Traffic Deduction Exploring Sensor Datas Intra-Correlations in Train Track Monitoring WSN

Zhi Liu†, Toshitaka Tsuda†, Hiroshi Watanabe⋆
† Global Information and Telecommunication Institute, Waseda University, Tokyo, 169-8555 Japan
⋆ Department of Communications and Computer Engineering, Waseda University, Tokyo, 169-8555 Japan
Email: { {liuzhi,tsuda-toshitaka}@aoni.waseda.jp, hiroshi.watanabe@waseda.jp}

Abstract—**WSNs are good options to help monitor the scene of interest and notify the unusual happening to control center. But sensors' high sampling rates lead to tremendous network traffic over the bandwidth-limited and energy-critical WSNs, hence how to reduce the network traffic while maintaining the unusual events monitoring function becomes important. In this paper, we investigate the intra-correlations of the data generated by each sensor at different time instances. And we propose a traffic deduction algorithm exploring the sensor data's intra-correlations which could reduce the data volume significantly and guarantee the parameters needed for unusual detection are delivered.**

I. Introduction

Disaster notification becomes a hot research topic recently. Japan is one of the countries most affected by natural disasters, such as the earthquakes[1] and tsunami[2]. Meanwhile, landslide [3] because of the heavy rain or typhoon may cause serious problems to the society and human beings, e.g. they may destroy the nearby infrastructure facilities such as the train tracks. Hence how to monitor the scene of interest and detect the unusual in time to avoid unnecessary losses becomes a very important and emergent issue.

Wireless sensor networks (WSNs) [1–3] have been used in many industrial and consumer applications (e.g. industrial processing monitoring and control [4], machine and building health monitoring [5]) due to WSNs' efficiency, low-cost, 24-hour sensing ability, etc. Scene monitoring and unusual event detection [6, 7] is one typical application, and there are WSNs that have already been deployed to monitor the scene of interest and detect the unusual events automatically and continuously.

WSNs are composed by the sensor nodes and the sensor nodes are connected using the wireless networks. One critical issue of the WSNs is that the sensor node is with limited battery life hence the transmission capability of the sensor nodes are not that strong, which leads to a bandwidth-limited connection networks. Therefore the tremendous sensor data samples generated is a heavy burden for WSNs, how to manage the huge amount of sensor data becomes an import issue.

The train track monitoring system is studied, where wireless sensor network is deployed in *Japan Railway Technical Research Institute* (RTRI) for the purpose of measuring transmission capability of the Wi-SUN sensor module. In this system, there are two terminals (East and West) and the train travels between them. Eleven Wi-SUN sensors[4] are placed along the train track. The senor is developed by *National Institute of Information and Communications Technology* (NICT), Japan, and each sensor detects the acceleration in the x-direction, y-direction and z-direction periodically. The sensors are running on two modes in term of broadcasting the data samples to the router: normal mode and power-saving mode. The sensors in normal model broadcast the data samples to the router every fixed time duration. Sensors in power-saving modes only broadcast the data when the captured scene incurs a value larger than the pre-assigned threshold value. If the sensor data captured is below the threshold, the sensor nodes in power-saving mode do not broadcast their captured data. Since normally there is no 'unusual' (defined as when the captured value is larger than the threshold) happening, this rare transmission greatly saves their sensor energy consumptions, where the sensor energy is critical in WSNs. The control center will collect all the transmitted data, where the data analysis is performed.

In this paper, we analyze the parameters used in this setup, and calculate the correlations between these parameters in different direction (or in x-direction, y-direction and z-direction), using *Pearson product-moment correlation coefficient*. We find there exists a near linear intra-correlation inside these data and hence we propose a scheme to reduce the redundant data traffic while maintaining the detection ability. Please note that the method could also be applied to other WSNs, especially

[1]Japan Meteorological Agency | Earthquake Information http://www.jma.go.jp/en/quake
[2]Japan Meteorological Agency|Tsunami Information http://www.jma.go.jp/en/tsunami
[3]'LANDSLIDE IN JAPAN' Homepage. http://www.tuat.ac.jp/ sabo/lj
[4]http://www.wi-sun.org/

which uses the same sensor nodes.

The rest of this paper is organized as follows: related work in the field of the WSNs and traffic reduction are introduced in Section II. Section III introduces the WSN system, and show the parameters that are applied for the 'unusual' detection. We then analyze correlation of these parameters and propose the traffic reductions scheme in Section IV. Section V concludes this paper.

II. Related works

This section introduces the WSNs and the sensor data reduction methods in literature.

A. Wireless sensor networks

WSNs [1–3] are mature topics in literature and they have been widely used in many industrial and consumer applications such as the industrial processing monitoring and control [4], machine and building health monitoring [5] due to their efficiency, low-cost, 24-hour sensing ability, etc. WSNs are more popular and necessary in countries like Japan, where the nature disasters happen quite often. By deploying the sensor nodes in the scene of interest and connecting them using wireless such as Wi-Fi [8], WSNs can be utilized to help monitor the scene of interest and detect the unusual events. After collecting the data, the control center can then analyze the captured sensor data to judge whether there is unusual and disaster happening or not, hence the scene of interest can be monitored by the sensor nodes[5].

B. Sensor data reduction

Given the tremendous data samples generated by the sensors result a huge burden for the WSNs. There are many works discussing how to reduce the data traffic such as [9–12]. While most of the related work uses the in-network processing (process the data at the node before transmission) or filter out the unnecessary information before forward to the next hop, while is similar with the sensor in power-saving mode. Some typical work is listed as follows: [9] proposed a sensor-oriented data reduction for estimation with WSNs, where censoring and quantization was applied. Since in-network data fusion can reduce data redundancy, [10] introduced an adaptive data fusion method for energy efficient routing.

Different from these work, this paper investigates the intra-correlation of the generated acceleration data of the sensors in the x-direction, y-direction and z-direction. Given the detection parameters used in this paper are highly (almost linear) correlated, we propose a novel scheme to reduce the redundant transmission while maintaining the detection performance.

III. System Overview

This section introduces the train track monitoring system and the parameters used for the detection.

[5]Local sensor data processing is also feasible if the sensor nodes have certain level computation capabilities, and this could reduce the network traffic.

A. System overview

The train track monitoring system studied in this paper is shown in Fig. 1. The main objective of the system is to examine the transmission performance when the Wi-SUN sensors are placed near the real railway track. Since these sensors are sending actual data, we can utilize these data to develop an efficient Wi-SUN sensor systems. There are two terminals (East and West) in this system, and the train travels between the two terminals on the track as illustrated in the figure. The carriage of the train is empty. Wireless sensor nodes are placed along the train tracks to help monitor the train track. The sensors nodes used are the Wi-SUN sensors and there are eleven sensors as illustrated in Fig. 1. Each sensor detects the acceleration value in the x-direction, y-direction and z-direction, i.e. in the horizontal direction and vertical direction as shown in Fig. 2. Please note that each sensor has its own original point (0,0,0), and (0,0,0) is not set to be the original position. Sensors are connected with each other using Wi-SUN. The Wi-SUN router is connected with the control center using 3G. The corresponding value generated by the sensor nodes can then be observed at the control center after the transmission over the WSNs.

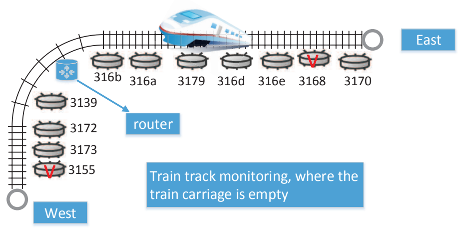

Fig. 1. Illustration of train track unusual monitoring system.

Sensor nodes involved in this system are running on two modes: normal and power-saving. Sensor nodes running on normal mode capture the nearby environment and broadcast the captured acceleration data to the data sink via the relay nodes periodically. Sensor nodes in power-saving mode also sense the nearby environment, but the sensed data is only delivered to the data sink when the captured data is above the pre-defined threshold. This is to say that the sensor nodes won't turn on its antenna when the acceleration data is below the threshold. This could greatly save the sensor nodes' energy consumptions, where the energy consumption is critical in WSNs. As illustrated in Fig. 1, sensor 3155 and sensor 3168 (marked by red 'V') are the sensors in power-saving mode and all the other sensor nodes are in normal mode.

B. Unusual notification parameters

Sensor nodes' captured acceleration value is normally bounded, and we call the event that lead to significantly different acceleration values from the normally

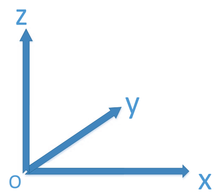

Fig. 2. Illustration of train track unusual monitoring system.

generated acceleration data as 'unusual' event. We fist define the parameters used for the detection for the normal mode sensor nodes. The normal mode sensor node broadcasts N consecutive samples every minute. Instead of using the raw data to perform the unusual detection, we use the mean, vibration and local variance as the unusual notification parameters. And they are defined as follows:

Since normal mode sensors broadcasts N consecutive samples one time during one minute, we denote this transmission period to be with instance number i. The data transmitted by sensor node j during this transmission windows are denoted as $d^1_{i,j,x}, \dots d^N_{i,j,x}$, where the x stands for the x-direction and this x could be replaced by y and z to represent the data samples in the corresponding direction. Hence the mean of these N samples $m_{i,j,x}$ can be calculated as

$$m_{i,j,x} = \frac{\sum_{k=1}^{k=N} d^k_{i,j,x}}{N} \tag{1}$$

The vibration indicates the samples' value changes, and is defined as follows

$$v_{i,j,x} = \frac{\sum_{k=1}^{k=N-1} |d^{k+1}_{i,j,x} - d^k_{i,j,x}|}{N-1} \tag{2}$$

The last parameter we use for the unusual event detection is local variance, which indicates the data samples' local changes. We use the local variance instead of the variance of all the N samples's variance since the sensor node is very sensitive. Then the local variance could be calculated as

$$l_{i,j,x} = \sum_{k=1}^{k=N} \sum_{r=max(1,k-\delta)}^{r=min(N,k+\delta)} \frac{(d^r_{i,j,x} - d^k_{i,j,x})^2}{Nn_r} \tag{3}$$

where δ defines the local, and δ is set to be 2 in the experiment. n_r is related with r and it tells how many neighboring samples are taken into consideration.

Please note these three equations are defined for the data samples in the x-direction, where the y-direction and z-direction notification parameters $m_{i,j,z}$, $v_{i,j,z}$, $l_{i,j,z}$ and $m_{i,j,z}$, $v_{i,j,z}$, $l_{i,j,z}$ can be calculated similarly by replacing the x-direction acceleration value into corresponding y-direction and z-direction value.

Next we show how to calculate the notification parameters for power-saving mode sensors. Since these sensors only transmit data when the sensed data is above the threshold and they will transmit all the captured data, we need to group them into the 'transmission windows'. In this paper, we count every N samples as a group to make this consistent with the normal mode sensors, and then we can use similar method as the normal mode sensors to generate the unusual notification parameters.

IV. Traffic Reduction

In this section, we will investigate the intra-correlations among the notification parameters in different directions of the same sensor. And then we propose a traffic reduction algorithm.

A. *Intra-correlation calculation*

Pearson product-moment correlation coefficient (PMCC) is employed for correlation calculation. And PMCC can be calculated as

$$\rho_{X,Y} = \frac{E[(X - m_X)(Y - m_Y)]}{\lambda_X \lambda_Y} \tag{4}$$

where X and Y represent the two data set, E stands for the expectation or the mean. m_X and m_Y are the mean value of data set X and data set Y, respectively. λ_X and λ_Y denotes the standard deviation of data set X and Y.

The real train track monitoring system was implemented in RTRI, Japan as shown in the Figure 1. The data we used in this experiment was collected between 10:00am and 20:00pm, May 20th 2015 by RTRI, where the train travels between the two terminals from 10:22:59am to 11:46:45am, from 14:12:11pm to 14:53:51pm and from 16:00:57pm to 16:35:00pm. Please note that the train stopped periodically for couple of minutes during this time period. The sensor nodes in normal mode send the data every one minute with the sampling period to be 0.05s. But the sensors do not transmit all the data to the control center but instead only 18 (N=18) consecutive generated samples during one minute. But the 'work' and 'sleep' time instances are different for different sensors, so at each time instance there are data generated by some sensor nodes. Power-saving mode sensors' sampling period is also 0.05s.

We then calculate the intra-correlation between the notification parameters of different directions, and the results are shown in Fig. 3. Where we can observe that the x-direction parameters and y (or z)-direction parameters are highly correlated, y-direction parameters and z-direction notification parameters are also highly correlated. In another word, there is redundancy in the sensor data, which leads to the possibility of reducing the data traffic.

B. *Traffic reduction*

Given the high intra-correlation of the notification parameters, we propose an algorithm for the traffic reduction while maintaining the detection performance. The general idea is that if u-direction parameters are

ID	mean (x,y)	local variance(x,y)	vibration (x,y)	mean (x,z)	local variance (x,z)	vibration(x,z)	mean (y,z)	local variance (y,z)	vibration (y,z)
316b	0.9972	0.9687	0.984	0.9964	0.9422	0.976	0.9958	0.9147	0.9617
3139	0.9982	0.9982	0.9968	0.998	0.9976	0.9889	0.9978	0.9972	0.9939
316a	0.998	0.998	0.9979	0.998	0.9979	0.9978	0.9979	0.9979	0.9979
316d	0.9965	0.9847	0.9816	0.9945	0.9707	0.9774	0.9945	0.9734	0.9679
316e	0.9963	0.9611	0.9814	0.9957	0.9348	0.969	0.9945	0.9184	0.9519
3170	0.9973	0.997	0.9939	0.9974	0.9972	0.9963	0.9974	0.9973	0.9949
3172	0.9953	0.9629	0.9866	0.9932	0.9472	0.9816	0.9916	0.9116	0.9719
3173	0.9907	0.9619	0.9798	0.9917	0.9556	0.9802	0.9836	0.9215	0.9639
3179	0.9958	0.9524	0.9773	0.9941	0.9325	0.9722	0.9921	0.8842	0.9527
3155	0.9985	0.9966	0.9989	0.9977	0.9937	0.9984	0.9969	0.9928	0.9986

Fig. 3. Illustration of intra-correlation coefficients. x/ y/ z stands for the corresponding value in x/y/z direction, respectively. mean (x, y) denotes the correlations between the mean of x-direction data and mean of the same sensors y-direction data. Similarly, local variance (x, y) and vibration (x, y) stand for the correlations between the local variance, vibration of the x-direction data and local variance, vibration of the same sensors y-direction data, respectively. The local variance and vibration could be calculated as introduced in the Motivation section. Other parameters such as mean (y,z) are defined similarly.

correlated with w-direction parameters (where u, w could be x, y or z), and the resulted PMCCis larger than τ, we can omit the corresponding direction's data samples. τ is pre-defined by the system and determines how many we sacrifice the notification performance, where $0 \leq \tau \leq 1$. The larger the τ is, the higher accuracy the system demands. We let $\rho_{m_{j,x},m_{j,y}}$ stands for the intra-correlation between sensor j's mean of x-direction data and mean of y-direction data. Similarly we can define $\rho_{v_{j,x},v_{j,y}}$, $\rho_{l_{j,x},l_{j,y}}$. By calculating the intr-correlation of different directions' data, we then can transmit part of all the data values to the control center. The detailed method is illustrated in 1, where u,w,s could be different direction's data samples among the x, y, z direction.

Algorithm 1 Traffic reduction method illustration

calculate all the intra-correlation values
if $\rho_{m_{j,u},m_{j,w}} \geq \tau$, $\rho_{v_{j,u},v_{j,w}} \geq \tau$, $\rho_{l_{j,u},l_{j,w}} \geq \tau$ and $\rho_{m_{j,u},m_{j,s}} \geq \tau$, $\rho_{v_{j,u},v_{j,s}} \geq \tau$, $\rho_{l_{j,u},l_{j,s}} \geq \tau$ **then** ▷ intra-correlation is $\geq \tau$
    sensor broadcasts u direction value alone.
else
    if $\rho_{m_{j,u},m_{j,w}} \geq \tau$, $\rho_{v_{j,u},v_{j,w}} \geq \tau$, $\rho_{l_{j,u},l_{j,w}} \geq \tau$ and $\rho_{m_{j,u},m_{j,s}} < \tau$, $\rho_{v_{j,u},v_{j,s}} < \tau$, $\rho_{l_{j,u},l_{j,s}} < \tau$ **then**
        sensor broadcasts u and s direction's value.
    else
        sensor broadcasts u, w, and s direction's value.
    end if
end if

V. Conclusion

Sensors' high sampling rates lead to tremendous network traffic over the bandwidth-limited and energy-critical WSNs, and is a serious problem for WSNs. This paper studies the train track monitorign system and targets to reduce the data traffic while maintaining the notification performance. We investigate the intra-correlations of the data generated by each sensor. And we propose a traffic deduction algorithm exploring the sensor data's intra-correlations which could reduce the data volume significantly and guarantee that the parameters needed for unusual detection are delivered.

VI. Acknowledgement

The research results have been achieved by"Research and Development on Fundamental and Utilization Technologies for Social Big Data", the Commissioned Research of National Institute of Information andCommunications Technology (NICT), JAPAN.

We also would like to thank Mr. Nagateru Iwasawa, Mr. Akio Hada from Signaling and Transport Information Technology Division, Railway Technical Research Institute, Japan, who provide us the experiment data.

References

[1] J. Yick, B. Mukherjee, and D. Ghosal, "Wireless sensor network survey," *Computer networks*, vol. 52, no. 12, pp. 2292–2330, 2008.

[2] I. F. Akyildiz, W. Su, Y. Sankarasubramaniam, and E. Cayirci, "Wireless sensor networks: a survey," *Computer networks*, vol. 38, no. 4, pp. 393–422, 2002.

[3] J. N. Al-Karaki and A. E. Kamal, "Routing techniques in wireless sensor networks: a survey," *Wireless communications, IEEE*, vol. 11, no. 6, pp. 6–28, 2004.

[4] V. C. Gungor and G. P. Hancke, "Industrial wireless sensor networks: Challenges, design principles, and technical approaches," *Industrial Electronics, IEEE Transactions on*, vol. 56, no. 10, pp. 4258–4265, 2009.

[5] G. J. Pottie and W. J. Kaiser, "Wireless integrated network sensors," *Communications of the ACM*, vol. 43, no. 5, pp. 51–58, 2000.

[6] K. K. Khedo, R. Perseedoss, A. Mungur *et al.*, "A wireless sensor network air pollution monitoring system," *arXiv preprint arXiv:1005.1737*, 2010.

[7] K. Lorincz, D. J. Malan, T. R. Fulford-Jones, A. Nawoj, A. Clavel, V. Shnayder, G. Mainland, M. Welsh, and S. Moulton, "Sensor networks for emergency response: challenges and opportunities," *Pervasive Computing, IEEE*, vol. 3, no. 4, pp. 16–23, 2004.

[8] F. Ohrtman and K. Roeder, *Wi-Fi Handbook: Building 802.11 b Wireless Networks*. McGraw-Hill New York, NY, 2003, vol. 67.

[9] S. Santini and K. Romer, "An adaptive strategy for quality-based data reduction in wireless sensor networks," in *Proceedings of the 3rd international conference on networked sensing systems (INSS 2006)*, 2006, pp. 29–36.

[10] H. Luo, J. Luo, Y. Liu, and S. K. Das, "Adaptive data fusion for energy efficient routing in wireless sensor networks," *Computers, IEEE Transactions on*, vol. 55, no. 10, pp. 1286–1299, 2006.

[11] Y. Zhu, E. Song, J. Zhou, and Z. You, "Optimal dimensionality reduction of sensor data in multisensor estimation fusion," *Signal Processing, IEEE Transactions on*, vol. 53, no. 5, pp. 1631–1639, 2005.

[12] E. J. Msechu and G. B. Giannakis, "Sensor-centric data reduction for estimation with wsns via censoring and quantization," *Signal Processing, IEEE Transactions on*, vol. 60, no. 1, pp. 400–414, 2012.