Acta Technica *62* (2017), No. 7A, 1023–1032 

# Anomaly event-driven based cluster topology detection algorithm

Yongxiong Zhang[1], Liangming Wang[2], Dong Li[3]

**Abstract.** It is a hot topic in wireless sensor network to effectively use the energy of sensor nodes and improve the probability of anomaly event detection. To this end, the Anomaly Event-Driven Based Cluster Topology Detection (AED-CTD) algorithm is proposed. The AED-CTD algorithm establishes a cluster through the location of the anomaly event, and then detects the event by the nodes in the cluster. When establishing clusters, the detection probability of nodes to events and the residual energy of nodes are considered. Only the nodes with the residual energy greater than the energy threshold can join the cluster. At the same time, the dynamic energy threshold is used to balance the energy consumption. The experimental data show that AED-CTD algorithm has a low missed detection rate. At the same time, the energy consumption of AED-CTD algorithm and CCM and GEP-ADS algorithm is reduced by nearly 4.1% and 5.8% respectively.

**Key words.** Wireless sensor networks, Anomaly event, Cluster, Energy, Threshold.

## 1. Introduction

Wireless Sensor Networks (WSNs) has been widely used in various applications [1-3], such as environmental monitoring, health care and battlefield investigation. WSNs realize the monitoring of the application environment by sensing the environment data in real time through the sensor nodes. Detection of anomaly events is one of the most important applications of WSNs[3]. In case of anomaly events, such as fire, pollutants leakage, the sensor nodes will timely sense the anomaly event location, and transmit this information to sink. However, due to the disturbance of the external environment and the influence of the characteristics of the sensor nodes,

[1]Department of Economics and Trade, Guangzhou College of Technology and Business, Guangzhou 510850, China

[2]School of Computer Science & Engineering, South China University of Technology, Guangzhou 510006, China

[3]School of Software Engineering, South China University of Technology, Guangzhou 510006, China

the missed detection of anomaly events may occur. Among them, the sensor nodes are limited in their own resources, such as storage capacity, data processing capacity and especially energy constraints. Since WSNs are often deployed in the field severe environment, it is not operable to recharge the energy or replace the battery for the sensor nodes [5]. Therefore, reducing the energy consumption of sensor nodes, increasing the energy utilization rate of nodes, and prolonging the detection time of the environment have become research priorities of WSNs. Reference [6] uses machine learning algorithms to detect anomaly events and cluster the sensor nodes through k-Nearest Neighbor algorithm. In Reference [7], the descent iteration algorithm is used to detect anomaly events, and the detection probability is improved by adjusting the parameters. However, these detection algorithms do not take into account the problem of node energy. To solve the problem of anomaly events detection, an Anomaly Event-Driven Based Cluster Topology Detection (AED-CTD) algorithm is proposed in this paper. The AED-CTD algorithm first calculates the detection probability of sensor nodes for anomaly events, and then establishes a cluster through this probability and the residual energy of the nodes. That is, each event corresponds to one cluster, which is responsible for the detection of anomaly events.

## 2. GDET-CH algorithm

Upon detection of anomaly events, it will sense the environmental data and transmit the data to sink, as shown in Figure 1.

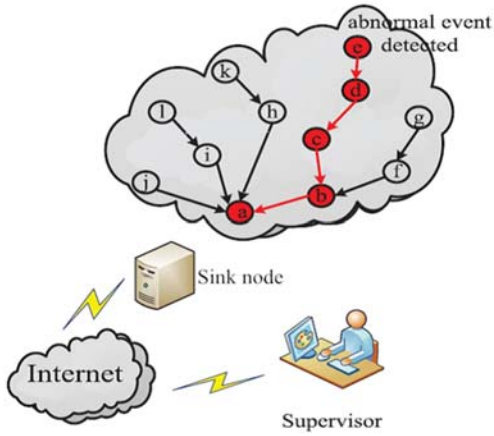


Fig. 1. Data transmission routing

Refer to the energy transmission model shown in Figure 2, Figure 2 (a) indicates the transmitter module and Figure 2 (b) shows the receiver module. The total energy consumed per bit in transmission is $E_{bit}$, as shown in Formula (1):

$$E_{bit} = \frac{P_{PA} + P_c}{R_b}. \tag{1}$$

Where, $P_{PA}$ is the energy consumption of the amplifier unit, $P_c$ is the energy consumption of the receiver circuit unit, and $R_b$ is the data transmission rate.

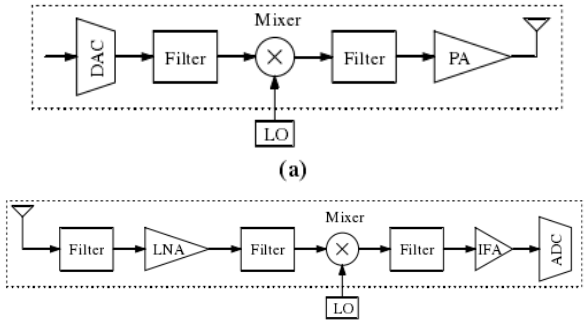


Fig. 2. Energy transmission module

And $P_{PA}$ can be defined as:

$$P_{PA} = (1 + \alpha)\, P_{out}\,. \tag{2}$$

Where, $\alpha = (\xi/\eta) - 1$. $\xi$ is the peak-to-average ratio, $\eta$ is the energy consumption rate of the wireless frequency power amplifier, and $P_{out}$ is the transmitting power, which is defined as shown in the Formula (3):

$$P_{out} = \bar{E}_b R_b \frac{(4\pi d)^2}{G_t G_r \lambda^2} M_\ell N_f\,. \tag{3}$$

Where, $\bar{E}_b$ represents the energy consumed per bit received by the receiver, and $d$ is the transmission distance. $G_t$ and $G_r$ respectively indicate the gain of transmitting and receiving antenna. $\lambda$ is the carrier wavelength, $N_f$ represents the receiver noise, and $N_f = N_r/N_o$. $M_\ell$ is link additive noise.

The definition of reference variable $\rho$ is as shown in Formula (4):

$$\rho = (1 + \alpha)\, \bar{E}_b R_b \frac{(4\pi)^2}{G_t G_r \lambda^2} M_\ell N_f\,. \tag{4}$$

According to the Formula (11), it can be known that $P_{PA} = \rho d^2$. In addition, the power consumption of the transmission circuit is further divided into the power consumption of the transmission and receiving circuit, that is $P_c = P_{ctx} + P_{crx}$. The definition of $P_{ctx}$ is as shown in Formula (5).

$$P_{ctx} = P_{DAC} + P_{filt} + P_{mix}\,. \tag{5}$$

Where, $P_{DAC}$, $P_{filt}$ and $P_{mix}$ are respectively the power consumption of the D/A converter, filtering and mixer modules.

The definition of $P_{ctx}$ is shown in Formula (6).

$$P_{rtx} = P_{LNA} + P_{IFA} + P_{mix} + P_{ADC}\,. \tag{6}$$

Where, $P_{LNA}$, $P_{IFA}$ and $P_{ADC}$ are respectively the power consumption of the low noise amplifier, the intermediate frequency amplifier and the D/A converter.

## 3. AED-CTD algorithm

Considering the random distribution of $n$ nodes in the sensing area of $A \times B$. At the initial stage of deployment, each sensor node has a certain data processing and communication capability. It is assumed that $\mathbf{S}$ is the set of all the sensor nodes within the network, as shown in Formula (7):

$$\mathbf{S} = \{S_1, S_2, \cdots, S_n\} \ . \tag{7}$$

Where, $S_i$ represents the $i$th sensor node.

Orienting at specific event and choosing a cluster of nodes to detect the event has become one of the research hotspots in WSNs. One of the core tasks of sensor nodes is to monitor and collect important sudden event information.

It is assumed that $k$ events occur at the time $t$, and the location of these events is known. To this end, it is assumed that the locations of these events are respectively expressed as $\left(e_x^\ell, e_y^\ell\right)$, and $\ell = \{1, 2, \cdots, k\}$. $(e_x, e_y)$ represents the two-dimensional coordinates of the event.

The AED-CTD algorithm is based on events and to establish cluster nodes that can sense the event. The cluster nodes are established mainly through the distance from the occurrence of events, and then the probability of event sensing is established by distance. Finally, the cluster nodes are formed according to the probability.

### *3.1. Cluster nodes*

In order to simplify the description, it is assumed that the cluster nodes at the time $t$ consist of $n_b$ sensor nodes, and between different clusters, $n_b$ is different. At the same time, it is assumed that the $k$ events occurring at the time $t$ form $k$ clusters. The definition of the sensor node set $N_e^k$ of the $i$th cluster is shown in Formula (8):

$$N_e^i|_t = \left\{ s_e^i | \hat{e} = 1, 2, \cdots, n_b^i \right\} \ . \tag{8}$$

Where, $n_b^i$ represents the number of sensor nodes of the $i$th cluster.

In addition, all clusters formed at the time $t$ do not overlap each other, that is:

$$N_e^1|_t \cap N_e^2|_t \cap \cdots \cap N_e^k|_t \in \varphi \, . \tag{9}$$

The purpose of the AED-CTD algorithm is to establish $k$ clusters of the events combined with the node sensing probability.

### *3.2. Event sensing probability*

When the cluster is established, the distance from the sensor node to the location of the event is first considered. It is assumed that the location of the $\ell$th event is

$\left(e_x^\ell, e_y^\ell\right)$, and $\ell = \{1, 2, \cdots, k\}$. The location of the sensor node $S_i$ is $\left(s_x^i, s_y^i\right)$, the distance $d(\ell, i)$ from the node $S_i$ to the event $\ell$ is:

$$d(\ell, i) = \sqrt{\left(e_x^\ell - s_x^i\right)^2 + \left(e_y^\ell - s_y^i\right)^2}. \tag{10}$$

It is assumed that the sensing radius of the sensor node is $R_s$, and the sensing range of each node is a circle that takes the node position as the center and with $R_s$ as the radius. Considering the external environment interference, there exist errors of the sensing radius of the nodes. Assuming that the sensing error is $R_u$, the sensing range of the node is shown in Figure 3.

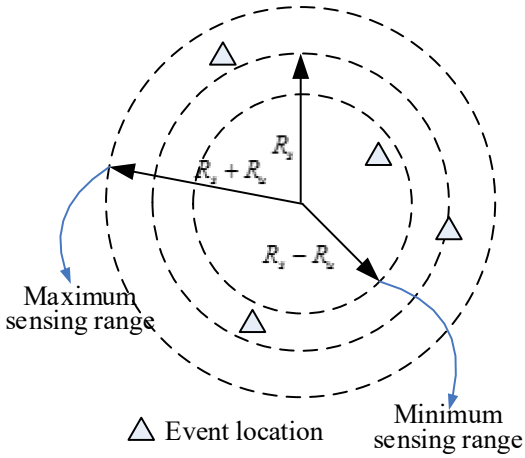


Fig. 3. Node sensing area

Using the probability coverage model shown in Reference [8], the probability $\sigma_{i,\ell}(e_\ell)$ of sensing event $e_\ell$ by the sensor node $s_i$ can be expressed as shown in Formula (11):

$$\sigma_{i,\ell}(e_\ell) = \begin{cases} 0, R_s + R_u \le d(\ell, i) \\ \exp\left(-\lambda a^\beta\right), R_s - R_u \le d(\ell, i) \le R_s + R_u \\ 1, R_s - R_u \ge d(\ell, i) \end{cases} \tag{11}$$

Where, $R_u$ represents the sensing error, and $\alpha = D(\ell, i) - (R_s - R_u)$. $\lambda$ and $\beta$ are respectively the system parameters used to measure the detection probability.

### *3.3. Energy factor*

When choosing the cluster head, the GDET-CH algorithm fully considers the residual energy of the node. When the residual energy $E_{re}$ of the node is greater than the threshold $E_{th}$, it may become a cluster head. The threshold $E_{th}$ setting plays a key role. To this end, the adaptive threshold mechanism is adopted.

First, $\psi = \{\psi_1, \psi_2, \cdots, \psi_m\}$ is referred to indicate the scope of the node energy,

and $\psi \in [0,1]$. With the proceeding of the work of the node, the energy of the node must be reduced. In other words, at first, the threshold $E_{th}$ can be higher because most nodes have more energy. However, after a period of time, the node energy will decrease. If the node threshold $E_{th}$ is too high or does not change, the energy of most nodes will be lower than the threshold $E_{th}$.

To this end, the adaptive threshold mechanism is adopted. First $\hbar$ thresholds are set by $\psi$, which are respectively expressed as $E_{th}^1$, $E_{th}^2$, $\cdots$, $E_{th}^{\hbar}$. Their definitions are shown in Formula (12):

$$E_{th}^{\hbar} = \frac{|\psi_1 - \psi_{\hbar}|}{\lambda^h}, h = 1, 2, \cdots, \hbar\,. \tag{12}$$

Where, $\lambda$ is a adjustable parameter.

At first, $E_{th}^1$ is used as a threshold. When all the nodes inside the network have served as the cluster head, use $E_{th}^2$ as the threshold, etc. In this way, the threshold can be consistent with the energy consumption of the network

### *3.4. Cluster formation process*

First define the matrix $\mathbf{D_1}$ of $k \times p$ dimensions, which indicates the probability of each event occurring by all nodes, as shown in Formula (13):

$$\mathbf{D_1} = \begin{bmatrix} \sigma_1(1,1) & \sigma_1(1,2) & \cdots & \sigma_1(1,p) \\ \sigma_1(2,1) & \sigma_1(2,2) & \cdots & \sigma_1(2,p) \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1(k,1) & \sigma_1(k,2) & \cdots & \sigma_1(k,p) \end{bmatrix}. \tag{13}$$

Where, $p$ represents the maximum number of nodes that all clusters can include, that is $p = \max\left\{n_b^i | i = 1, 2, \cdots, k\right\}$.

It can be known from the Formula (13) that, the first row of the matrix $\mathbf{D_1}$ indicates the probability that all nodes of the sensor node can detect the first event location. To this end, $\sigma_{\mathbf{1}}(\ell)$ is used to represent the distance from all nodes to the location of the $\ell$th event, that is $\sigma_{\mathbf{1}}(\ell) = \left\{\sigma_{1(\ell,1)}, \sigma_{1(\ell,2)}, \cdots, \sigma_{1(\ell,p)}\right\}$, and $\ell = 1, 2, \cdots, k$.

Next, the AED-CTD algorithm sorts the matrix $\mathbf{D_1}$, in the first row, the matrix is sorted from small distance to large distance. That is, $\hat{\sigma}_{\mathbf{1}(\ell)} \leftarrow Sort\left\{\sigma_{\mathbf{1}(l)}\right\}$, and $\ell = 1, 2, \cdots, k$. At the same time, the ordinal number vector $\mathbf{p}_{\mathbf{1}(k)}$ is used to represent the ordinal number of each node. Accordingly, the sorted matrix $\mathbf{D_1}$ is expressed as $\hat{\mathbf{D}}_{\mathbf{1}}$, and the ordinal number matrix is $\mathbf{P_1}$.

The pseudo code in the cluster head generation process is shown in Figure 4. First, the matrixes $\mathbf{D_1}$, $\mathbf{P_1}$ and $\mathbf{Q}_s$ are initialized, and the vectors $\sigma_{\mathbf{1}}$ and $\mathbf{p_1}$ are also initialized. Where, $\mathbf{Q}_s$ is the node state matrix, and the dimension is $k \times p$.

Then, from the first event and according to the Formula (10), the distance of each node to the event is calculated, that is $\sigma_{\mathbf{1}(1)}$, and then sort it again. Then the above process is repeated from second event until all sensor networks have calculated the distance from each event.

Next, based on the sorted matrix $\hat{\mathbf{D}}_1$, the first node with the largest probability of sensing event in each row of the matrix $\hat{\mathbf{D}}_1$ are found and included in the cluster. Considering that each cluster includes a maximum of $p$ nodes. In the formation of a cluster, each cluster is considered to only include $m$ nodes, and $m \leq p$. To this end, a variable $m$ is used to limit the number of nodes in the cluster.

At the beginning of the first row, the initial value of $\tau$ is 1, and the nodes with the residual energy greater than the threshold are included in the cluster, and add 1 to the $\tau$ value until $\tau > m$. Then, start from the second row and repeat the processes above until $k$ clusters are established.

Cluster formation algorithm

```
Input: n, E_th, E_re^i, i = 1, 2, ..., n
initial : D_1 ← φ, σ_1 ← φ, P_1 ← φ, p_1 ← φ
for ℓ ← 1 to k do
    for i ← 1 to n do
     σ_1(ℓ) ← σ_1 where σ_1 is calculated from (10)
    end for
     σ_1(ℓ) ← Sort(σ_1(ℓ))
     D̂_1(ℓ) ← d̂_1
     P_1(ℓ) ← p_1
end for
 τ ← 1
for ℓ ← 1 to k do
    for i ← 1 to n do
        if E_re^i ≥ E_th and τ ≤ m then
            N_a^ℓ ← i
           τ = τ + 1
      End if
   End for
End for
```

Fig. 4. Pseudo code for generating cluster algorithm

## 4. Performance simulation

### *4.1. Simulation parameters and performance indexes*

In order to better estimate the performance of GDET-CH algorithm, the simulation platform is established by Matlab software. The sensing area is selected as $100m \times 100m$ and the number of sensor nodes is 100. The initial energy of each node is $E_0 = 50J$. Other related energy parameters are as shown in Table 1.

Table 1. Energy parameters

| Parameter | Value | Parameter | Value |
|---|---|---|---|
| $B$ | 10kHz | $P_{IFA}$ | 30mW |
| $P_{mix}$ | 30.3mW | $P_{LNA}$ | 20mW |
| $P_{fNA}$ | 2.5mW | $P_{ADC}$ | 6.566mW |
| $N_f$ | 10dB | $P_{DAC}$ | 15.435mW |
| $M_L$ | 40dB | $M_L$ | 0.35 |

In addition, the occurrence rate $\lambda$ of anomaly events varies from 1 to 5. The simulation time is 100 seconds, each experiment is repeated 50 times independently, and the mean is taken as the final experimental data.

In addition, the performance of AED-CTD is analyzed by using the missed detection rate and energy consumption, and it is compared with the CCM [9] and GEP-ADS [10] algorithms. CCM and GEP-ADS are selected as references because they are typical anomaly event detection algorithms.

### *4.2. Missed detection rate*

The change curve of the missed detection rate with the occurrence rate of abnormal events is shown in Figure 5. It can be known from Figure 5 that, with the increase of the occurrence rate $\lambda$ of anomaly events, the missed detection rate is on the rise. In the range from 1 to 3 of $\lambda$, the missed detection rate of AED-CTD decreased by 40%-66.7% and 29%-55% respectively compared with the CCM and GEP-ADS algorithms. This is mainly because AED-CTD establishes clusters driven by event, and determines whether nodes join clusters based on event sensing probability and residual energy, which has good detection rate.

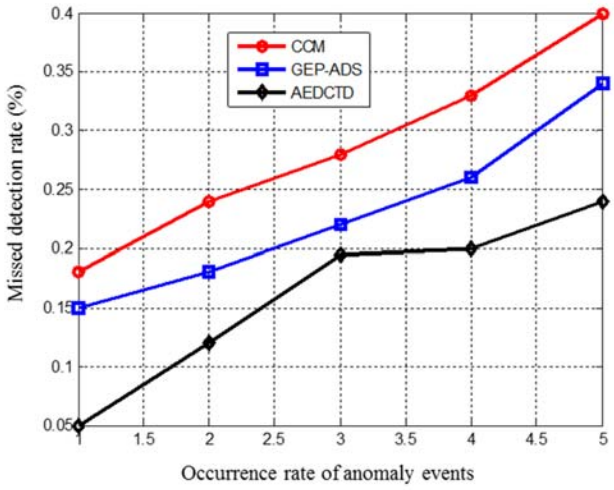


Fig. 5. Change curve of missed detection rate with the occurrence rate of anomaly events

### *4.3. Energy consumption*

Figure 6 analyses the energy consumption of the three algorithms. As shown in Figure 6, the energy consumption is decreasing with the increase of the missed detection rate. When the missed detection rate remains unchanged, the energy consumption of the AED-CTD algorithm is the least. Compared with CCM and GEP-ADS algorithms, the energy consumption of AED-CTD algorithm decreases by 4.1% and 5.8% respectively, which is mainly because that the AED-CTD algorithm is based on event-driven cluster, which reduces the communication cost and energy consumption.

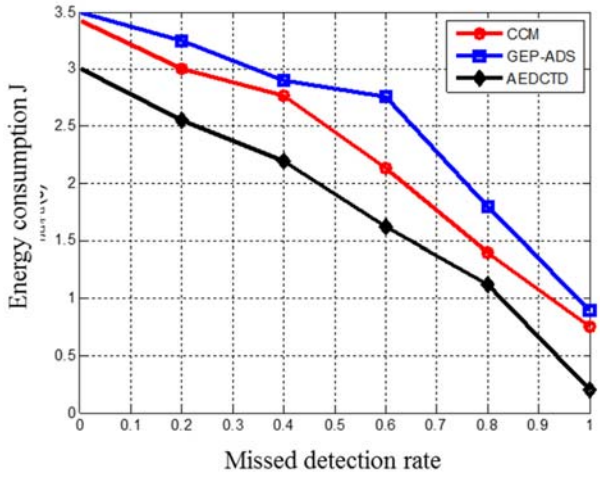


Fig. 6. Change curve of energy consumption with missed detection rate

## 5. Conclusion

Anomaly event detection is one of the applications of WSNs. To this end, this paper proposes the Anomaly Event-Driven Based Cluster Topology Detection (AED-CTD) algorithm. The AED-CTD algorithm establishes clusters at the event location. When the cluster is established, the energy of the sensor node and the detection probability of the sensor node to the event are considered. The experimental data show that the proposed AED-CTD algorithm reduces energy consumption and improves the detection rate of anomaly events.

## Acknowledgement

The industrialization and popularization of the family information platform of the Guangdong Ministry of education in 2013 (project number: 2013B090200055).

**References**


[1] L. R. Stephygraph, N. Arunkumar, V. Venkatraman: *Wireless mobile robot control through human machine interface using brain signals*, 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials, ICSTM 2015 - Proceedings, (2015), art. No. 7225484, 596–603.

[2] N. Arunkumar, V. S. Balaji, S. Ramesh, S. Natarajan, V. R. Likhita, S. Sundari: *Automatic detection of epileptic seizures using independent component analysis algorithm*, IEEE-International Conference on Advances in Engineering, Science and Management, ICAESM-2012, (2012), art. No. 6215903, 542–544.

[3] Y. Du, Y. Z. Chen, Y. Y. Zhuang, C. Zhu, F. J. Tang, J. Huang: *Probing Nanostrain via a Mechanically Designed Optical Fiber Interferometer*. IEEE Photonics Technology Letters, *29* (2017), 1348–1351.

[4] W. S. Pan, S. Z. Chen, Z. Y. Feng: *Automatic Clustering of Social Tag using Community Detection.* Applied Mathematics & Information Sciences, *7* (2013), No. 2, 675–681.

[5] Y. Y. Zhang, E. Mintzer, and K. E. Uhrich: *Synthesis and Characterization of PEGylated Bolaamphiphiles with Enhanced Retention in Liposomes*, Journal of Colloid and Interface Science, *482* (2016), 19–26.

[6] N. Arunkumar, K. M. Mohamed Sirajudeen: *Approximate Entropy based ayurvedic pulse diagnosis for diabetics - A case study*, TISC 2011 - Proceedings of the 3rd International Conference on Trendz in Information Sciences and Computing, (2011), art. No. 6169099, 133–135.

[7] Y. Y. Zhang, A. Algburi, N. Wang, V. Kholodovych, D. O. Oh, M. Chikindas, and K. E. Uhrich: *Self-assembled Cationic Amphiphiles as Antimicrobial Peptides Mimics: Role of Hydrophobicity, Linkage Type, and Assembly State*, Nanomedicine: Nanotechnology, Biology and Medicine, *13* (2017), No. 2, 343–352.

[8] N. Arunkumar, K. R. Kumar, V. Venkataraman: *Automatic detection of epileptic seizures using new entropy measures*,Journal of Medical Imaging and Health Informatics, *6* (2016), No. 3, 724–730.

[9] R. Hamza, K. Muhammad, N. Arunkumar, G. R. González: *Hash based Encryption for Keyframes of Diagnostic Hysteroscopy*, IEEE Access (2017), https://doi.org/10.1109/ACCESS.2017.2762405

[10] D. S. Abdelhamid, Y. Y. Zhang, D. R. Lewis, P. V. Moghe, W. J. Welsh, and K. E. Uhrich: *Tartaric Acid-based Amphiphilic Macromolecules with Ether Linkages Exhibit Enhanced Repression of Oxidized Low Density Lipoprotein Uptake*, Biomaterials, *53* (2015), 32–39.


Received May 7, 2017