

CLAIMS:

1. A method for quantitative unified analysis of unstructured threat intelligence data, the method to be performed by a computer system comprising:
 - receiving unstructured threat intelligence data from molecular level sources, wherein the unstructured threat intelligence data comprises security events generated from the molecular level sources; uniting the received unstructured threat intelligence data by translating the received threat intelligence data into a uniformed meta-format and storing the translated data in a database, wherein each translated data includes at least a time component;
 - determining isolated data points from the translated data stored in the database by grouping the translated data into groups, wherein the determining comprises:
 - determining a first group of translated data from the translated data where data in the first group has time components contained within a first time period into a first group,
 - determining at least one cluster of data from the translated data in the first group using unsupervised machine learning techniques,
 - identifying at least one sparse cluster from the at least one clusters having a total number of data points below a predefined threshold, and
 - categorizing the data in each of the at least one sparse clusters, identified as having the total number of data points below the predefined threshold, as isolated data points;
 - determining data outliers from the isolated data points;
 - generating a learned function using the data outliers; and
 - providing the generated learned function to at least one of a plurality of upstream sources to update a security rule of the at least one of the plurality of upstream sources using the provided generated learned function, wherein the updated security rule set is used to filter threats to the at least one of the plurality of upstream sources.
2. The method according to claim 1 wherein the unstructured threat intelligence data further comprises triggers generated from atomic level sources, wherein the security events generated from the molecular sources are based on the triggers generated from atomic level sources.
3. The method according to claim 1 wherein the determining the isolated data points from the translated data stored in the database comprises:
 - grouping the translated data having time components contained within a first time period into a first group and grouping the translated data having time components contained

within a second time period into a second group, wherein the second time period comprises a period of time that is longer than the first time period;

clustering the translated data in the first and second groups using unsupervised machine learning techniques;

plotting the clusters and identifying clusters having a total number of data points below a predefined threshold; and

categorizing the data in the identified clusters as isolated data points.

4. The method according to claim 1 wherein the step of determining data outliers from the isolated data points comprises: generating a fixed sized window;

identifying isolated data points that may be concurrently located within the fixed sized window; and

classifying the identified data points as data outliers.

5. The method according to claim 1 wherein the generating the learned function comprises:

generating training sequences using the data outliers; and

training supervised machine learning algorithms using the generated training sequences, wherein the learned function comprises at least one trained supervised machine learning algorithm.

6. The method according to any of claim 1 wherein the unsupervised machine learning techniques comprises mining metadata from the normalized data.

7. The method according to any one of claim 1 wherein the unsupervised machine learning techniques comprises identifying relationships in data of the normalised data by identifying semantic distances between the normalised data.

8. The method according to any one of claim 1 wherein the unsupervised machine learning techniques comprises using statistical data to determine a relationship between the normalised data.

9. A computer system for quantitative unified analysis of unstructured threat intelligence data, the method to be performed by a computer system comprising:

circuitry configured to receive unstructured threat intelligence data from molecular level sources, wherein the unstructured threat intelligence data comprises security events generated from the molecular level sources; circuitry configured to unite the received

unstructured threat intelligence data by translating the received threat intelligence data into a uniformed meta-format and storing the translated data in a database, wherein each translated data includes at least a time component;

circuitry configured to determine isolated data points from the translated data stored in the database by grouping the translated data into groups, the circuitry configured to determine comprises:

circuitry configured to determine a first group of translated data from the translated data where data in the first group has time components contained within a first time period into a first group,

circuitry configured to determine at least one cluster of data from the translated data in the first group using unsupervised machine learning techniques,

circuitry configured to identify at least one sparse cluster from the at least one clusters having a total number of data points below a predefined threshold, and

circuitry configured to categorize the data in each of the at least one sparse clusters, identified as having the total number of datapoints below the predefined threshold, as isolated data points;

circuitry configured to determine data outliers from the isolated data points;

circuitry configured to generate a learned function using the data outliers; and

circuitry configured to providing the generated learned function to at least one of a plurality of upstream sources to update a security rule of the at least one of the plurality of upstream sources using the provided generated learned function, wherein the updated security rule set is used to filter threats to the at least one of the plurality of upstream sources.

10. The computer system according to claim 9 wherein the unstructured threat intelligence data further comprises triggers generated from atomic level sources, wherein the security events generated from the molecular sources are based on the triggers generated from atomic level sources.

11. The computer system according to any one of claim 9 wherein the circuitry configured to determine the isolated data points from the translated data stored in the database comprises:

circuitry configured to group the translated data having time components contained within a first time period into a first group and to group the translated data having time components contained within a second time period into a second group, wherein the second time period comprises a period of time that is longer than the first time period;

circuitry configured to cluster the translated data in the first and second groups using unsupervised machine learning techniques;

circuitry configured to plot the clusters and identify clusters having a total number of data points below a predefined threshold; and

circuitry configured to categorize the data in the identified clusters as isolated data points.

12. The computer system according to claim 9 wherein the circuitry configured to determine data outliers from the isolated data points comprises:

circuitry configured to generate a fixed sized window;

circuitry configured to identify isolated data points that may be concurrently located within the fixed sized window; and

circuitry configured to classify the identified data points as data outliers.

13. The computer system according to claim 9 wherein the circuitry configured to generate the learned function comprises:

circuitry configured to generate training sequences using the data outliers; and

circuitry configured to train supervised machine learning algorithms using the generated training sequences, wherein the learned function comprises at least one trained supervised machine learning algorithm.

14. The computer system according to claim 9 wherein the unsupervised machine learning techniques comprises mining metadata from the normalised data.

15. The computer system according to claim 9 wherein the unsupervised machine learning techniques comprises identifying relationships in data of the normalised data by identifying semantic distances between the normalised data.

16. The computer system according to claim 9 wherein the unsupervised machine learning techniques comprises using statistical data to determine a relationship between the normalised data.