Information Technology for Big Data Sensor Networks Stability Estimation

Volodymyr Petrivskyi\textsuperscript{a} \textsuperscript{i}, Georgi Dimitrov\textsuperscript{b} \textsuperscript{j}, Viktor Shevchenko\textsuperscript{a}, Oleksiy Bychkov\textsuperscript{a}, Magdalena Garvanova\textsuperscript{b}, Galina Panayotova\textsuperscript{b}, Pavel Petrov\textsuperscript{c}

\textsuperscript{a} Taras Shevchenko National University of Kyiv, Ukraine, https://pst.knu.ua/
\textsuperscript{b} University of Library Studies and Information Technologies, Sofia, Bulgaria, https://www.unibit.bg/
\textsuperscript{c} University of Economics, Varna, Bulgaria, https://www.ue-varna.bg/

Abstract:
Sensors and sensor networks are widely used in various industries and human activities and main components of the Internet of Things and Industrial Internet of Things concepts. Nowadays, huge amounts of data, called Big Data, can be transported and collected using sensor networks. This article presents approaches for providing sensor networks' stability estimation for transporting Big Data and collecting respective use cases. The proposed estimation method allows to detect sensor network's components that decrease data transport system stability. Consequently, additional connections or sensors can be added to increase stability. In the case of data collection, the solution consists of finding the most vulnerable sensor and an optimal position for the additional sensor with given intersection levels with other sensors. Simulation results confirm the feasibility and effectiveness of the proposed approaches.

Article Info:
Received: 07 June 2020
Revised: 08 Sep 2020
Online: 22 Sep 2020

Keywords:
sensors, sensor network, Big Data, sensor network stability, coverage radius, data loss

Creative Commons BY-NC 4.0
Introduction

According to the great number of advantages and opportunities sensors take a great part in all human activities nowadays. Main sensors’ advantages are their size, applications, low energy consumption, usability. Sensors are used for Big Data collecting and measurement tasks (temperature, pressure, flow rate, etc.), analysis, scanning, detecting. Also, sensors are parts and widely used in the Internet of Things and Industrial Internet of Things concepts. According to Internet of Things is the network of physical objects accessed through the Internet that contain embedded technology to interact with internal states or the external environment. The areas of sensor networks application are environmental monitoring, home intelligence, health care sphere, industrial processes control, agriculture sphere. A huge amount of data, called Big Data, can be transported and collected using sensor networks. Collected with sensors and sensor networks Big Data can further be used for processes modelling and prognosis. But like any network sensor networks are at risk. The main threats for sensors networks are connection loss, different types of attacks, trusted and unauthorized sensors, wireless communication channels. Pass loss modeling for wireless sensor networks presented in the paper. In the article authors proposed sensor networks active defense model based on the game theory. A dynamic clustering technique for maximizing wireless sensor networks stability presented in the paper. In papers coverage and connectivity techniques with deployment algorithms in wireless sensor networks are presented. Also, in the article authors distinguish sensing range and communication range of the sensor, but in our case, it is single range called “coverage radius”. Optimal sensor’s placement problem described in papers. It should also be taken into account that sensors can be static and movable. Sensors’ movability can also cause data and connection losses, decreasing coverage radiuses, signal’s deterioration. Therefore, sensor networks modelling and fault tolerance testing are a pressing issue in modern conditions. Results of the research can be used for detecting vulnerable sensor networks’ components and finding optimal placement for additional sensors with the given coverage zones intersection levels.

Sensor networks vulnerability estimation

Suppose we have a sensor network that consists of \( n \) sensors. It is obviously there are connections between sensors. Describe these connections using the appropriate matrix \( A \) with elements:

\[
a_{i,j} = \begin{cases} 
1, & \text{connection is available} \\
0, & \text{connection is unavailable} 
\end{cases}, \\
\]

where \( i = \overline{1, n}, j = \overline{1, n}, n \) – number of sensors.

Estimate the importance of each sensor network’s element and mark it \( c_i, i = \overline{1, n}, n \) – number of sensors in the network. Described estimations can get any value, but for the forward calculations all estimations must be normalized, that is \( 0 \leq c_i \leq 1 \). Determine the importance estimation of each connection
and form appropriate matrix $K$. Elements of this matrix will be calculated in the next form:

$$k_{i,j} = a_{i,j} \left( \frac{c_i + c_j}{2} \right),$$

(2)

where $i = 1, n, j = 1, n$, $n$ – number of sensors.

Let consider two main cases:

1. Data passing;
2. Data collecting.

Consider the first case. Suppose we have a sensor network that consists of $n$ sensors and performs data passing from the first sensor or source to the last sensor or target. Mark source sensor as $s_1$ and target sensor as $s_n$ (Figure 1).

**Figure 1: $s_1 ... s_n$ system example.**

Form the matrix of transitions between sensors, mark it $H$. Elements of this matrix will be equal:

$$h_{i,j} = \begin{cases} 
1, & \text{data pass from i sensor to j sensor} \\
0, & \text{data do not pass from i sensor to j sensor}
\end{cases},$$

(3)

Using Dijkstra’s algorithm and transition matrix $H$ let find the shortest path from $s_1$ to $s_n$. This way will be equal $s_{\text{min}} = \{s_1, ..., s_n\}$. In case when $s_{\text{min}} < n$ sensor network can be tested for resistance in failure of elements or loss of communication between sensors. Mark as $m$ number of sensors that are used in the shortest path $s_{\text{min}}$, therefore $m = |s_{\text{min}}|$. Using Dijkstra’s algorithm or Lee algorithm find solution of the pathfinding problem with transition matrix $H$ and transition weight matrix $K^\ast$. Elements of the matrix $K^\ast$ will be equal:

$$k_{i,j}^\ast = (1 - k_{i,j}),$$

(4)

where $i = 1, n, j = 1, n$, $k_{i,j}$ – elements of the matrix $K$ (2).

For Lee algorithm using graph that can be built using transition matrix $H$ must be planar. The solution of the described above problem will be vector $s^\ast$ with elements that have the smallest weight. In our case, using transformation (4),
these elements are sensors that have the biggest importance estimation for selected sensor network. Using $s^*$ form transition vector $h^*$ elements of which are transitions between elements of the $s^*$ vector. Select the first element $h^*_{1j}$ that is also certain element $h^*_{ij}$ in the transition matrix $H$ and replace $h^*_{ij}$ by zero. After replacement find again the shortest path from $s_1$ to $s_n$. In the case of the way existing value of the $h^*_{ij}$ in the $H$ matrix that is equal to zero should be replaced by its previous value and the next element from vector $h^*$ should be considered. Otherwise, case of the way from the source to the target absenting means transition $h^*_{ij}$ vulnerability. Find the ratio of the number of failures during the test $c_0$ to the number of transitions in the shortest path $s^*$. The obtained value means the probability of system failure in case of communication loss. Mark $c_h$ as a network’s transition loss estimation and calculate it as an inverted to communication loss probability value in the next form:

$$c_h = 1 - \frac{c_0}{m_{t}},$$

(5)

where $c_0$ – number of loses during the test, $m_t$ – number of transitions in the shortest path $s^*$.

For the next test select second element from the $s^*$ and replace by zero all values in the second row and column of the transition matrix $H$. After this replacement find the shortest path from the source $s_1$ to the target $s_2$. If there is a solution, return previous values of elements of the matrix $H$ and select next element from the first solution $s^*$. The first and the last elements of the $s^*$ are not considered as they are the source and the receiver and in case of their absence the data passing is impossible. In the absence of a solution after replacement, it should be concluded that it is necessary to ensure the transmission of information in the absence of this sensor. It can be adding new sensors, changing sensor’s importance, etc. After calculating number of elements $s^*$ at zero value of which the connection was lost, we estimate stability of the system in the sensor’s loss case in the same with the formula (5) way in the next form:

$$c_s = 1 - \frac{s_0}{m_{s} - 2},$$

(6)

where $s_0$ – number of loses during the test, $m_s$ – number of sensors in the shortest path $s^*$.

In the formula (6) $m - 2$ means that we don’t take into the account source and target sensors.

Total sensor network’s stability estimation means simultaneous stability to loss of transition and loss of the sensor. This value can be calculated as the probability of simultaneous occurrence of two independent events in the next way:

$$c = c_sc_h.$$

(7)

The value $c = 1$ means system stability with respect to the data transmission components.
Consider the second case. Suppose we have a set of sensors \( \{s_1, s_2, \ldots, s_n\} \) that are used for data collecting. Consider the same type of sensors and two-dimensional case. Same sensors type means same observation or accumulation data type for all sensors. Described above network can schematically be illustrated on the next figure (Figure 2):

![Figure 2: 4-Sensors collecting data system with coverage radiuses.](image)

Mark as \( \bar{S} \) square that is covered by sensors \( S = \{s_1, s_2, \ldots, s_n\} \). According to the article 23 total coverage area of sensors \( S_i \) and \( S_j \) can be calculated using next formula:

\[
S_{ij} = S_i + S_j - S_i \cap S_j,
\]  

(8)

where \( i = 1, n - 1, j = i + 1 \).

Total coverage square can be calculated using formula (8) for the pair of sensors \( S_i \) and \( S_j \) at the first step. At the next step mark area from the previous step as \( \bar{S}_i \) and calculate coverage area with the next \( S_j \) sensor (Figure 3).

![Figure 3: Formula (8) using steps.](image)

Exclude the first sensor \( s_1 \) from consideration and calculate square \( \bar{S} \) as described above. Mark calculated square as \( \bar{S}_1 \). Assume that data amount obtained by the sensor corresponds to the radius of its coverage. According to this assumption find accumulated data loss without sensor \( s_1 \) in the next way:

\[
l_1 = 1 - \frac{\bar{S}_1}{\bar{S}},
\]  

(9)

where \( \bar{S}_1 \) – square covered without sensor \( s_1 \), \( \bar{S} \) – total square.
Using described above approach calculate loss estimations for all sensors get corresponding set $L = \{l_1, l_2, ..., l_n\}$. Select the largest element among the elements of this set. This element means the biggest loss in case of this sensor’s absenting. To decrease such loss additional sensor can be included into the data collecting set. As a result, there is a problem of the correct location of the additional sensor. In the article information technology that allows to increase sensors’ operation time by regulating their coverage radiuses is presented. Described approach can be modified for finding sensor’s $s = s(x, y, r)$ location with maximum coverage radius and given intersection with another sensors level. Suppose we have a set of movable sensors $S = \{s_1, s_2, ..., s_n\}$, where each sensor has following parameters $s_i = s_i(x, y, r)$, where $x$ and $y$ are sensor’s coordinates, $r$ – coverage radius, $i = 1, n$. Suppose that sensor’s coverage radius is a constant value. Mark $c$ as sensors’ $\{s_1, s_2, ..., s_k\}$ minimum allowed intersection level with new sensor $s$. This statement can be described:

$$|r - r_j| = c,$$

where $j = 1, k$.

Schematically value $c$ can be illustrated on the next figure (Figure 4):

![Figure 4: Value c meaning.](image)

It is also necessary to require that the coverage area of the added sensor is in the area under consideration. Mark borders of the area as $x_{min}, x_{max}, y_{min}, y_{max}$, therefore constraints can be presented in the next form:

$$\begin{align*}
x_{min} + r & \leq x \leq x_{max} - r \\
y_{min} + r & \leq y \leq y_{max} - r.
\end{align*}$$

Let assume that coordinates and coverage radiuses are already known and have constant values. As follows, problem of the sensor’s $s = s(x, y, r)$ placement with given intersection level $c$ can be formulated in the next form:

$$\begin{align*}
\sqrt{(x - x_1)^2 + (y - y_1)^2} & = r_1 + r - c \\
\sqrt{(x - x_2)^2 + (y - y_2)^2} & = r_2 + r - c \\
& \vdots \\
\sqrt{(x - x_k)^2 + (y - y_k)^2} & = r_k + r - c \\
x_{min} + r & \leq x \leq x_{max} - r, y_{min} + r & \leq y \leq y_{max} - r,
\end{align*}$$

146
where \( k \) – the number of sensors according to which the new sensor should be positioned.

The difference \( r - c \) in the right part of the (12) can be modified in the case of variant intersection level for each sensor as follows:

\[
r - c_i,
\]

(14)

where \( i = \overline{1,k} \).

The solution of the (12)-(14) problem is pair of new sensor’s coordinates \((x^*, y^*)\). Obviously, problem (12)-(14) can have more than one solution. In such cases additional constraints should be taken into the account.

**Computer simulating results**

Computer simulation was performed with an assumption that the terrain is flat and there is no impact on sensing and communication. Consider sensor network that consists of 8 sensors and serves for data transmission. Connection matrix \( A \), sensors’ importance vector \( c \) and transition matrix \( H \) for this network have next values:

\[
A = \begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 1 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 1 & 0
\end{pmatrix},
\]

\[
c = (1, 0.8, 0.5, 1, 0.7, 0.5, 0.9, 1),
\]

\[
H = \begin{pmatrix}
0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0
\end{pmatrix}.
\]

Using formula (2) find transition weight matrix \( K \) and as a result with formula (4) transition weight matrix \( K^* \) will be equal:

\[
K = \begin{pmatrix}
0 & 0.9 & 0.75 & 0 & 0 & 0 & 0 & 0 \\
0.9 & 0 & 0 & 0.9 & 0 & 0 & 0 & 0 \\
0.75 & 0 & 0 & 0.75 & 0 & 0 & 0 & 0 \\
0 & 0.9 & 0.75 & 0 & 0.85 & 0.75 & 0.95 & 0 \\
0 & 0 & 0 & 0.85 & 0 & 0 & 0 & 0.85 \\
0 & 0 & 0 & 0.75 & 0 & 0 & 0 & 0.75 \\
0 & 0 & 0 & 0.95 & 0 & 0 & 0 & 0.95 \\
0 & 0 & 0 & 0 & 0.85 & 0.75 & 0.95 & 0
\end{pmatrix},
\]

147
Described with these matrixes sensor network with transition weights can schematically be illustrated on the next figure (Figure 5):

\[
K^* = \begin{pmatrix}
0 & 0.1 & 0.25 & 0 & 0 & 0 & 0 & 0 \\
0.1 & 0 & 0 & 0.1 & 0 & 0 & 0 & 0 \\
0.25 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0 \\
0 & 0.1 & 0.25 & 0 & 0.15 & 0.25 & 0.05 & 0 \\
0 & 0 & 0 & 0.15 & 0 & 0 & 0 & 0.15 \\
0 & 0 & 0 & 0.25 & 0 & 0 & 0 & 0.25 \\
0 & 0 & 0 & 0.05 & 0 & 0 & 0 & 0.05 \\
0 & 0 & 0 & 0 & 0.05 & 0.25 & 0.05 & 0 \\
\end{pmatrix}.
\]

Figure 5: Sensor network.

The shortest path from \(s_1\) to \(s_8\) according to transition matrix \(H\) and transition weight matrix \(K^*\) is \(s^* = \{s_1, s_2, s_4, s_7, s_8\}\) with transitions \(h^* = \{h_{1,2}, h_{2,4}, h_{4,7}, h_{7,8}\}\). Using vectors \(s^*\) and \(h^*\) results of the described above tests according to formulas (5)-(7) are:

\[
c_h = 1, \ c_s = 0.33, \ c = c_h c_s = 0.33.
\]

According to the test results, presented sensor network is not stable with respect to the data transmission sensors, especially \(s_4\) sensor. To increase network’s stability transition from \(s_2\) to \(s_5\) or from \(s_3\) to \(s_5\) should be added. With one of the proposed transitions estimation value \(c_s\) increases to 1 and as a result total estimation \(c\) also increases to 1 that means sensor network’s stability with respect to the data transmission components. With the proposed method using the stability of the given network was increased by 67%.

Consider collecting data sensor network that consists of 4 sensors with parameters that are presented in the Table 1. Make some assumptions that the terrain is flat, rectangular with size 100x100 and does not depend on the value measured by the sensors.
Table 1. Computer simulating input dataset.

<table>
<thead>
<tr>
<th>s</th>
<th>(x, y)</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>s_1</td>
<td>(20, 55)</td>
<td>19</td>
</tr>
<tr>
<td>s_2</td>
<td>(35, 40)</td>
<td>7</td>
</tr>
<tr>
<td>s_3</td>
<td>(50, 50)</td>
<td>13</td>
</tr>
<tr>
<td>s_4</td>
<td>(80, 45)</td>
<td>15</td>
</tr>
</tbody>
</table>

Sensors’ positions and covered areas can be illustrated on the next figure (Figure 6):

![Sensors with coverage radiuses.](image)

**Figure 6: Sensors with coverage radiuses.**

Total covered square using (8) is equal $\tilde{S} = 2446.8$. According to the formula (9) estimation loss vector is equal $L = \{0.44, 0.07, 0.21, 0.28\}$. So additional sensor should be added in case of the first sensor’s loss. Let find new position of the additional sensor with the same intersection levels as they are for $s_1$ $c_{1,2} = 5$ and $c_{1,3} = 1$ using (13)-(15):

$$\begin{cases} \sqrt{(x - 35)^2 + (y - 40)^2} = 16 \\ \sqrt{(x - 50)^2 + (y - 50)^2} = 30 \end{cases}$$

$$19 \leq x \leq 81, 19 \leq y \leq 81, x \neq 20, y \neq 55.$$

The solution of this system is $\begin{cases} x = 41.7 \\ y = 20.1 \end{cases}$. Additional sensor’s position with covered area can be illustrated on the next figure (Figure 7):
Figure 7: First additional sensor with coverage area (dashed line).

Also, according to the input dataset and Figure 7 last sensor is not intersecting with the other sensors. This fact means absence data transition between all sensors in the presented network. This problem can be solved with the additional sensor. Let find new position of the additional sensor with the coverage radius $r = 15$ and intersection levels $c_3 = c_4 = 1$ with taking into the account previous solution (Figure 7) using (13)-(15):

$$
\begin{align*}
\sqrt{(x - 50)^2 + (y - 50)^2} &= 26, \\
\sqrt{(x - 80)^2 + (y - 45)^2} &= 28,
\end{align*}

15 \leq x \leq 85, \quad 50 \leq y \leq 85.
$$

The solution of this system is $x = 66.9, \quad y = 69.7$. Additional sensor’s position with covered area can be illustrated on the next figure (Figure 8):

Figure 8: Second additional sensor with coverage area (dash-dot line).
With the additional sensor (dash-dot line) total coverage square increased by 28.7 % and is equal 3151.3.

Conclusions
Thus, in the article method for testing sensor networks in two use cases is presented. The method is considered in the case of operating large amounts of data. In the case of Big Data transport sensor network tests allow calculating estimations for connection loss and sensor loss situations. As a result, vulnerable components can be detected. For the Big Data collecting sensor network systems case test allows us to find the most valuable sensor that means the biggest loss without this sensor. Also, an approach that allows us to find an optimal additional sensor’s placement with a given intersection level is described. Sensor’s movement can also cause data and connection loss, so given approaches can be used for such situations. Computer simulation was performed with an assumption that the terrain is flat and there is no impact on sensing and communication. Terrain and the possibility of communication interference will be taken into account in further research. Computer simulating results which are presented confirm the feasibility and effectiveness of using the proposed approaches and appropriate software.

Acknowledgements
This work is supported by PPNIP-2020-03/09.03.2020 “Extraction, processing and analysis of parametric data from external devices” and BG05M2OP001-1.001-0003 “Center of Excellence in Informatics and Information and Communication technologies.”

References


About the Authors

Volodymyr Petrivskyi, PhD student of Software System and Technologies Department of Taras Shevchenko National University of Kyiv. Master of computer science and information technologies. Research interests: computer science, mathematical modelling, Internet of Things, sensors, sensor networks, software development.
Georgi P. Dimitrov, Deputy Dean of the Faculty "Information Science" in the University of Library Studies and Information Technologies (ULSIT), Sofia, Bulgaria. He is the author/co-author of over 130 scientific publications. His research interests involve Information and Communication Technologies, Data Science, Brain Wave, Cyber Security, Big Data, Deep Learning and etc.

Viktor Shevchenko, Dr.Sc., Professor of Software System and Technologies Department of Taras Shevchenko National University of Kyiv. Was a Director of Educational Scientific Institute of Defense of Information in State University of Telecommunication; Head of National Scientific Research Center for Defense Technologies and Military Security of Ukraine; Head of scientific support of the implementation of a number of information-analytical systems of defense planning and management of defense resources (6 of which have been adopted in Armed Forces). Research interests: Internet of Things, Internet of Everything, mathematical modeling, cybersecurity, sensors.

Oleksiy Bychkov, Dr.Sc., Associated Professor, Head of Software System and Technologies Department of Taras Shevchenko National University of Kyiv. Proposed a new algorithm for studying the stability of stationary states of hybrid automata. Solved an open problem about the necessary and sufficient conditions for the existence of a positive matrix, which is the solution of the matrix Lyapunov equation on a cone. Research interests: Internet of Things, Internet of Everything, mathematical modeling of physics of natural phenomena, computer science.

Magdalena Garvanova, Associate professor at the Department of Public Communications in the University of Library Studies and Information Technologies, Sofia, Bulgaria. She is an author of more than 120 scientific publications in Bulgarian, English, and Russian, including 4 monographs and 4 textbooks. Dr. Garvanova has extensive experience in conducting various international and national marketing, social and scientific studies, related to survey design, data collection, data processing and data analysis.

Galina Panayotova, Doctor of Mathematics and Professor of Mathematical modeling in the State University of Library Studies and Informational technologies and University “Prof. Dr. As. Zlatarov” – Burgas. Research interests - Nets and tissues in Finsler and Riemannian spaces; Application of differential geometry in the theory of PDE; Use of information technologies and applications in education; Mathematical and Computer Modeling. Prof. Panayotova is the author of more 80 scientific publications, books and textbooks.

Pavel Petrov, PhD, Associate Professor, Department of Informatics in University of Economics - Varna, Bulgaria. Proposed a new approach for building complex information systems for management of manufacturing enterprises. Research interests: real time web technologies, big data, banks online technologies, information security, informatics and computer science.