

CLAIMS

What is claimed is:

1. A method comprising:

identifying, by one or more processors, data sources, the data sources being configured to provide sensor data and timestamping of the sensor data as a first set of one or more time series, the sensor data comprising values of a first group of measurands;

determining, by one or more processors, that values of a second group of one or more of the measurands of a subset of the sensor data indicates an anomaly, wherein the subset of sensor data is provided by a given data source of the data sources and covering a time window, the given data source comprises a set of components;

determining, by one or more processors, a third group of one or more of the measurands that are root cause candidates of the anomaly using a set of one or more similarity techniques for comparing values of the second group of measurands and the third group of measurands in the time window, wherein the measurands of the third group are provided by the set of components;

for each similarity technique of the set of similarity techniques and for each measurand of the second group, assigning, by one or more processors, a set of coefficients to the measurand, wherein each coefficient of the set of coefficients is indicative of a comparison result of the each measurand with a measurand of the third group using the similarity technique;

determining, by one or more processors, using the sets of coefficients, whether a specific subset of the set of components of the given data source can be identified as a root cause of the anomaly; and

in response to determining that the specific subset of the set of components of the given data source can be identified as the root cause of the anomaly, providing, by one or more processors, the specific subset of components as the root cause of the anomaly.

2. The method of claim 1, further comprising:

in response to determining that no specific subset of the set of components of the given data source can be identified as a root cause of the anomaly, updating, by one or more processors, the third group of measurands;

for each similarity technique of the set of similarity techniques and for each measurand of the second group, assigning, by one or more processors, an updated set of coefficients to the measurand, wherein each coefficient of the updated set of coefficients is indicative of the comparison result of the each measurand with a measurand of the updated third group using the similarity technique; and

determining, by one or more processors, using the updated sets of coefficients, whether the specific subset of the set of components of the given data source can be identified as the root cause of the anomaly.

3. The method of claim 2, wherein the updating of the third group of measurands further comprises an action selected from the group consisting of:

removing, by one or more processors, one or more measurands from the third group;
and

adding, by one or more processors, one or more measurands to the third group.

4. The method of claim 3, wherein the removed or added measurands are provided by a selected component of the set of components, the method further comprising:

excluding, by one or more processors, the selected component from the set of components for a next iteration.

5. The method of claim 4, further comprising:

ranking, by one or more processors, the set of components in accordance with a predefined property of the set of components and in a monotonic sorting order, wherein the selected component is a first ranked component.

6. The method of claim 5, wherein the property is an overload degree.

7. The method of claim 1, wherein the specific subset of components is a single component of the set of components.

8. The method of claim 1, wherein the determining whether a specific subset of the set of components of the given data source can be identified as a root cause of the anomaly further comprises:

for each measurand of the second group and for each measurand of the third group, combining, by one or more processors, the respective set of coefficients, resulting in a combined coefficient; and

determining whether the specific subset of the set of components of the given data source can be identified as a root cause of the anomaly using the combined coefficients.

9. The method of claim 8, wherein combining the set of coefficients includes summing the set of coefficients.

10. The method of claim 1, further comprising:

normalizing, by one or more processors, in the time window, values of measurands of the second and third groups of measurands prior to comparing the values of the second group of measurands and the third group of measurands in the time window.

11. The method of claim 1, wherein the set of similarity techniques includes a Manhattan distance, Euclidean distance, Dynamic time warping (DTW) distance, Spearman and Pearson metrics.

12. The method of claim 1, wherein determining that values of a second group of one or more of the measurands of a subset of the sensor data indicates an anomaly further comprises:

receiving, by one or more processors, an event ticket from the data source, the event ticket indicative of the anomaly.

13. The method of claim 1, wherein determining that values of a second group of one or more of the measurands of a subset of the sensor data indicates an anomaly initiates in response to receiving an event ticket.

14. The method of claim 1, wherein a measurand of the second group comprises a measurand selected from the group consisting of (i) a measurand of the first group of measurands and (ii) a combination of measurands of the first group.

15. The method of claim 1, wherein each time series of the first set of time series includes values of a respective measurand.

16. A computer system comprising:
- one or more computer processors;
 - one or more computer readable storage media; and
 - program instructions stored on the computer readable storage media for execution by at least one of the one or more processors, the program instructions comprising:
 - program instructions to identify data sources, the data sources being configured to provide sensor data and timestamping of the sensor data as a first set of one or more time series, the sensor data comprising values of a first group of measurands;
 - program instructions to determine that values of a second group of one or more of the measurands of a subset of the sensor data indicates an anomaly, wherein the subset of sensor data is provided by a given data source of the data sources and covering a time window, the given data source comprises a set of components;
 - program instructions to determine a third group of one or more of the measurands that are root cause candidates of the anomaly using a set of one or more similarity techniques for comparing values of the second group of measurands and the third group of measurands in the time window, wherein the measurands of the third group are provided by the set of components;
 - for each similarity technique of the set of similarity techniques and for each measurand of the second group, program instructions to assign a set of coefficients to the measurand, wherein each coefficient of the set of coefficients is indicative of a comparison result of the each measurand with a measurand of the third group using the similarity technique;
 - program instructions to determine, using the sets of coefficients, whether a specific subset of the set of components of the given data source can be identified as a root cause of the anomaly; and
 - in response to determining that the specific subset of the set of components of the given data source can be identified as the root cause of the anomaly, program instructions to provide the specific subset of components as the root cause of the anomaly.

17. The computer system of claim 16, further comprising program instructions, stored on the computer readable storage media for execution by at least one of the one or more processors, to:

in response to determining that no specific subset of the set of components of the given data source can be identified as a root cause of the anomaly, update third group of measurands;

for each similarity technique of the set of similarity techniques and for each measurand of the second group, assign an updated set of coefficients to the measurand, wherein each coefficient of the updated set of coefficients is indicative of the comparison result of the each measurand with a measurand of the updated third group using the similarity technique; and

determine using the updated sets of coefficients, whether the specific subset of the set of components of the given data source can be identified as the root cause of the anomaly.

18. The computer system of claim 16, wherein the specific subset of components is a single component of the set of components.

19. A computer program product comprising:

one or more computer readable storage media and program instructions stored on the one or more computer readable storage media, the program instructions comprising:

to program instructions to identify data sources, the data sources being configured to provide sensor data and timestamping of the sensor data as a first set of one or more time series, the sensor data comprising values of a first group of measurands;

program instructions to determine that values of a second group of one or more of the measurands of a subset of the sensor data indicates an anomaly, wherein the subset of sensor data is provided by a given data source of the data sources and covering a time window, the given data source comprises a set of components;

program instructions to determine a third group of one or more of the measurands that are root cause candidates of the anomaly using a set of one or more similarity techniques for comparing values of the second group of measurands and the third group of measurands in the time window, wherein the measurands of the third group are provided by the set of components;

for each similarity technique of the set of similarity techniques and for each measurand of the second group, program instructions to assign a set of coefficients to the measurand, wherein each coefficient of the set of coefficients is indicative of a comparison result of the each measurand with a measurand of the third group using the similarity technique;

program instructions to determine, using the sets of coefficients, whether a specific subset of the set of components of the given data source can be identified as a root cause of the anomaly; and

in response to determining that the specific subset of the set of components of the given data source can be identified as the root cause of the anomaly, program instructions to provide the specific subset of components as the root cause of the anomaly.

20. The computer program product of claim 19, further comprising program instructions, stored on the one or more computer readable storage media, to:

in response to determining that no specific subset of the set of components of the given data source can be identified as a root cause of the anomaly, update third group of measurands;

for each similarity technique of the set of similarity techniques and for each measurand of the second group, assign an updated set of coefficients to the measurand, wherein each coefficient of the updated set of coefficients is indicative of the comparison result of the each measurand with a measurand of the updated third group using the similarity technique; and

determine using the updated sets of coefficients, whether the specific subset of the set of components of the given data source can be identified as the root cause of the anomaly.