Role-based Deceptive Detection and Filtering in WSNs

Shinan Wang
Wayne State University
shinan@wayne.edu

Kewei Sha
Oklahoma City University
ksha@okcu.edu

Weisong Shi
Wayne State University
weisong@wayne.edu

1. PROBLEM STATEMENT

With more and more real applications of WSNs have been deployed, which are in charge of either monitoring parameters or event detection, we envision that the success of the WSNs is decided by the quality of the collected data [2]. Further more, the quality of the collected data is mainly affected by the deceptive data, which includes redundant data, very similar repeated readings that provide less information, and false data, wrong readings resulted from inaccurate components, unreliable wireless communication or malicious attacks. On one hand, redundant data should be filtered because less information is provided but considerable resources are wasted. On the other hand, false data should also be filtered to improve data accuracy.

Hence, the major concern of improving the data quality is to detect and filter deceptive data. However, because of limited resources, it is a big challenge to implement deceptive data detection, which is further complicated when system has high dynamics. Several schemes have been developed to solve this issue [1, 3], but lots of them concerns rather on the systems than the data itself. Moreover, few of them targets deceptive data detection in high dynamic systems. Thus, we are targeting to detect and filter the deceptive data in high-dynamic event-driven WSNs from the data itself point of view by proposing a *Role-based Deceptive Detection and Filtering* (RD^4) mechanism in this paper. The detail of our approach is listed as follows

2. OUR APPROACH

First, each sensor picks up a role from the role set based on the specific features of the sensor, such as storage size, computation ability, communication ability and trustable level. For each event it sensed or received, the sensor issues a confidence score to the event, which is denoted as $csr(E,T)_{ij}$, indicating the truth level of this event, where E specifies the event, and i and j are the identity of the role and the identity of the sensor, while T means the score will be valid for T time slots. Moreover, different roles have different maximum confident scores that can be issued toward an event.

In RD^4, the confidential score defined above is calculated based on the accumulated signal strength during a certain time period ($[0,T_0]$), depicted as $ASS(E,T_0)_{ij}$, of the corresponding event E at sensor j with role i. Here, the signal strength of an event E, denoted as $SS(E)_{ij}$, can be defined as the amount of changes of a monitoring physical parameter within a unit time period. For example, if we try to detect an event of sudden changes in temperature at a computing node in a high performance computing system. The signal strength will be the amount of temperature changing within each minute. Thus, if the function to specify the changing rate of a monitoring physical parameter is $p(t)$, we can define $SS(E)_{ij}$ as

$$SS(E)_{i,j} = p(t)dt \tag{1}$$

Based on the defined $SS(E)_{ij}$, the accumulated signal strength can be defined as

$$ASS(E,T_0)_{ij} = \int_0^{T_0} SS(E)_{ij} = \int_0^{T_0} p(t)dt \tag{2}$$

Having (2), we design a function f that maps the accumulated signal strength to a confidential score, to be specific, $csr(E,T)_{ij} = f(ASS(E,T_0)_{ij})$. When an event is detected at a sensor, the sensor will set up a timer T, also used as the first lifetime period of the event, to the detected event. Then the sensor will try to confirm whether it is a real event before the event expires. The decision is made based on the confidence score, which can come from two sources. One is the

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
Copyright 200X ACM X-XXXXX-XX-X/XX/XX ...$5.00.

observation by the sensor itself, for which we use accumulated signal strength detected by the senor, and the other is the reported signal strength about the same event from other sensors. Then the sensor j assigns a confidence score, $csr(E,T)_{ij}$, to the detected event, E, within a lifetime of T as follows.

$$csr(E,T)_{ij} = f(ASS(E,T_0)_{ij})$$
$$= \begin{cases} ASS(E,T_0)_{ij} & ASS(E,T_0)_{ij} \leq CSR(E,T)_{ij} \\ CSR(E,T)_{ij} & ASS > CSR(E,T)_{ij} \end{cases} \quad (3)$$

If the calculated confidence score exceeds the pre-set bound to confirm an event, the sensor that detects the event will broadcast the event report to nearby sensors in the system; otherwise, after a certain time T the event is confirmed as false and discarded. In this way, we can image that the true event will be propagated very fast but it will be dropped after a while either because the farther sensor cannot detect the event or not so many sensor are sensing the reports. On the other hand, false event report will be discarded from the beginning for lack of evidence.

3. CASE STUDY

The RD^4 mechanism is a general mechanism to detect deceptive data. In this section, we adapt the RD^4 mechanism to detect false accident report in the context of a high dynamic system - vehicular networks.

We define the role set of sensors for this application as following: Road Side Units(RSUs), public vehicles such as police cars, school buses and so on, regular vehicles like personal owned cars, and vehicle itself. Thus, $R = \{R_{rsu}, R_{pub}, R_{reg}, R_{self}\}$. For each role R_i, it can assign a maximum confidence score, CSR_{ij}, to an accident report it detects or confirms based on the role.

Based on the reality that traffic will be blocked so that the vehicles will slow down when an accident happens. We use the vehicle velocity deceleration as the signal strength defined in the model: $p(t) = a(t) = dv/dt$, where $a(t)$ is the acceleration rate, and v is the velocity of the vehicle. Then, the signal strength observed by vehicle j, $ASS(E,T_0)_{ij}$, can be calculated as $ASS(E,T_0)_{ij} = \int_0^T a(t) = v_T - v_0 = \Delta v$.

Based on the above definition, we simulate RD^4 by extending a traffic simulator which simulates the movements, such as acceleration, deceleration, lane changing, of the vehicles. The primary results are shown below:

In figures 1 and 2, the x-axis is the traffic density on the road and the y-axis shows the recall. we can easily observe that the RD^4 mechanism detects 99.9% false accident reports in most cases and more than 95.7% true accident reports are confirmed while about 5% reports are misclassified as false accident reports.

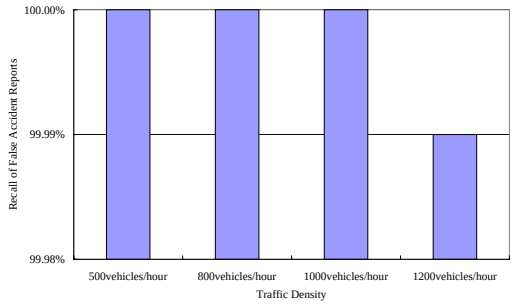

Figure 1: Recall of false accident reports.

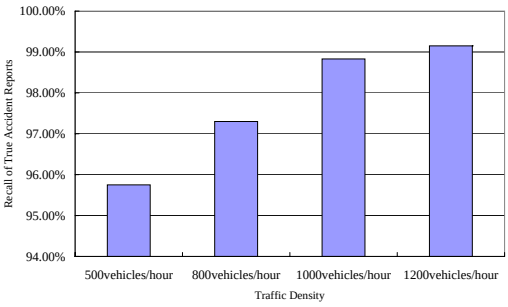

Figure 2: Recall of true accident reports.

4. CURRENT STATUS AND FUTURE WORK

Since we classify the deceptive data into two different categories, redundant data and false data and the sensing systems are classified as passive monitoring and event detection, the RD^4 mechanism we proposed in this paper, focusing on false data detection in event detection sensing systems, is just one part of our general framework to detect and filer deceptive data. Other aspects of our project, including how to discover redundant data and filter it at lower costs and how to utilize the spatial-temporal relation of the readings to detect and filter both redundant and false data, will be our future work.

In addition, if we count on various multiple parameters monitored by different types of sensors to build the event detection sensing system, he RD4 mechanism could be extended to a multi-modality version, which can guarantee more efficient and effective deceptive data detection and filtering. This will also be our future work.

5. REFERENCES

[1] V. Hodge and J. Austin. A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22:85–126, 2004.

[2] K. Sha and W. Shi. Consistency-driven data quality management in wireless sensor networks. *Journal of Parallel and Distributed Computing*, 68(9):1207–1221, 2008.

[3] S. Shekhar, C. Lu, and P. Zhang. A unified approach to spatial outliers detection. *Geoinformatica*, 7(2), 2003.