Network Service Disruption upon Natural Disaster: Inference Using Sensory Measurements and Human Inputs

Supaporn Erjongmanee
Georgia Institute of Technology
Atlanta, GA 30332
gtg730d@mail.gatech.edu

Chuanyi Ji
Georgia Institute of Technology
Atlanta, GA 30332
jic@ece.gatech.edu

ABSTRACT

Natural disasters cause large-scale network service interruption which corresponds to unreachability of networks. This problem relates to how networks respond under extreme conditions, and it is neither well-studied nor well-understood. To infer network service disruption, challenges arise, i.e., how to use heterogeneous data that include sensory measurements and human inputs?

This work shows an important role of data mining and machine learning in inferring large-scale network service disruption upon Hurricane Katrina. We present a joint use of large-scale sensory measurements from Internet and a small number of human inputs for effective network inference. Specifically, data mining includes (a) unsupervised learning, i.e., clustering and feature extraction of sensory measurements and (b) semi-supervised learning of both sensory measurements and human inputs.

The approaches are evaluated on network service disruption induced by Hurricane Katrina at subnet level. Our result shows that clustering reduces the spatial dimensionality by 81%, and sensory measurements are temporally extracted down to two features. The subnet statuses inferred by the classifier derived from semi-supervised learning show interesting facts of network resilience and provide the spatial and the temporal maps of network service disruption that can be used to assist disaster response and recovery.

To our understanding, this is the first work of data mining and machine learning using sensory measurements and human inputs for inference of large-scale network service disruption upon a large-scale natural disaster.

Categories and Subject Descriptors

J.2.8 [**Computer Applications**]: Internet Applications; H.2.8 [**Database Management**]: Database Application—*data mining, feature extraction*

1. INTRODUCTION

Internet is composed of a large number of heterogeneous sub-networks (subnets). Subnets can become unreachable after the occurrence of natural disasters, resulting in large-scale network service disruption. To provide the reliability and the reachability of networks, measurements of subnets are collected for performance and service monitoring. In a general setting, devices that perform data collection, e.g., border routers, can be regarded as "sensors," and measurements collected can be considered as sensory measurements.

Besides sensory measurements, this work introduces a novel use of human inputs to aid the inference of network service disruption. Human inputs correspond to human reports on network outages. While sensory measurements can be plenty, human inputs are generally available in a small number. This work shows how to apply data mining and machine learning to sensory measurements and human inputs to perform knowledge discovery of network service disruption upon natural disasters.

1.1 Challenges and Contribution

As analytical models of network services are unavailable, sensory measurements are imperative for inferring service disruption. However, several challenges arise and hinder the advance of this inference application.

The first challenge is that sensory measurements are large-scale. A monitored network generally consists of thousands of subnets, resulting in measurements of a high spatial dimension. For example, in this work, we consider 1009 time-series sensory measurements from 1009 subnets.

Another challenge is complex temporal patterns in sensory measurements. Networks exhibit unknown transient behaviors in response to a disaster, and the corresponding measurements generally have bursty temporal characteristics that are complex for inference.

The third challenge is the heterogeneity of data. In addition to sensory measurements, human inputs provide a distinct type of data. A human input is a "network-911-call" that a disaster responder makes to report network outages. In general, a report is made at a particular time instance but aftermath and delayed. Human inputs are usually available in a small number of subnets. The other data in this work are geographic locations and network addresses of subnets.

The current state of art in inferring service disruption relies solely on sensory measurements. Human inputs, although available, have been excluded from inference. An open issue is how to jointly use a large number of sensory measurements and a small number of human inputs for more effective inference of large-scale service disruption.

This work uses offline sensory measurements and human inputs from a large-scale natural disaster, i.e., Hurricane Katrina. We first provide a problem formulation in the context of machine learning. We then apply unsupervised learning, i.e., clustering and feature extraction to unlabeled data that are time-series sensory measurements. This reduces the spatial dimension of time-series by more than 80% and the temporal dimension down to two features. After that, semi-supervised learning is performed by combining human inputs with unlabeled data for inference. Because human inputs are delayed and thus not in a usable form of labels, we show how human inputs are converted into labeled data by indexing the unreachability pattern in time-series measurements. We then apply semi-supervised learning algorithm to both labeled and unlabeled data to infer service disruption.

The main contributions of this work lie into two aspects. First is the application of data mining and machine learning to a novel networking problem, i.e., inference of large-scale network service disruption upon a natural disaster. Second is the demonstration of the need and the effectiveness of learning from both heterogeneous sensory- and human-data.

The paper is organized as followed. The rest of Section 1 presents background and heterogeneous data. Section 2 provides problem formulation. Sections 3 and 4 respectively show the use of unsupervised and semi-supervised learning to sensory measurements and human inputs. Results are presented in Section 5. Section 6 discusses related works, and Section 7 concludes the paper.

1.2 Hurricane Katrina

Hurricane Katrina was the most severe hurricane that flooded Louisiana, Mississippi, and Alabama in 2005 and caused large-scale disruption in telecommunication networks. Network connectivity was critical but either unavailable or unstable experienced by disaster responders [1, 2]. There were a few public reports that showed the disaster impact to communications at Internet scale [3, 4] but there has not been any study on detailed service disruption at subnet level.

1.3 Network Monitoring

Network service disruption can be characterized as unreachability of subnets. This service disruption has been studied for day-to-day network operations [5] or by using simulation to infer large-scale network failures [6]. But the 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 [7, 8]. Internet consists of interconnected autonomous systems (AS), and the routing protocol among ASes is the Border Gateway Protocol (BGP) [9]. Each AS is served by at least one Internet service provider (ISP) and is composed

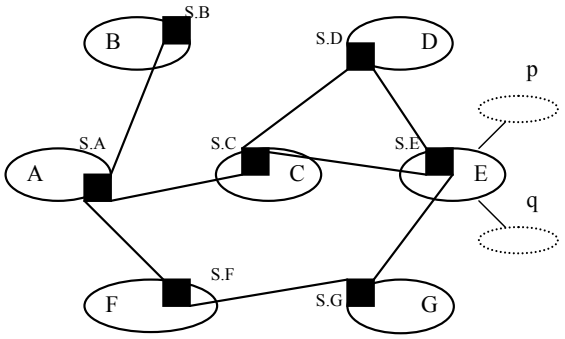


Figure 1: Example of AS network.

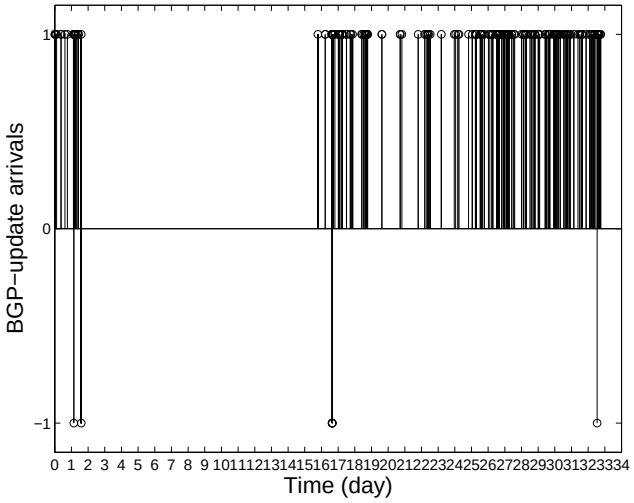


Figure 2: Example of time-series sensory measurements. (1 = BGP announcement, -1 = BGP withdrawal)

of one or several subnets identified by prefixes (network addresses)[1]. 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 prefixes 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

[1]We shall use subnet and prefix interchangeably.

of multiple withdrawals followed by new announcements is a symptom rather than a one-to-one mapping of network service interruption [7, 8].

BGP update messages in this work are collected and stored by Oregon Route Views [10] 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.

1.4 Large-Scale and Heterogeneous Data

We obtain the real sensory measurements and the real human inputs from Hurricane Katrina. In particular, we choose sensory measurements to be BGP update messages that can provide remote monitoring of service disruption when local measurements are not directly available due to the evacuation and the limited accessibility to the disaster area.

Geographic locations are pertinent for selecting subnets in the disaster area to study. We identify geographic locations of subnets from Whois database [11] and select 1009 subnets from 48 ASes in the disaster area. This results in 1009 time-series sensory measurements, one per subnet. Figure 2 shows an example of time-series sensory measurements.

We choose our study duration as 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 belong to the same subnets but between August 1-28, 2005 for comparison; this study period is called the pre-Katrina interval.

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 [4]. The other nine human inputs are network outage reports from customers of our collaborating ISP. Human inputs provide valuable and mostly accurate information on network outage status but 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.

2. 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 node i, $Z_i(t) = 1$ if node i is outage (unreachable); $Z_i(t) = -1$ if 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[2].

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 1.4, 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[3], 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 data mining and 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 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 of network service disruption upon Hurricane Katrina and discuss

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

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

Table 1: List of prefix subsets with (a) Geographic location (LA = Louisiana, MS = Mississippi, AL = Alabama) (b) Number of prefixes, and (c) Reduction percentage

Set	1	2	3	4	5	6	7
(a)	LA	LA	LA	LA	LA	MS	AL
(b)	166	53	49	115	180	232	214
(c)	84.3	56.6	67.4	81.8	76.7	81.5	90.7

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

Sub net	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

the further use of the results and the applications in future network study.

3. UNSUPERVISED LEARNING

We now perform unsupervised learning to extract features from 1009 time-series sensory measurements belong 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.

3.1 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 subnet i and 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 prechosen. 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 [12] 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 $\hat{T}$.

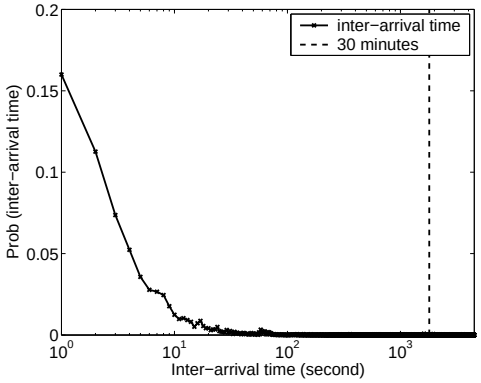


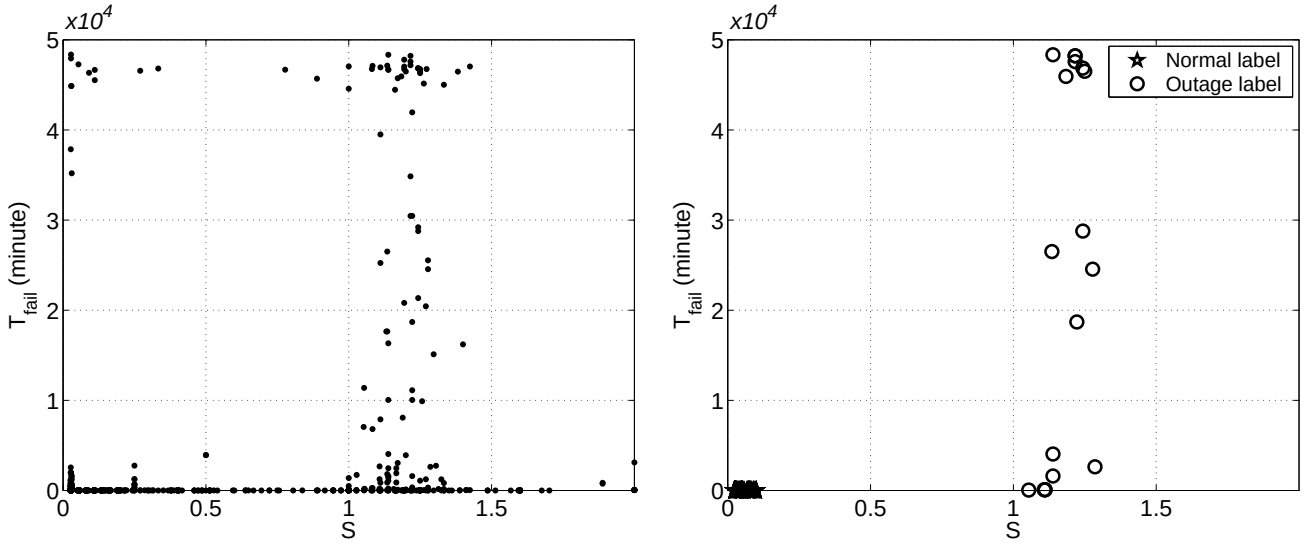
Figure 3: Empirical distribution of BGP withdrawal inter-arrival time.

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 to handle measurements with small similarity measures. The reduction percentages are also obtained for smaller prefix sets by separating 1009 prefixes into seven subsets based on the customers of seven local ISPs in the disaster area, and the reduction percentage of each subset is shown in Table 1. In details, each cluster contains the prefixes 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 belong to these subnets are similar.

3.2 Temporal Feature Extraction

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

As described in Section 1.3, 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 a burst of withdrawals. A burst characterizes a 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. This duration can be used to infer whether a subnet is unreachable upon a disaster or not. Thus, a burst



(a) Unlabeled data (b) Labeled data

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

of withdrawals followed by new announcements and a succeeding unreachable duration form a BGP-burst pattern.

The inference of network events from a BGP-burst pattern has been studied for day-to-day network operations [8]. For instance, a BGP-burst pattern with a short unreachable duration can be caused by a temporary service disruption, i.e., a change of routes or routing policies, and a prefix 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 belong 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 percentage 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.

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 [13] but can be up to 15 minutes [14]. 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 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.

3.3 Feature Statistics

Statistics of S and T_{fail} belong 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 determine the appropriate unreachable duration of Katrina service disruption. Section 4 shows how to use human inputs to derive the threshold to determine the suitable duration of this service disruption.

4. SEMI-SUPERVISED LEARNING

We extract 217 (S, T_{fail}) features from time-series measurements belong 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.

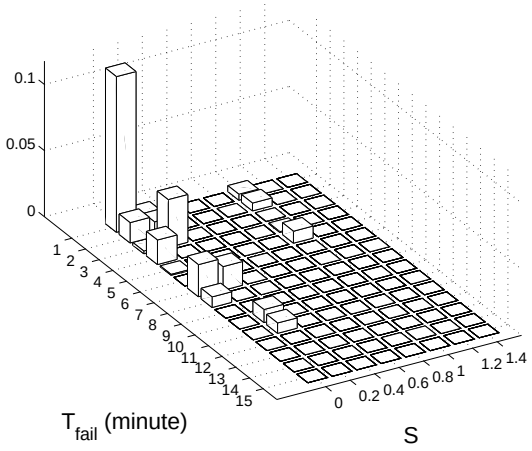


Figure 5: Empirical probability distribution of S and T_{fail} from the pre-Katrina interval.

4.1 Labeling Human Inputs

As a human input can be delayed, there can be more than one network service disruption and thus more than one BGP-burst pattern in the time-series measurements of this human input prior to the time of the human report. This shows that converting a delayed human report to a human label can be a complex process such that which BGP-burst pattern, if there is more than one, exactly corresponds to a service disruption that human reports. This work selects, for simplicity, the BGP-burst pattern immediately preceding a human report[4]. 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%) percentage of peering BGP routers send out BGP withdrawals pertaining to a prefix while the rest of peering BGP routers can still reach this prefix. Moreover, with small T_{fail}, this can be interpreted that a prefix quickly becomes reachable after a BGP burst. Hence, prefixes 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 summary, 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, $\{(S_i, T_{fail_i}), -1\}_{i=1}^{460}$.

4.2 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 of labeled data along with unlabeled data can reduce classification error from using only unlabeled data [15]. There are three major algorithms used in semi-supervised learning (see [16] and references in there), i.e., the generative models, the transductive support vector machine, and the 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 [17] that only relies on labeled and unlabeled data.

Our goal is to train the (S, T_{fail}) classifier to determine whether prefixes are unreachable or not. To avoid overfitting, 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 the 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 as in Section 4.1, 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 [18]).

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

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

subject to

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

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

$$\xi_i \geq 0,\ \xi_j^* \geq 0 \qquad (4)$$

where $\frac{2}{||w||}$ is the margin width of the classifier where $\sum_{i=1}^{k}\xi_i$ 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 [17] 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, $\{y_j^*\}_{j=1}^{m}$.

4.3 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 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 [17] for details on a choice of parameters).

Let S^* and T_{fail}^* be the thresholds such that if any prefix has features $S > S^*$ and $T_{fail} > T_{fail}^*$, this prefix is inferred

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

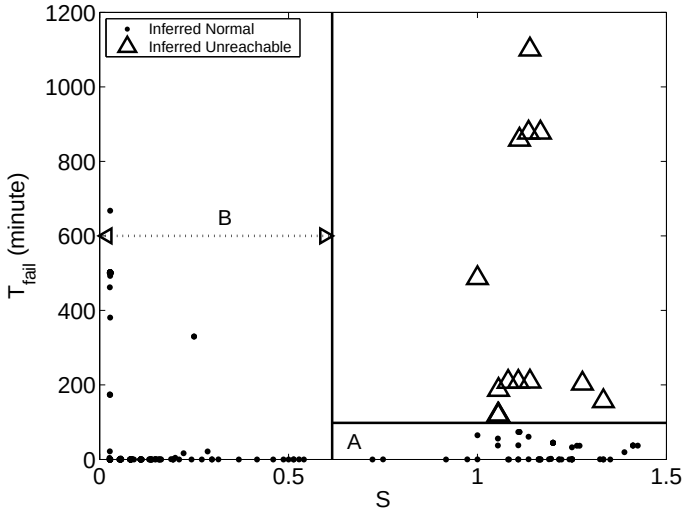


Figure 6: Scatter plot of inferred S and T_{fail}. (Solid vertical line: $S = S^*$, Solid horizontal line: $T_{fail} = T^*_{fail}$.) Plot only shows values of T_{fail} up to 1200 minutes.

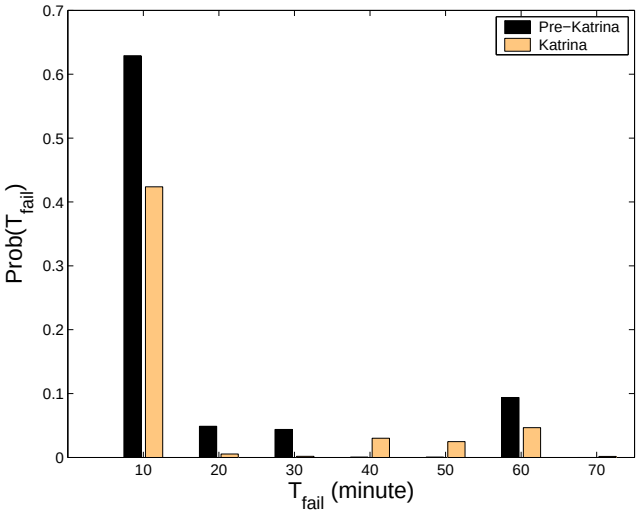


Figure 7: Empirical probability distributions of T_{fail} from the pre-Katrina and the Katrina intervals.

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 belong 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 prefixes are valid.

5. INFERRED SERVICE DISRUPTION

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

5.1 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 S and T_{fail} belong to the inferred unreachable subnets.
- Normal region that has either $S{\leq}S^*$ or $T_{fail}{\leq}T^*_{fail}$. This region contains S and T_{fail} belong to the inferred reachable subnets.

In normal region, 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 Hurricane Katrina.

Region A is located where $S > S^*$ and $T{\leq}T^*_{fail}$. The features in this region correspond to the prefixes that after Katrina, multiple peering BGP routers responded with bursty BGP withdrawals, but these prefixes only experienced brief T_{fail} and resumed reachability soon after. The empirical probability distribution of T_{fail}, where $T_{fail}{\leq}T^*_{fail}$, presented in Figure 7, shows that 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[5]. This shows that Katrina caused network to respond differently from day-to-day network operations.

Region B is located where $S{\leq}S^*$. The features in this region correspond to the prefixes that only a small number of peering BGP routers responded to Katrina. Comparing among S statistics, we find that there are more S with values between $[0.1, 0.5]$ in the Katrina than the pre-Katrina interval. We also study some corresponding prefixes and find that these prefixes maintained the reachability statuses; hence, there might have been parts of Internet that were not highly affected and responded to Katrina.

We quantify the percentages of prefixes in these four regions as shown in Table 3. The results show that 25% of prefixes are inferred as outages, and there are 42% of prefixes from both regions A and B. With prefixes that maintained reachabilities or responded with brief disruption duration, this provides the signs of network resilience and suggests the meaningful direction to investigate the prefixes in regions A and B. This can result into an in-depth understanding of network resilience and responses upon a large-scale disaster.

5.2 Spatial-Temporal Damage Maps

We now obtain the spatial damage map presented in Figure 8. The spatial map shows network service disruption of different degree, based on the average disruption duration of the inferred unreachable prefixes in each geographic location. The worst service disruption occurred at subnets near the coast of Louisiana. Nevertheless, our results show

[5]For both intervals, probabilities of T_{fail} >80 minutes and $T_{fail} < T^*_{fail}$ are very small.

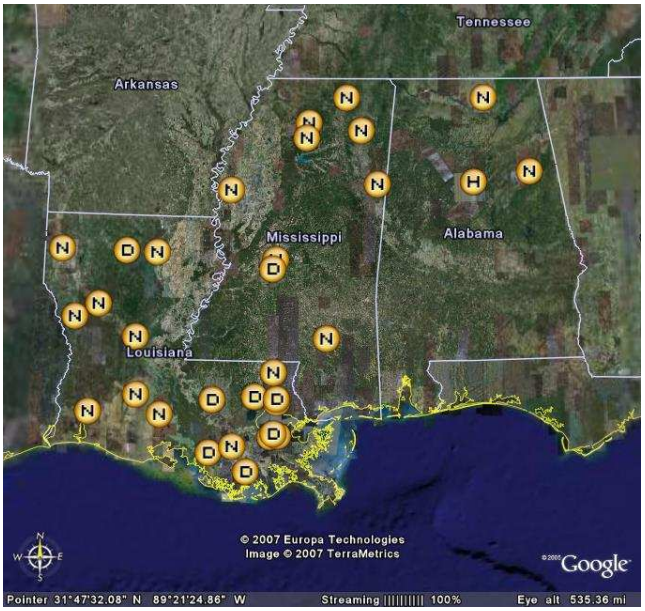


Figure 8: Impact degree of network service disruption. (N): T_{fail}<T^*_{fail}, (H): $T^*_{fail} < T_{fail} <$ 24 hours, and (D): $T_{fail} \geq$ 24 hours.

Table 3: Percentages of prefixes in four regions.

Region	Percentage of prefixes
Normal	75
Outage	25
A	12
B	30

that not all subnets in the entire disaster area suffered from service disruption. This suggests that there were available network resources in the area that 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. This results in the temporal map shown in Figure 9. The temporal map shows 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 substantial service disruption (45.67%) occurred on August 29, 2005 before the landfall, and this service disruption will be discussed in a future study.

Communications are critical after disasters. The application of this work can be used in the future to infer network service disruption upon other disasters. The use of remotely-monitoring sensory measurements gives the advantage of the service disruption inference when the disaster area is physically inaccessible. Moreover, because ISPs do not generally disclose information on unreachability of their service networks, this application can be used to determine service disruption across different ISPs. Furthermore, this application can be further developed to use with online sensory measure-

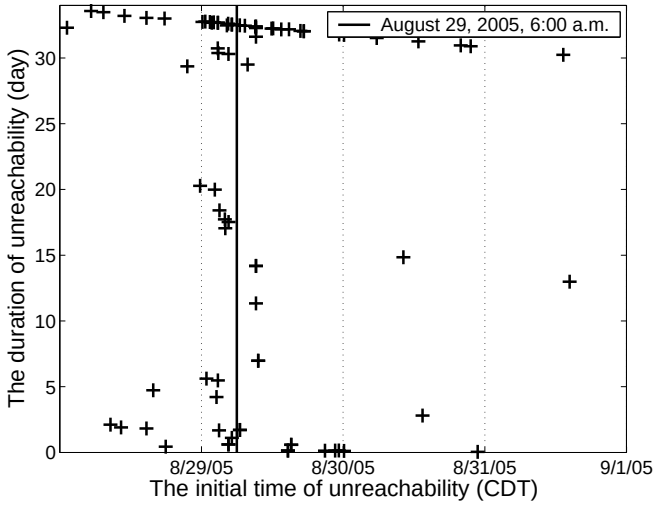


Figure 9: Initial and duration time of inferred network service disruption upon Katrina. ("+" = an inferred service disruption)

ments, and thus, it can infer service disruption in a timely fashion to assist rescue and recovery effort after disasters.

6. 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 [19], and the Code-Red and the Nimda worm attacks in 2003 [20]. In [3], 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 [3, 19]. 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 different methods. For example, Andersen et al. applied the clustering algorithm to BGP update messages to infer a BGP topology [21] while Chang et al. temporally and spatially clustered ASPATHs to identify the cause of path changes [22]. Xu et al. proposed the algorithm to infer significant BGP events by applying the principal component analysis (PCA) to BGP updates [8].

Semi-supervised learning has been widely studied [15, 16] and has been applied in many applications such as text classification [17, 23], remote sensing [24], and image processing [25]. Nonetheless, semi-supervised learning has yet been applied in networking problem in previous studies.

7. CONCLUSION

This work has introduced data mining and machine learning to a new networking application as inference of large-scale network service disruption caused by Hurricane Katrina using sensory measurements and human inputs.

We have found that data mining has played a vital role in learning large-scale and complex sensory measurements in two aspects. First is 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. Second is that semi-supervised learning makes use of a large number of sensory measurements and a small number of human inputs to derive the classifier of network service disruption upon Katrina.

The results show that 25% of subnets are inferred as unreachable. We also present the spatial and the temporal damage maps that are practical values to disaster response and recovery. A large fraction, i.e., 42% of prefixes are found to be either maintained or briefly resumed reachability after Katrina. This suggests the interesting directions for obtaining a deeper understanding of network resilience and responses under a large-scale disaster. These results would have been difficult to obtain without data mining, and this shows the usefulness of our approaches. Our application that is based on publicly available sensory measurements can be used to remotely monitor and localize the reachable network resources after large-scale disasters in the future.

This network application has presented challenges to the existing data mining as well as networking approaches. For example, how to use data mining with a large and complex data set in real time? How to in-depth study network resilience in response to a large-scale disaster? These provide some of future directions for our study and motivate developments of more advanced data mining applications.

8. ACKNOWLEDGEMENTS

The authors would like to thank Jere Stokely and Neale Hightower for their technical help with data and comments, Cheng Guang, Derrick Dy and Phong Do for help with data processing, Anwar Walid and Zesheng Chen for many helpful discussions. This paper is supported by NSF SGER-Katrina and ECS 0334759.

9. REFERENCES

[1] U.S. House of Representatives. A Failure of Initiative: Final Report of the Select Bipartisan Committee to Investigate the Preparation for and Response to Hurricane Katrina. Congressional Reports H. Rpt. 109-377, Washington, D.C., 2005.

[2] K. J. Martin. 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. U.S. House of Representatives, 2005.

[3] J. Cowie, A. Popescu, and T. Underwood. Impact of Hurricane Katrina on Internet Infrastructure. Renesys Corporation, 2005.

[4] T. Underwood. http://www.merit.edu/mail.archives/nanog/2005-08/msg00938.html.

[5] N. Feamster, et al. Measuring the Effects of Internet Path Faults on Reactive Routing. In *Proceedings of ACM SIGMETRICS on Measurements and Modeling of Computer*, pages 126-137, 2003.

[6] A. Sahoo, K. Kant, and P. Mohapatra. Characterization of BGP Recovery Time under Large-Scale Failures. In *Proceedings of IEEE International Conference on Communications,* pages 949-954, 2006.

[7] A. Feldmann, et al. Locating Internet Routing Instabilities. *ACM SIGCOMM Computer Communication Review,* 3(4): 205-218, 2004.

[8] K. Xu, J. Chandrashekar, and Z.-L. Zhang. Inferring Major Events from BGP Update Streams. Technical Report 04-043, University of Minnesota, 2004.

[9] Y. Rekhter, T. Li, and S. Hares. Border Gateway Protocol 4 (RFC 1771).

[10] University of Oregon. Route Views Project. http://archive.routeviews.org.

[11] Whois Database. http://www.arin.net/whois.

[12] D. L. Davies and D. W. Bouldin. A Cluster Separation Measure. *IEEE Transactions on Pattern Recognition and Machine Intelligence,* 1(2): 224-227, 1979.

[13] C. Labovitz, G. R. Malan, and F. Jahanian. Internet Routing Instability. *IEEE/ACM Transactions on Networking,* 6(5): 515-528, 1998.

[14] C. Labovitz, et al. Delayed Internet Routing Convergence. *IEEE/ACM Transactions on Networking,* 9(3): 293 - 306, 2001.

[15] V. Castelli and T. Cover. The Relative Value of Labeled and Unlabeled Samples in Pattern Recognition with an Unknown Mixing Parameter. *IEEE Transactions on Information Theory,* 42(6): 2101-2117, 1996.

[16] O. Chapelle, B. Scholkopf, and A. Zien. *Semi-Supervised Learning*. MIT Press, 2006.

[17] T. Joachims. Transductive Inference for Text Classification using Support Vector Machines. In *Proceedings of International Conference on Machine Learning,* pages 200-209, Bred, Slovenia, 1999.

[18] C. J. C. Burges. A Tutorial on Support Vector Machine for Pattern Recognition. *Data Mining and Knowledge Discovery,* 2(2): 121-167, 1998.

[19] Committee on the Internet under Crisis Conditions: Learning from September 11 National Research Council. *The Internet under Crisis Conditions*. The National Academies Press, 2003.

[20] L. Wang, et al. Observation and Analysis of BGP Behavior under Stress. In *Proceedings of ACM SIGCOMM on Internet Measurement Workshop,* pages 183-195, 2002.

[21] D. Andersen, et al.. Topology Inference from BGP Routing Dynamics. In *Proceedings of ACM SIGCOMM Internet Measurement Workshop,* pages 243-248, Marseille, France, 2002.

[22] D. F. Chang, R. Govindan, and J. Heidemann. The Temporal and Topological Characteristics of BGP Path Changes. In *Proceedings of IEEE International Conference on Network Protocols,* pages 190-199, 2003.

[23] K. Nigam. Using Unlabeled Data to Improve Text Classification. Doctoral Thesis CMU-CS-01-126, Carnegie Mellon University, 2001.

[24] B. Shahshahani and D. Landgrebe. 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, 1994.

[25] J. Li and C. S. Chua. Transductive Inference for Color-based Particle Filter Tracking. In *Proceedings of International Conference Image Processing,* pages III 949-952, 2003.