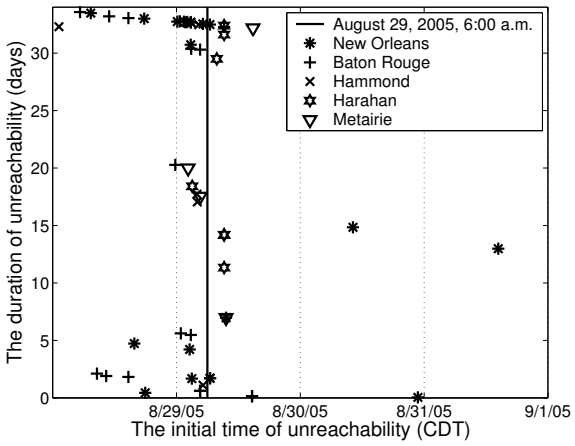


Fig. 9. Initial and duration time of inferred unreachable subnets in cities near the coast of Louisiana.

ferent methods. For example, Andersen et al. applied clustering algorithm to BGP update messages to infer a BGP topology [28] while Chang et al. temporally and spatially clustered ASPATHs to identify the cause of path changes [29]. Xu et al. proposed the algorithm to infer significant BGP events by applying the principal component analysis (PCA) to BGP updates [13].

Human data has been used in supervised learning to infer root causes of network failures using probabilistic models [30–32]. These studies use data from day-to-day operations. Moreover, they rely on complete knowledge of network status and complete underlying inference models. However, natural disasters are rare events. Thus, knowledge of network-service disruption caused by natural disasters is incomplete, and no underlying model of service disruption is available.

Semi-supervised learning has been widely studied [20, 21] and applied in many applications such as text classification [22, 33], remote sensing [34], and image processing [35, 36]. Nonetheless, semi-supervised learning has yet been applied in networking problem in previous studies.

The preliminary version of this work [37] provides an initial study of this problem. This work provides the in-depth study of the problem, the detailed inference results, and the analysis of the results.

8 Conclusion

Natural disasters caused large-scale network-service disruption. This work has presented the inference techniques of network-service disruption after a natural disaster and has analyzed the inference results. We have studied unreachabilities

of subnets from Hurricane Katrina by obtaining real sensory measurements and real human inputs.

We have introduced the joint use of sensory measurements and human inputs to infer network-service disruption. First, we have used clustering and feature extraction to reduce data dimensionality of sensory measurements. The results show that clustering has reduced the spatial dimension of sensory measurements by 81%, and feature extraction has reduced the temporal dimension down to two informative features. Then, we have applied semi-supervised learning to both sensory measurements and human inputs to derive the classifier of unreachable subnets.

We have inferred 25.87% of total subnets as unreachable. We have also presented the spatial and the temporal damage maps that are practical values to disaster response and recovery. Moreover, we have analyzed these Katrina unreachable subnets and found that 49.21% of unreachable subnets occurred after the landfall while substantial unreachabilities (45.67%) took place during six hours before the landfall.

Among all unreachable subnets, 72.83% were unreachable longer than four weeks. This shows that Katrina network-service recovery was extremely slow. Our collaborative ISP have suggested that inaccessibility to the physical area, intensive damage on networking infrastructure, and small demand of network services from customers are the main factors that slowed the recovery.

We have found that 57.09% of unreachable subnets were at the network edges where important emergency organizations such as hospitals and government agencies are mostly located. Because the Internet is highly involved in every aspect of people's daily lives, network customers need and demand reliable network services and network reachabilities. Therefore, it is important to understand how large-scale natural disasters affect networks. This provides future directions for our study to gain further in-depth knowledge on network vulnerability and resilience to natural disasters.

Acknowledgments. The authors would like to thank Nikil Jayant for his help with the project, Cheng Guang, Derrick Dy and Phong Do for help with data processing, Clarence Agbi with software development, Anwar Walid and Zesheng Chen for many helpful discussions. The authors also thank Fred Juang, George Riley, and Magnus Egerstedt for their helpful comments. The support from NSF SGER-Katrina and ECS 0334759 is gratefully acknowledged.

References

1. BBC News. (December 27, 2006). Asia Communications Hit by Quake. from `http://news.bbc.co.uk/2/hi/asia-pacific/6211451.stm`
2. Brown, M., Popescu, A., Underwood T., & Zmijewski, E. (February 2008). Aftershocks from the Taiwan Earthquake: Shaing up Internet Transit in Asia. Paper presented at the NANOG42.
3. California Wildfires Affect Internet Service. from `http://www.satellitefamily.com/news-california-wildfires-affect-internet-service.asp`

4. Zmijewski, E. (September 4, 2008). Gustav: 3 Days Later. from http://www.renesys.com/blog/2008/09/gustav-3-days-later.shtml
5. Beven-II, J. L., Avila, L. A., Blake, E. S., Brown, D. P., Franklin, J. L., Knabb, R. D., et al. (March 2008). Annual Summary-Atlantic Hurricane Season of 2005. Miami, FL: Tropical Prediction Center, NOAAN/NWS/National Hurricane Center.
6. Martin, K. J. (September 29, 2005). Written Statement of Kevin J. Martin, Chairman Federal Communications Commission, at the Hearing on Public Safety Communications from 9/11 to Katrina: Critical Public Policy Lessons, before the Subcommittee on Telecommunications and the Internet: House Committee on Energy and Commerce, U.S. House of Representatives.
7. U.S. House of Representatives. (2005). A Failure of Initiative: Final Report of the Select Bipartisan Committee to Investigate the Preparation for and Response to Hurricane Katrina (Congressional Reports No. H. Rpt. 109-377): The Select Bipartisan Committee to Investigate the Preparation for and Response to Hurricane Katrina.
8. Cowie, J., Popescu, A., & Underwood, T. (2005). Impact of Hurricane Katrina on Internet Infrastructure: Renesys Corporation.
9. Underwood, T. (2005). from http://www.merit.edu/mail.archives/nanog/2005-08/msg00938.html
10. Feamster, N., Andersen, D., Balakrishnan, H., & Kaashoek, M. F. (June, 2003). Measuring the Effects of Internet Path Faults on Reactive Routing. Paper presented at the Proc. of ACM SIGMETRICS on Measurements and Modeling of Computer, San Diego, CA.
11. Sahoo, A., Kant, K., & Mohapatra, P. (2006). Characterization of BGP Recovery Time under Large-Scale Failures. Paper presented at the Proc. of IEEE International Conference on Communications (ICC), Istanbul, Turkey.
12. Feldmann, A., Maennel, O., Mao, Z. M., Bergerm A., & Maggs, B. (2004). Locating Internet Routing Instabilities. ACM SIGCOMM Computer Communication Review, 3(4), 205-218.
13. Xu, K., Chandrashekar, J., & Zhang, Z.-L. (2004). Inferring Major Events from BGP Update Streams (Tech. Rep. No. 04-043). Minneapolis, MN: Department of Computer Science and Engineering, University of Minnesota.
14. Rekhter, Y., Li, T., & Hares, S. (1995). Border Gateway Protocol 4 (RFC 1771).
15. University of Oregon. Route Views Project. from http://archive.routeviews.org.
16. Whois Database. from http://www.arin.net/whois
17. Davies D. L., & Bouldin, D. W. (April, 1979). A Cluster Separation Measure. IEEE Transactions on Pattern Recognition and Machine Intelligence, 1(2), 224-227.
18. Labovitz, C., Malan, G. R., & Jahanian, F. (October, 1998). Internet Routing Instability. IEEE/ACM Transactions on Networking, 6(5), 515-528.
19. Labovitz, C., Ahuja, A., Bose, A., & Jahanian, F. (June, 2001). Delayed Internet Routing Convergence. IEEE/ACM Transactions on Networking, 9(3), 293-306.
20. Castelli, V., & Cover, T. M. (November, 1996). The Relative Value of Labeled and Unlabeld Samples in Pattern Recognition with an Unknown Mixing Parameters. IEEE Transations on Information Theory, 42(6), 2102-2117.
21. Chapelle O., Scholkopf B., & Zien, A. (2006). Semi-Supervised Learning: MIT Press.
22. Joachims, T. (June, 1999). Transductive Inference for Text Classification using Support Vector Machines. Paper presented at the Proc. of International Conference on Machine Learning (ICML), Bred, Slovenia.
23. Burges, C. J. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. Data Mining and Knowledge Discovery, 2(2), 121-167.

24. Warrick, J. (December 10, 2005). Crisis Communications Remain Flawed. Washington Post. from http://www.washingtonpost.com/wp-dyn/content/article/2005/12/09/AR2005120902039.html
25. Committee on the Internet Under Crisis Conditions: Learning from the Impact of September 11. (2003). The Internet Under Crisis Conditions : Learning from September 11. Washington, D.C.: The National Academies Press.
26. Wang, L., Zhao, X., Pei, D., Bush, R., Massey, D., Mankin, A., et al. (November, 2002). Observation and Analysis of BGP Behavior under Stress. Paper presented at the Proc. of ACM SIGCOMM Internet Measurement Workshop on Internet Measurements, Marseille, France.
27. Brown, M., Popescu, A., & Zmijewski, E. (Februay 2008). Middle East Meltdown: A Global BGP Perspective. Paper presented at the Asia Pacific Regional Internet Conference on Operational Technologies (APRICOT), Taipei, Taiwan.
28. Andersen, D., Feamster, N., Bauer, S., & Balaskrishman, H. (November, 2002). Topology Inference from BGP Routing Dynamics. Paper presented at the Proc. of ACM SIGCOMM Internet Measurement Workshop on Internet Measurements, Marseille, France.
29. Chang, D. F., Govindan, R., & Heidemann, J. (November, 2003). The Temporal and Topological Characteristics of BGP Path Changes. Paper presented at the Proc. of IEEE International Conference on Network Protocols (ICNP), Atlanta, GA.
30. Kandula, S., Katabi, D., & Vasseur, J. P. (August 2005). Shrink: A Tool for Failure Diagnosis in IP Networks. Paper presented at the ACM SIGCOMM Workshop on Mining Network Data (MineNet), Philadelphia, PA.
31. Lee, G. J., & Poole, L. (September 2006). Diagnosis of TCP Overlay Connection Failures using Bayesian Networks. Paper presented at the ACM SIGCOMM Workshop on Mining Network Data (MineNet), Pisa, Italy.
32. Bahl, P., Chandra, R., Greenberg, A., Kandula, S., Maltz, D. A., & Zhang, M. (August 2007). Towards Highly Reliable Enterprise Network Services via Inference of Multi-Level Dependencies. Paper presented at the ACM SIGCOMM, Kyoto, Japan.
33. Nigam, K. (2001). Using Unlabeled Data to Improve Text Classification (Doctoral Thesis. No. CMU-CS-01-126). Pittsburgh, PA: School of Computer Science, Carnegie Mellon University.
34. Shahshahani, B., & Landgrebe, D. (September, 1994). The Effect of Unlabeled Samples in Reducing the Small Sample Size Problem and Mitigating the Hughes Phenomenon. IEEE Transactions on Geoscience and Remote Sensing, 32(5), 1087-1095.
35. Li, J., & Chua, C. S. (September, 2003). Transductive Inference for Color-based Particle Filter Tracking. Paper presented at the Proc. of International Conference on Image Processing (ICIP), Barcelona, Spain.
36. Balcan, M.-F., Blum, A., Choi, P. P., Lafferty, J., Pantano, B., Rwebangira, M. R., et al. (August, 2005). Person Identification in Webcam Images: an Application of Semi-Supervised Learning. Paper presented at the International Conference on Machine Learning Workshop on Learning with Partially Classified Training Data, Bonn, Germany.
37. Erjongmanee, S., & Ji, C. (August, 2008). Network Service Disruption upon Natural Disaster: Inference Using Sensory Measurements and Human Inputs. Paper presented at the Proc. of International Workshop Knowledge Discovery from Sensor Data (Sensor-KDD), Las Vegas, NV.