DRISTI: Distributed Real-Time In-Situ Seismic Tomographic Imaging

Goutham Kamath Wen-Zhan Song Paritosh Ramanan Lei Shi Junjie Yang
Department of Computer Science, Georgia State University, USA
School of Electric and Information Engineering, Shanghai University of Electric Power, China
Email: {gkamath1,pramanan1,lshi1}@student.gsu.edu; wsong@gsu.edu; jyang@suep.edu

***Abstract*—Seismic tomographic imaging is a complex process for imaging the subsurface geological structures. It involves massive data acquisition, signal processing and computing. Traditionally, the voluminous data is logged in each station then manually gathered to a centralized location for post processing. It may take months to see the subsurface image. To see real-time subsurface dynamics, we developed a Distributed Real-Time In-Situ Seismic Tomographic Imaging (DRISTI) system. Unlike the traditional system, it has smart in-situ processing and computing capability that can compute seismic images in real-time. Additionally, each station is equipped with a low-power radio unit that enables collaborative signal processing via a mesh network. We demonstrate the capability of this system by performing experiments with real seismic data-traces as well as outdoor experiments. Evaluation shows that DRISTI can generate the results that matches the ones obtained from a traditional centralized processing [18]. The result from an outdoor hammer shock field experiment using DRISTI system align with the scientific facts and it is able to image the subsurface characteristics.**[1]

***Keywords*—Seismic Monitoring, In-network computing, Seismic Testbed, Distributed Least-Squares, Inverse Problem**

I. Introduction

Seismic activities such as an earthquake, volcano and tsunamis are some of the most feared natural disasters. These have caused incalculable damage to both human life and properties. Today we have advanced earthquake early-warning systems and are implemented in several places such as Japan and California [2]. This system is based on P-wave (primary) arrival time and issues a warning when it detects a P-wave above a certain magnitude. This gives us a few minutes to evacuate before a destructive S-wave (secondary) wave arrives [11]. This system fails if there is a shallow earthquake, that gives a small time difference between P and S wave. To have a better warning system, we need to understand the seismic phenomena by observing the fault movements, the tectonic plate movements and the subsurface velocity in real time. To understand the dynamics of these complex systems scientists often generate a 3D image of the earth's interior as shown in Fig.1(d). This process is called seismic tomographic imaging and is similar to a CT scan.

Seismologists use geophones (vibration sensor) that convert the ground movement (displacement) into a voltage. These are continuously sampled at 50-500 Hz and recorded into a data-logger. Data-logger along with geophones, battery and other sensors are together called seismic station (node). Due to high sampling rate, each station records up to few gigabytes of data each day. Transferring such voluminous data from all the stations through a low power wireless radio is virtually impossible. For this reason, seismologists manually gather all the raw data from each station and perform the tomography calculations at a centralized server. The manual becomes tedious and often take months to generate subsurface images. It also restricts seismologists from deploying large number of stations to obtain high-resolution images. Earthquakes and their timings are an important markers for tomography imaging. From our preliminary analysis of the seismic raw data, we observed a sparse pattern in the detection of the earthquake at the stations. Due to this it would be sufficient to identify these sparse earthquake signature within each station instead of sending the entire raw data. To achieve this, it is necessary for us to perform smart in-situ processing and collaborative signal processing. We also require innovations both in hardware design and algorithms, as the existing system lacks computing capabilities and the algorithms rely on centralized architecture.

A. Background

Seismic signal recorded at each station is shown in Fig.1(a). The sudden spike in the signal represents detection of earthquake wave (P-wave) at that station. The time at which the peak occurs is called the *Arrival Time* (AT). Each stations record different AT for the same earthquake, as the propagation of the earthquake waves through the earth's interior is influenced by its underlying densities which varies with distance. Therefore, AT becomes an important marker to obtain the seismic image. The vertical line in Fig.1(a) shows the AT at four different stations.

Once, the AT is calculated we can use their difference to estimate the earthquake location (x,y,z) along with the time of its origin. This process is called *Event Location* (Fig.1(b)). Now, using the earthquake and the station location we can predict the path of the ray traveled which is called as a *Ray Tracing* (RT) (Fig.1(c)). Discretization of the RT over a regular grid generates a large sparse linear equations and solution to this equation gives us the required seismic image as shown in Fig.1(d). Detailed information regarding seismic tomography can be found in [5], [10].

B. Methodology

In order to balance the system lifetime and network coverage, we adopt a hierarchical network architecture that consists of low-end nodes (referred to as *edge node*) and high-end nodes (referred to as *coordinators*). Each edge nodes run a real-time AT picking algorithm to detect the occurrence of an earthquake. Edge node then sends the timing and the

[1]Our research is partially supported by NSF-CNS-1066391, NSF-CNS-0914371, NSF-CPS-1135814 and NSF-CDI-1125165.

[2]http://www.shakealert.org

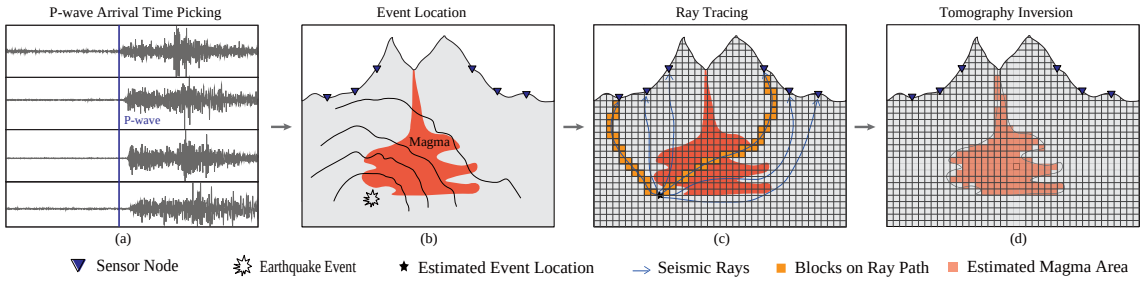


Fig. 1. Illustration of various process in seismic monitoring a) P-wave arrival time picking b) Earthquake hypo-center location c) P-wave ray tracing d) Tomography Inversion

node id to its immediate coordinator node. Due to sparse occurrence of earthquake and also due to smaller size of information sent, the transmission costs drastically reduces from gigabytes to few megabytes per day. At the coordinator node, these AT information from the corresponding edge node is used to estimate earthquake location, perform ray tracing and imaging subsurface anomaly all in a distributed way. Using this architecture, a large quantity of edge nodes can be deployed to monitor an area with longer lifetime, while a small number of coordinators can be used to perform computationally intensive distributed processing to obtain real-time seismic imaging.

C. Contributions

Developing a real-time seismic monitoring system that performs in-network computation and communication in a harsh environment face great challenges [14]. Apart from developing real-time algorithms suitable for processing seismic data, we need stations that can withstand harsh weather condition, easy to deploy and and low cost. In this paper, we develop a Distributed Real-Time In-Situ Seismic Tomographic Imaging (DRISTI) system. We demonstrate the capability of this system by performing experiments with real seismic data-traces as well as an outdoor experiment. We perform indoor testing of DRISTI system using real data trace obtain from Parkfield, CA (San Andreas fault). Evaluation shows that the DRISTI system can generate the results that matches the ones obtained from a traditional centralized processing [18]. We perform an outdoor experiment using our DRISTI system near Lilburn, GA, USA. The result from this experiment align with the scientific facts and it is able to image the subsurface characteristics.

Our key contributions in this paper are, i) Design of a smart seismic station that consists of tiny computational unit and a low power radio in addition to sensing and data-logging features present in traditional seismic station. ii) Lightweight distributed algorithms designed to sample seismic raw data and process them in real time to generate high resolution seismic images. iii) An indoor testbed to evaluate various algorithms and perform end to end system validation using data traces from real world. iv) A GUI at the control center to assist the operator in monitoring the health of the deployed stations. GUI also provides tools to visualize seismic images in real-time which can help operator take decisions quickly.

This paper is organized as follows: In section II we present the related works. Section III describes the system architecture and the hardware components in detail. We describe briefly the algorithms used for seismic tomography in section IV. We perform evaluation and validation of DRISTI system in section V. Few lessons learned from our work is presented in VI. Finally, we conclude our paper in VII.

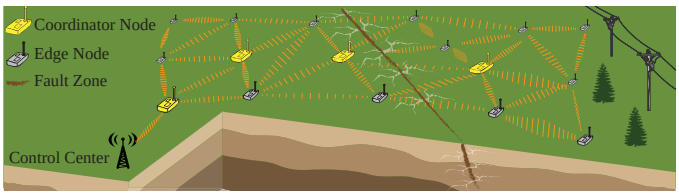


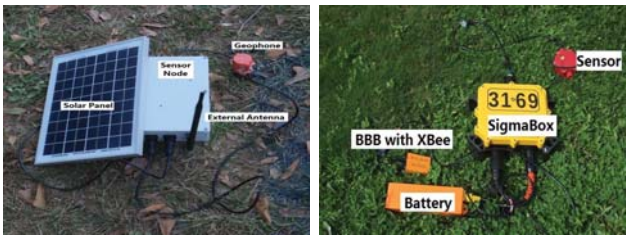
Fig. 2. Illustration of DRISTI system consisting of coordinator node (Yellow) and Edge Node forming a mesh network and deployed to monitor a fault zone.

II. Related Works

The first real deployment of sensor networks for seismic monitoring was carried out on Tungurahua Volcano in Ecuador by team of researcher from Harvard [15]. The team successfully collected three days of acoustic data using four MICA2 nodes. Later in 2005 the same group deployed sixteen Tmote nodes equipped with seismic and acoustic sensors on Raventador volcano in Ecuador which ran for three weeks [16]. These experiments that ran for a few weeks was consider to be major success in sensor network community considering the complexity of monitoring in a harsh environment. Time synchronization was performed through the use of GPS and The Flooding Time Synchronization Protocol (FTSP). FTSP exhibited unexpected behavior and nodes reported an inaccurate time synchronization. Sensor nodes had an inappropriate mark of time in data. It was determined that the accuracy of event detection was only 1% and the reliability of its network and working time were relatively low. The working time of the node was only 69%. Some of the lessons learned from these experiments have motivated us to design DRISTI system making it more stable and robust for long term monitoring.

In 2008, our group carried out a scientific experiment on Mt. St Helens under the OASIS project [14]. The objective was to demonstrate a long term (up-to half year) sustainable WSN in a challenging environment. This design mainly focused on robust network services such as node sustainability, network connectivity, time synchronization and data collection. In this system, a hybrid mode was used for time synchronization, where each node was equipped with a GPS receiver, and the node synchronized by default using this GPS, if the GPS signal disappears, the system switched to FTSP mode. Once this system received the GPS signal with right characteristics it switched back to GPS. This solved the issue faced in [16], therefore, in the DRISTI system we rely on the similar synchronization scheme.

Authors in [9] developed an in-network arrival time picking algorithm for seismic monitoring. They tested their algorithm on 12 TelosB motes loaded with real data traces from 12 nodes of OASIS system. Their implementation involved a lightweight earthquake detection algorithm for detecting a window of size



(a) Custom designed low cost Edge Node designed for sampling and processing seismic signals

(b) Extended SigmaBox (X-Sigma) which acts as a coordinator node; has increased computation capability and reliable communication module.

Fig. 3. Hardware design of the edge and the coordinator seismic station used in DRISTI system.

3 seconds and a picking algorithm to process these samples for accurate picking. The goal of their experiment was to design efficient picking and they did not implement earthquake location or tomography algorithm for end-to-end evaluation. Their architecture unlike ours does not support long term testing due to memory constraints. Their system was tested only in a indoor settings and was not integrated for outdoor field deployment. We extend their work and make the system not only pick the arrival time but also generate tomography image in a real-time using distributed signal processing technique.

Some of the general purpose testbed like Motelab[17] that are widely used to test WSN's are not suitable to carry out seismic monitoring experiments. First, these testbeds are for public use and have time and memory constraints that restricts us from performing long term testing. Second, these testbed architecture are not suitable for rapid deployment and are only limited to an indoor environment. DRISTI design provides flexibility to operate both inside the lab and also to deploy rapidly in an outdoor environment. This hot-swappable architecture is key to this design and helps us to test various features during our project.

III. Overview of DRISTI

In this section, we present a detailed design of our DRISTI system. The system consists of three main components. i) Edge Node ii) Coordinator Node and the iii) Control Center as shown in Fig. 2. The hierarchical design helps us to scale the system, balance the system lifetime and network coverage. The system is able to perform collaborative signal processing via mesh network formed using low power radio.

A. Edge Node

Some of the important design features of the edge nodes are low-cost, lightweight and easy to deploy. In our proposed system, edge nodes processes seismic signals in real-time. These nodes are deployed in large numbers, therefore it needs to be low-cost and easy to deploy. To satisfy these requirements, we design a custom seismic station which acts as an edge node (Fig 3(a)). This design adopts a stackable principles that makes it easy to customize. The design has four types of stacks: computing board, communication module, seismic sensor and the energy module. Each stack can be added or removed independently without effecting others. This design helps us for rapid prototyping and development based on different scenarios. Each station has sensor board that digitizes data at 24 bit at 50-500Hz, and a L1 GPS is embedded for precise timestamps. It consists of smart energy management module to manage the power supply from the solar panel and rechargeable battery. This smart module is designed with a goal of a perpetual lifetime. Each station measures 9.5 x 6.5 x 3.5 and weighs about 5lbs which helps us in rapid deployment. MSP430 micro-controller is used as the low power computational device. This device samples the raw data in real-time and detects the AT using lightweight algorithm discussed later. The earthquake time stamp along with node id is sent to the coordinator node using a wireless module Xbee PRO S3B. This low power radio module is developed by Digi Inc and consists of a programmable Freescale MC9S08QE32 micro-controller, an Analog Devices ADF7023 radio transceiver and an RF power amplifier. The Xbee module is a robust, low power, wireless communication module which operates at 902-928 MHz giving an indoor range of 300m and an outdoor Line-Of-Sight range of 1 km, while guaranteeing a data rate of 200Kbps. This particular Xbee module runs the proprietary Digi mesh protocol which offers mesh network capability making it highly attractive for seismic monitoring applications. The advantage of using a module like Xbee lies in the fact that the key networking features at the MAC layer are abstracted making it much easier to develop applications without bothering much with lower level details. Xbee can be configured to work in five power levels with the default and the highest level providing a transmit power of 24dBm (250mW) drawing a current of 215 mA.

B. Coordinator Node

Coordinator node is the second component of the DRISTI system and is designed to perform data intensive calculation and collaborative estimation. Some of the salient features of coordinator nodes are i) Greater processing capability ii) Extended Storage feature and iii) Reliable communication module. Since the edge node does not have these features we cannot use them as the coordinator nodes. Note that the coordinator node also samples seismic signal similar to edge node but have higher processing capabilities. We design coordinator node using existing state of the art seismic acquisition unit called SigmaBox. SigmaBox consists of 3 channel signal acquisition capability (x,y,z), 24 bit Digital to Analog converter unit, GPS module, storage (64 GB), ethernet interface and rugged case with watertight connectors. Original sigmabox unit do not have the computational and communication capabilities. We embed BeagleBone Black (BBB)a tiny computer and Xbee Pro S3B into an existing sigmabox unit. We call this smart seismic acquisition unit *XSigma*. BBB consists of ARM CPU at 1Ghz with 512MB RAM and supports 32GB memory card to perform in-network processing of seismic data. These tiny units are low power and can be customized to interact with various other peripheral devices using its general purpose I/O pins. These unit also run different flavor of Linux such as Ubuntu or Angstrom.

C. Control Center

The idea behind the control center is to act as a congregation point which collects all the data relayed by the different nodes sprawled across the network and present it for decision making and further analysis. Some of the features of the control centers are: i) It comprises of communication module allowing for easy interoperability and compatibility with the rest of the network radios. ii) It acts as a center for scientists or operator to monitor data and to take corresponding action. iii) It features graphical user interface (GUI) that helps to monitor the health of a mesh network and also to visualize

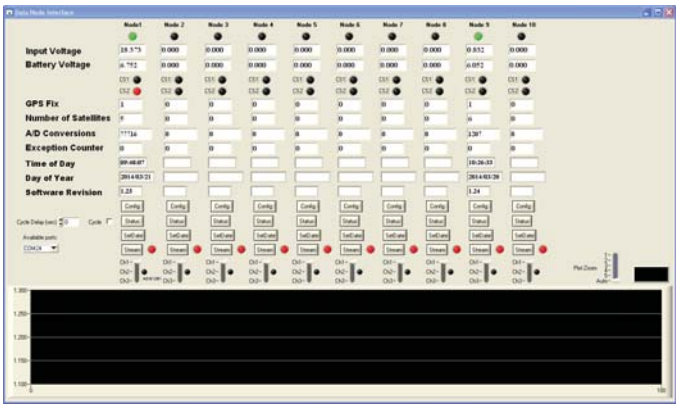


Fig. 4. Graphical User Interface at the control center to monitor the mesh network and also to analyze seismic activity in real-time.

various seismic related data such as the earthquake location and the tomography imaging. We provide GUI (Fig 4) to visualize the seismic data and take actions in real time. GUI also allows operator to control and monitor the health of DRISTI system. Through this GUI, users can also download raw data samples by specifying certain time period. This feature can be very useful for seismologist to manually visualize the historic data remotely and avoid going to the deployment site which could prove to be hazardous.

IV. Algorithm Implementation

DRISTI is designed as a generic seismic monitoring system and it consists of various hardware architecture stack. We can run various seismic tomography monitoring algorithm such as eikonal tomography [8], double difference tomography [19] and first arrival time tomography [6] using DRISTI systems framework. In this paper we use principles of the first arrival time tomography to demonstrate the working of DRISTI system. Traditional arrival time algorithms are centralized and we have developed distributed algorithms for tomographic imaging [3], [4], [13]. In this section we will briefly describe the algorithm and its implementation details for the completeness.

Arrival time tomography has three important steps:i) arrival time picking ii) Earthquake location iii) ray tracing and tomography inversion. In DRISTI system arrival time picking algorithm runs both on the edge and the coordinator nodes. Earthquake location algorithm is computationally more intensive and is run on coordinator nodes. It takes arrival time as the input from the corresponding edge node and returns the location and the origin time of the earthquake (x,y,z,t). Once the location is estimated, we perform ray tracing followed by tomography inversion within the coordinator node. Detailed explanation regarding this process is provided in [4], [5]. We will now provide a detailed descriptions of these algorithms.

A. Arrival Time Picking

Traditionally, the arrival time picking was done at the control center using all the gathered data. Therefore, existing centralized algorithms are not suitable to run on low power stations. Seismic sensors record continuous vibration generated from earth's interior. When a sudden vibration (e.g earthquake) occurs, there will be a change in an amplitude and the frequency of the signal as shown in Fig. 5. Our goal is to detect this change and accurately pick the time when the first change occurs. To process these high fidelity signals we need accurate real-time algorithms. We achieve this by designing

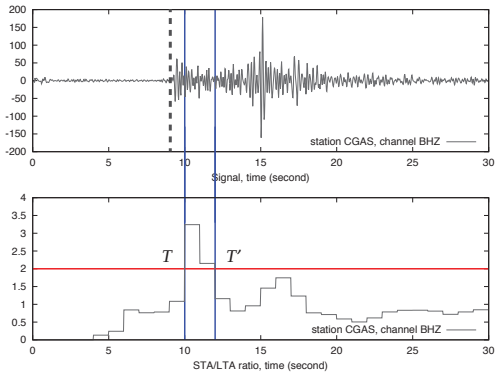


Fig. 5. Arrival time picking algorithm running on DRISTI using real-data set. Red line indicates the STA/LTA threshold and when the threshold is above certain value a window of 3 second is extracted (Blue line). MLE is run on this 3 second window to calculate exact time. Trace details: Station: CGAS, Date/Time: 17:39:20 to 17:39:50 Feb 7, 2002, Location: Parkfield, CA, USA

a two tier search scheme i) Coarse feature extraction using STA/LTA (short term average/long term average) [11]. This phase obtains a rough window of probable earthquake occurrence. ii) Fine grained picking using Maximum Likelihood Estimation (MLE) [] for accurate detection of p-wave time.

STA/LTA algorithm is fast and memory efficient and is based on the principle of RSAM (Realtime Seismic Amplitude Measurement), which is widely used in seismology. To describe this briefly, let us assume the sampling rate of the signal is m (samples per second), let $\{x_t, \cdots, x_{t+m-1}\}$ and $\{x_{t-m}, \cdots, x_{t-1}\}$ be the samples in the i-th and $(i-1)$-th second respectively, then $e_{i-1} = \frac{\sum_{j=t-m}^{t-1} x_j}{m}$ is the average of the $(i-1)$-th second. The i-th second RSAM r_i is calculated as, $r_i = \frac{\sum_{j=t}^{t+m-1}(x_j - e_{i-1})}{m}$. Once a rough window of probable earthquake is obtained, we detect the exact time the sensor received the earthquake signal using maximum likelihood estimation. This is a classic change detection algorithm and here we use variance of post-change and pre-change. Prior to any earthquake, sensor typically records signal consisting mainly of noise. This noise is effected by temperature, manufacture noise etc and is constant for given station. We calculate this prior to the deployment and set as pre-change parameter σ_1^2 in our algorithm. We denote σ_2^2 as the variance of post change samples and our goal is to calculate this unique value. Based on MLE theory, the variance of the seismic signals are related according to Eq 1. We can use classic MLE solvers to solve the Eq. 2. In our implementation, the STAW and LTAW are set to 1 and 4 seconds respectively; we use three seconds signal window size.

$$\frac{\partial \sum_{i=k+1}^{t} \left[\frac{1}{2} \ln \sigma_1^2 - \frac{1}{2} \ln \sigma_2^2 - \frac{x_i^2}{2} \left(\frac{1}{\sigma_2^2} - \frac{1}{\sigma_1^2} \right) \right]}{\partial \sigma_2^2} = 0 \qquad (1)$$

from this σ_2^2 can now be expressed as,

$$\sigma_2^2 = \frac{\sum_{i=k+1}^{t} x_i^2}{t-k} \qquad (2)$$

B. Earthquake Location

In this step our objective is to estimate the earthquake location based on arrival time. Based on DRISTI architecture, each arrival time calculated at the station is transmitted to its respective coordinator node using a low power radio (Xbee). At

the coordinator node, these arrival times are processed in real-time in order to estimate an earthquake location. Estimating an earthquake location plays a key role in understanding the dynamics of the earth's interior. Since, earthquakes are used as the source in the tomography inversion, estimating its location helps us to derive ray propagation map (ray tracing). It must be noted that localization of a particular earthquake at the coordinator is only possible if the arrival time picked by its edge node are relatively close. Various factors like, instrument noise, signal quality etc can cause false alarming and such pickings are often discarded for earthquake location. A minimum of four arrival time with a difference of few millisecond is required for accurate estimation of a particular earthquake location. Therefore, each coordinator nodes should have atleast three edge nodes in order to estimate an earthquake location.

At the coordinator node we use a modified version of the standard Geiger's method [2] for location estimation. For this method, we require a 1D initial reference model of the subsurface P-wave velocity of that region. To give an overview, let the arrival time picked at the edge node i for earthquake e is given by α_{Oie}. We call this arrival time as *observed time* as this is obtained from the actual observation at node i. The predicted arrival time α_{Pie} for node i can be obtained using 1D velocity model. Now, α_{Pie} is a function of x,y,z and q where {x,y,z} are the coordinates of the earthquake location and q is earthquake origin time. Our goal is to minimize the residual between observed and predicted time i.e. $t_{ie} = \alpha_{Oie} - \alpha_{Pie}$ = $\alpha_{Oie} - f(x, y, z, q)$. In order to solve this optimization problem on DRISTI we use iterative methods such as LSQR [12] or Bayesian ART [7].

C. Distributed Tomography Inversion Algorithm

Tomography inversion is the final and the most important step in real time seismic processing. This step provides us a 3D layout of the earth's interior. Interpretation of such images can provide us with a substantial information regarding the physical dynamics of the underlying region. Inversion process is divided into two phases. The first phase is called *Ray Tracing*. The idea behind this process is to estimate the path traveled by an earthquake wave from the origin to the corresponding node. This method is local to each coordinator node and requires no communication with other stations. We use ray bending method presented in [1] to perform ray tracing in each coordinator node. The traced ray paths, in turn, is used to probe a 3D velocity structure of the subsurface through tomographic inversion. This leads to the formulation of linear least squares problem where each row represents the path information of the wave traveling through 3D subsurface. Since each ray path intersects the model at only a relatively few number of voxels, the design matrix, A, is very sparse. If the unknown velocity model is represented by s and the travel time residual by t, the inversion aims to find the least squares solution $s = argmin_s \|t - As\|_2^2$.

In the DRISTI framework, each coordinator node generates partial design matrix A, due to the fact that each coordinator node receives only certain ray information. Therefore, at a coordinator node i we now only have A_i, t_i. Now, the key challenge lies in solving

$$As = t \tag{3}$$

where,

$$\mathbb{A} = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_P \end{pmatrix}; = \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_P \end{pmatrix}; A_i \in \mathbb{R}^{m_i \times n}; t_i \in \mathbb{R}^{m_i}$$

and $s \in \mathbb{R}^n$ is the unknown vector. We assume that, $\{A_i, t_i\}$ forms subsystem in each of the P coordinator node. All the traditional inversion methods assume a complete linear system at the centralized unit. This implies the transfer of all the partial matrix from coordinator node to a central unit. This aggregation imposes a bottleneck at the central unit and becomes virtually impossible because of the size of the design matrix. In order to avoid this we have to designed a robust decentralized inversion algorithm suitable to run on DRISTI system. The method should be iterative in nature and should not require the full design of matrix to be in memory at one time and can incorporate new information (ray paths), on the fly. We proposed an algorithm in [3] and has been tested for convergence and correctness. This algorithm is based on a well known iterative method called Bayesian ART (BART) [7] and is developed specifically to solve tomography inversion problem. Algorithm 1 provides the pseudo-code of Distributed Bayesian ART. Details regarding its convergence etc can be found in [3].

Algorithm 1 Distributed-Bayesian ART Algorithm

set $x_\ell^0 \in \mathbb{R}^n$ to an arbitrary value $\forall \ell \in \mathcal{V}$.
for $k \leftarrow 0$ until convergence or max iteration **do**
  for each $1 \leq l \leq P$ in parallel **do**
    $y_\ell = \text{BART}(A^\ell, b^\ell, x_\ell^k, \rho_\ell)$
    $\tilde{y}_{(\ell)}^0 \leftarrow \quad y_\ell$
  end
  for $t \leftarrow 0$ until convergence or max iteration **do**
    Node $i \in \mathcal{V}$ contacts $j \in \mathcal{N}_i$ and updates
    $\tilde{y}_{(j)}^{t+1} = \tilde{y}_{(i)}^{t+1} = \frac{\tilde{y}_{(j)}^t}{2} + \frac{\tilde{y}_{(i)}^t}{2}$
  end
  $x_{(\ell)}^{(k+1)} = \tilde{y}_{(\ell)}^t \quad 1 \leq l \leq P$
end

V. Evaluation and Validation

In this section, we evaluate DRISTI system with two main experimental setup. First, we validate the robustness of both the hardwares and algorithms by performing a long term test in an indoor environment. In this setup, we supply the DRISTI system with the real data traces and allow it to run for several days. We also perform load/stress test by simulating bursts of earthquakes in a very short duration. This allows us to test the robustness of the hardware and the algorithm in case of unusual circumstances. We report some of the advantages of performing the load/stress in identifying some crucial bug related to memory issue in our communication module. Finally, we perform extensive outdoor test of the DRISTI system at the location near Lilburn, GA, USA. This test was to evaluate the sensor sensitivity, edge and coordinator node data-logging capacity, algorithm behavior and communication protocols in a real environment.

A. Indoor Experimental Setup

For an indoor evaluation of DRISTI, we use eight edge nodes, two coordinator nodes and a control center. Each coordinator is responsible to receive message from four edge

nodes. For an indoor setup since we cannot create real seismic vibration that can be sensed via the geophones, we supply data traces from real testing site. The edge nodes are connected to a network switch which is in turn connected to a FTP server. This server is responsible to supply the file to all the edge nodes thereby simulating the occurrence of seismic data generation. The dataset we used for this experiment is obtained from Parkfield, CA between Oct 2, 2001 through Oct 10, 2002 [18] [3]. Parkfield, CA is a region located in coastal mountain range of Central California and is considered by the geologist as "Earthquake capital of the world". Researchers have installed large number of seismic station to study the fault tectonics (colliding of two plates). The dataset we used are from 32 stations deployed on the fault and in this experiment these data are mapped to our eight edge nodes. From this indoor setup, we can process single day data in less than a minute there by accelerating the whole experiment duration (one year data) to less than few days. This indoor setup resembles to that of real world seismic monitoring. Next we validate each steps in detail.

1) Arrival Time Picking

Fig. 6(b) shows the difference in the time picked between our algorithm with that of the state of the art picks from the data set [18]. X-axis is the picking difference between two methods and the y-axis is the number of pickings in that difference range. There are total 4478 pickings generated by the algorithm, which has a matching in the manual picking data for the Parkfield dataset. About 91% pickings from our algorithm are within 0.2 seconds of manual pickings. The mean and the standard deviation of the difference are 0.043 and 0.23. As is evident from Fig. 6(a), our algorithm validates the scientifically proven fact that all events occur on the fault line. These results confirm that the picking algorithm has sufficient accuracy and is robust as it has successfully processed one year data.

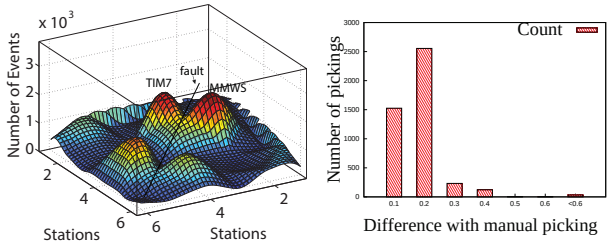
2) Earthquake Location Accuracy

As mentioned earlier, localization of events can be done only when the arrival time of a particular event obtained by various stations are relatively close. Using the arrival time result sent by edge nodes, Geiger's method was able to pick 433 events of which 290 corresponded to the events detected in [18]. Increased event identification could be perhaps due to small variation in the sensitivity parameter or 1D velocity model used. The mean value and standard deviation of the difference between the positions of our algorithm and [18] are (0.33, 0.26) km respect to X, (0.46, 0.37) km respect to Y and (0.39, 0.40) km respect to Z. Fig. 6(c) shows the event location result compared with the [18]. A mean error less than 0.5 km is sufficient enough to obtain satisfactory imaging result and our system is capable of calculating earthquake location with this accuracy.

3) Tomography Inversion

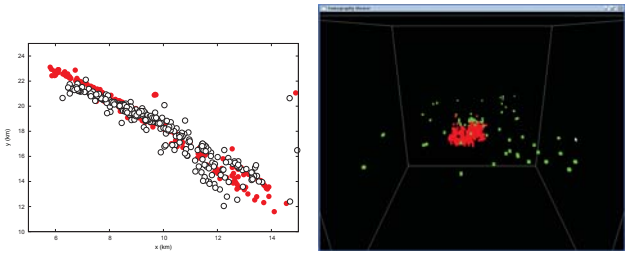
Once the location is estimated we use ray tracing methods to form the linear system. We use our distributed tomography algorithm in DRISTI to generate real-time tomography result of the parkfield region. Fig. 7 shows the P-wave velocity model (tomography image) of the san-andreas fault region. The tomography in the first row is obtained using centralized algorithm as presented in [18]. The second row is the result obtained from DRISTI system after running Distributed Bayesian ART with with $\rho = 0.25$ and $\lambda = 1$. From this experiment we were able to obtain the main feature of the fault and is comparable to results in [18].

[3]http://www.iris.edu/



(a) Number of earthquake detected by each station. We can observe large number of events detected by stations along the fault

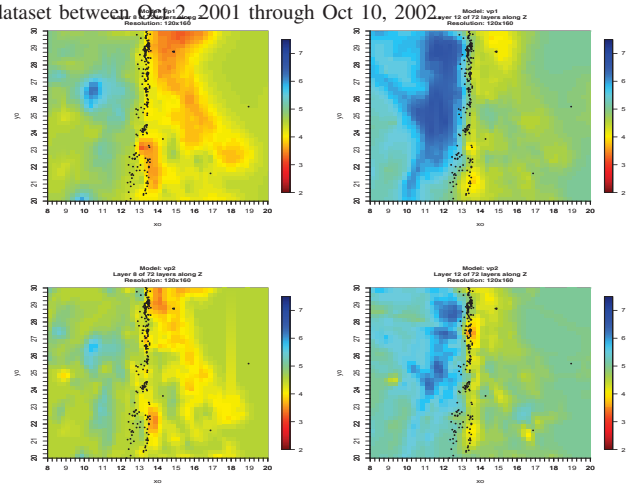
(b) Arrival time picking Error (Compared with manual picking [18])



(c) Comparison of earthquake locations calculated from our algorithm (black dots) to that of [18] (red dots).

(d) GUI plotting earthquake location (red) in real-time. Green dots indicate the station location

Fig. 6. Indoor simulation result on DRISTI system using Parkfield, CA dataset between Oct 2, 2001 through Oct 10, 2002.



(c) Vp at depth = 1km

(d) Vp at depth = 2km

Fig. 7. Horizontal slices of the P-wave velocity at depths of 1km and 2km. The vertical partition at the center shows the san-andreas fault of california and the black dots depicts the earthquake location. First row represents results from [18]; Second row results from our DRISTI system.

B. Outdoor Field Test

In this section we demonstrate robustness of DRISTI system by performing outdoor field test. Performing an end-to-end outdoor testing becomes tricky, as there is no mechanism to generate any physical events (earthquake) that can be sensed by the geophones. Geophysicists use a method called *Hammer Shock Test* to validate their seismic instruments. In

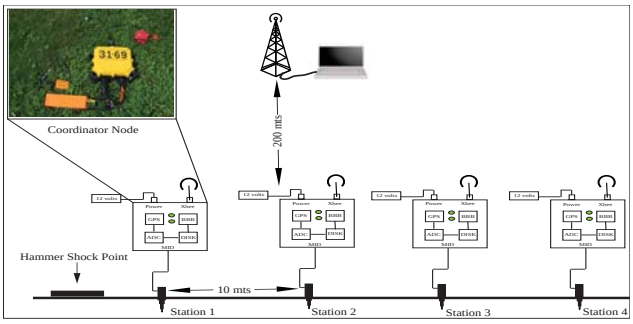


Fig. 8. Illustration of our deployment setup for hammer test at Lilburn, GA. Four coordinator nodes are aligned in a straight line and control center is set up at a distance of 200 mts.

this technique, a hammer shock is created at certain place and the shock wave generated is sensed by the geophones. One drawback of this scheme is that the shock waves are weaker and cannot penetrate deep inside the earth. This restricts us from obtaining the subsurface characteristics beyond 1-5 meters below the ground.

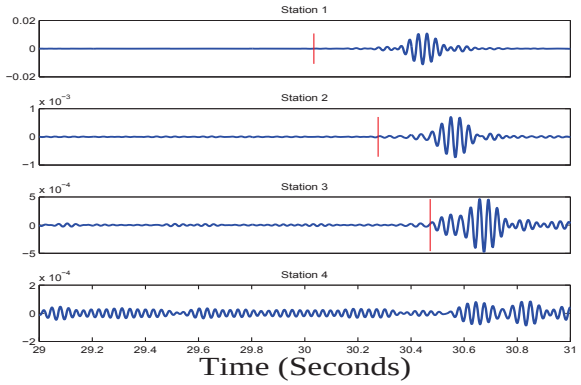
The test was conducted in a outdoor location near Lilburn, GA, USA. US Geological Survey and Earthscope has a seismic station Y52A [4] deployed in this region as a part of USArray project. Since the hammer shocks generated are weak we used only four coordinator node in this experiment that also acted as an edge node. These nodes where placed in a linear fashion spaced ten meters apart as shown in Fig. 8. Hammer shocks were created near the station 1 at a one minute interval for a total duration of 14 minutes. The shock wave created were sensed by geophones and recorded in the coordinator node. This was then processed in real-time using AT algorithm. This picked time was sent over to control center placed 200 meters away from the stations. The seismic vibration recorded in each station is shown in Fig. 9(a). Station 4 located 40 meters away from hammer shock location was not able to record any event due to reduced wave energy. This station however recorded some noise which we suspect to be mainly from nearby electric unit (Fig. 9(b)).

1) System Validations using Scientific Facts

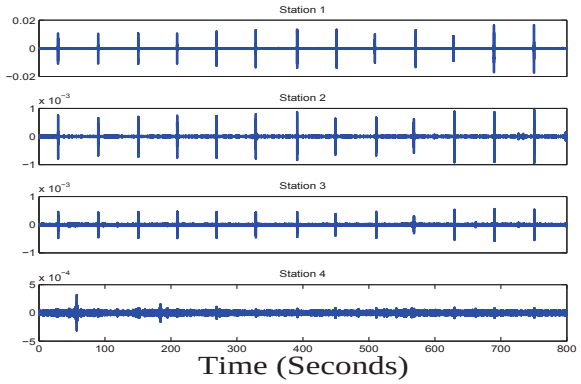
In this section, we validate our system by cross-checking the results obtained from the outdoor experiments with some of the scientific facts. i) Shock wave has higher amplitude near the source and as the distance increases from the source the wave energy decreases. This can be seen from Fig. 9(a) where the maximum amplitude was observed at station one (0.02) and decreased with the distance (station 2-4 smaller amplitude). ii) We know that shock wave travels linearly in a homogenous medium i.e comprised of similar structure like stone, sand etc. This principle can be observed from Fig. 9(a), where the wave reaches first at station 1 followed by others. The red vertical line indicates time of the arrival of the shock wave at each stations. iii) The surface wave speed is around 23-27 m/s and depends on soil composition, density etc. This can be observed from Fig. 10(b) where we obtained the mean velocity between station 1-2 as 25.4 m/s while between station 2-3 as 24.2 m/s.

We observed an anomaly at the 13th sec indicating the negative velocity which means the wave might have not traveled linearly. We suspect this to be a miss-picking or due to another

[4]http://usarray.seis.sc.edu/station.html?zip_or_station_code=Y52A&x=0&y=0



(a) Two second snapshot of picking across 4 stations. Red line indicate the time picking. We can observe the time shift in arrival of wave at each station.



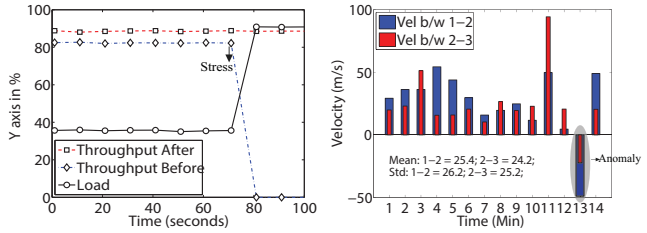
(b) Seismic signal recorded at 4 station during the whole test (14 sec). The blue vertical spike indicates the detection of shock. The amplitude of shock (energy) decreases from station 1 to 4

Fig. 9. Arrival time pickings result for hammer shock test. The testing was conducted for approx 14 min and shock was created at an interval of 1 sec

similar event created by external source. It should also be noted that the physical dynamics at the subsurface is complex and cannot be accurately understood by generating shock waves created by hammer. However, from this experiment we can confirm the correctness of the hardware and the algorithms in the DRISTI system. The arrival time picking algorithm in DRISTI is robust along with the data-logging and other peripheral units. The control center in this experiment was located 200 meters away from station and it was able to receive all the arrival time with zero packet loss. In future, we intend to run few long term experiments in much harsher environment such as forests/mountains in order to validate the communication robustness.

C. Stress Testing

On an average, few tens of earthquakes occur in a day. However, it is found that due to after shocks, the number could be higher (around 200) after a major earthquake. High rate of earthquake has direct effect on communication module as more arrival time will be picked and sent to the control center increasing the load at this module. So to test this, we created more than 500 earthquake events in a single minute using our indoor setup. Arrival time algorithm was able to detect all 500 events correctly. However, the Xbee module though exhibited good performance initially, experienced a dramatic drop in the throughout (Fig. 10(a)). Upon rigorous analysis using the indoor DRISTI system testbed we were able to arrive at the root cause of this problem. It so happened that, when



(a) Stress test on DRISTI system which helped us to detect crucial system bug. Throughput of the system before and after the bug.

(b) Velocity of the shock wave observed between each stations. Mean velocity between station 1-2 is 25.4 m/s and between 2-3 is 24.2 m/s.

Fig. 10. DRISTI system results during stress and hammer test

there was a burst of data, as was the case here, the serial communication buffer which is allocated a 4kB by the system was getting exhausted. We were able to rectify the problem by waiting until sufficient space was available in the serial buffer rather than continuously try to force send a packet. The latter approach had an unintended consequence of cascading failure to send, which was remedied by an intelligent wait and send mechanism followed by the former approach. The stress test therefore was instrumental in isolating and rectifying a critical error which could have proved problematic during an actual field test.

VI. Lessons Learned

Seismic tomographic imaging is a complex process and designing an end-to-end system requires a great set of domain knowledge. The parameters such as sensitivity threshold, regularization coefficient etc are data dependent. This needs to be carefully chosen with the help of a domain expert and also by running several experiment in various settings. After processing hammer shock data we observed that the data recorded can be easily corrupted by external noise, which in-turn can cause miss-picking. Occasionally we observed that seismic station lost the GPS signal, due to which the meta-data file got corrupted and caused the program to crash.

VII Conclusion

DRISTI system was designed to perform in-network processing of voluminous seismic signals and to transfer only certain relevant information in a large scale mesh network of stations. We used a two-tier hierarchical system design which used low-cost, low-power custom designed seismic station (edge node). These nodes basically sampled seismic signal and performed arrival time picking sending only relevant pick timings to its immediate coordinator node. The coordinator node is an extension of an existing seismic system. We added computation (BBB) and communication (Xbee) unit which made the stations smart and allowed them to communicate among themselves. We also implemented several algorithms such as p-wave detection and picking, earthquake location and tomography imaging algorithm. Both the hardware's and the algorithms where tested using real data traces in a indoor testing environment. The results obtained where satisfactory and matched closely to the centralized method presented in [18]. We also validated DRISTI system using outdoor hammer shock experiment and the system was validated for its robustness and accuracy. Stress test was performed on the system which was able to identify a bug in our communication module. Using DRISTI we are now able to obtain complex physical dynamics of the earth's interior in real-time. The design of this system can be extended to other cyber-physical system such as smart-grid, smart-home and perform in-network processing of large data (big-data analytics).

References

[1] R. Fischer and J. M. Lees. Shortest path ray tracing with sparse graphs. *GEOPHYSICS*, 58(7):987–996, July 1993.

[2] L. Geiger. Probability method for the determination of earthquake epicenters from the arrival time only. *Bull.St.Louis.Univ*, 8:60–71, 1912.

[3] G. Kamath, P. Ramanan, and W.-Z. Song. Distributed Randomized Kaczmarz and Applications to Seismic Imaging in Sensor Network. In *The 11th IEEE International Conference on Distributed Computing in Sensor Systems (IEEE DCOSS), Fortaleza, Brazil, 2015.*, 2015.

[4] G. Kamath, L. Shi, and W.-Z. Song. Component-Average Based Distributed Seismic Tomography in Sensor Networks. In *Distributed Computing in Sensor Systems (DCOSS), 2013 IEEE International Conference on*, pages 88–95, May 2013.

[5] G. Kamath and W.-Z. Song. *Tomography Imaging in Sensor Networks*, chapter 11, pages 283–301. Electronic and Optical Materials. Woodhead Publishing, Cambridge, United Kingdom, first edition, 2015.

[6] J. M. Lees and R. S. Crosson. Tomographic Inversion for Three-Dimensional Velocity Structure at Mount St. Helens Using Earthquake Data. *Journal of Geophysical Research*, 94(B5):5716–5728, 1989.

[7] J. M. Lees and R. S. Crosson. Bayesian Art versus Conjugate Gradient Methods in Tomographic Seismic Imaging: An Application at Mount St. Helens, Washington. *Institute of Mathematical Statistics*, 20:186–208, 1991.

[8] F.-C. Lin, M. H. Ritzwoller, and R. Snieder. Eikonal tomography: surface wave tomography by phase front tracking across a regional broadband seismic array. *Geophysical Journal International*, 177(3):1091–1110, June 2009.

[9] G. Liu, R. Tan, R. Zhou, G. Xing, W. Song, and J. Lees. Volcanic Earthquake Timing using Wireless Sensor Networks. pages 91–102, 2013.

[10] W. Menke. *Geophysical data analysis discrete inverse theory*. Elsevier/Academic Press, 2012.

[11] T. L. Murray and E. T. Endo. A real-time seismic-amplitude measurement system (rsam). volume 1966 of *USGS Bulletin*, pages 5–10. 1992.

[12] C. C. Paige and M. A. Saunders. LSQR: An Algorithm for Sparse Linear Equations and Sparse Least Squares. *ACM Trans. Math. Softw.*, 8(1):43–71, Mar. 1982.

[13] L. Shi, W.-Z. Song, M. Xu, Q. Xiao, J. M. Lee, and G. Xing. Imaging Seismic Tomography in Sensor Network. In *IEEE SECON*, 2013.

[14] W.-Z. Song, R. Huang, M. Xu, A. Ma, B. Shirazi, and R. Lahusen. Air-dropped Sensor Network for Real-time High-fidelity Volcano Monitoring. In *The 7th Annual International Conference on Mobile Systems, Applications and Services (MobiSys)*, June 2009.

[15] G. Werner-Allen, J. Johnson, M. Ruiz, J. Lees, and M. Welsh. Monitoring Volcanic Eruptions with a Wireless Sensor Network. *Second European Workshop on Wireless Sensor Networks (EWSN'05)*, Jan. 2005.

[16] G. Werner-Allen, K. Lorincz, J. Johnson, J. Lees, and M. Welsh. Fidelity and Yield in a Volcano Monitoring Sensor Network. In *Proc. 7th USENIX Symposium on Operating Systems Design and Implementation (OSDI)*, Nov. 2006.

[17] G. Werner-Allen, P. Swieskowski, and M. Welsh. MoteLab: A Wireless Sensor Network Testbed. In *Proceedings of the Fourth International Conference on Information Processing in Sensor Networks (IPSN'05), Special Track on Platform Tools and Design Methods for Network Embedded Sensors (SPOTS)*, Apr. 2005.

[18] H. Zhang, C. Thurber, and P. Bedrosian. Joint inversion for Vp, Vs, and Vp/Vs at SAFOD, Parkfield, California. *Geochem. Geophys. Geosyst.*, 10(11):Q11002+, Nov. 2009.

[19] H. Zhang and C. H. Thurber. Double-Difference Tomography: The Method and Its Application to the Hayward Fault, California. *Bulletin of the Seismological Society of America*, 93(5):1875–1889, Oct. 2003.