

CLAIMS

1. A server computer system comprising at least one hardware processor configured to:
- assign events of a training corpus to a plurality of event categories, the training corpus comprising a collection of events having occurred on a plurality of client systems; in response to assigning events to event categories, assign client systems of the plurality of client systems to a plurality of client clusters according to the plurality of event categories; and
- in response to assigning client systems to client clusters, transmit a client cluster membership indicator to an anomaly detector configured to determine whether a target event occurring on a target client system is indicative of a computer security threat;
- wherein assigning events to event categories comprises:
- selecting a plurality of events from the training corpus, the plurality of events having occurred on a client system of the plurality of client systems, arranging the plurality of events according to a time of occurrence to form an event sequence, and
- in response, assigning a selected event of the event sequence to a selected event category according to a first event preceding the selected event and further according to second event following the selected event within the event sequence;
- wherein assigning client systems to client clusters comprises assigning the client system to a selected client cluster according to an event profile determined according to a count of events occurring on the client system and belonging to the selected event category; and
- wherein the anomaly detector is configured to determine whether the target event is indicative of the computer security threat according to a behavior model trained on a client-cluster-specific sub-corpus of events, the client-cluster-specific sub-

corpus selected from the training corpus to include only events that have occurred on a plurality of members of a target cluster of the plurality of client clusters.

2. The computer system of claim 1, wherein assigning events to event categories comprises:

determining a position of the selected event in a multidimensional event embedding space according to the first and second events; and
assigning the selected event to the selected event category according to the position of the selected event in the event embedding space.

3. The computer system of claim 2, further configured to determine the selected event category according to a cluster of events occupying similar positions in the event embedding space.

4. The computer system of claim 2, wherein determining the position of the selected event in the event embedding space comprises training an event encoder to produce a set of coordinates indicating the position, and wherein training the event encoder comprises:
coupling the event encoder to an event decoder configured to receive the set of coordinates and to output a prediction indicator indicative of a likelihood that the event sequence comprises the first and second events; and
adjusting a set of parameters of the event encoder according to the prediction indicator.

5. The computer system of claim 2, wherein determining the position of the selected event in the event embedding space comprises training an event encoder to produce a set of coordinates indicating the position, and wherein training the event encoder comprises:

5 coupling the event encoder to an event decoder configured to receive the
6 set of coordinates and to output a predicted context of the selected
7 event, the predicted context comprising a first predicted event and
8 a second predicted event;
9 comparing the first predicted event to the first event and the second
10 predicted event to the second event; and
11 adjusting a set of parameters of the event encoder according to a result of
12 the comparison.
13

1 6. The computer system of claim 1, wherein the event profile comprises a plurality of
2 components, each component determined according to a proportion of events
3 belonging to each event category of the plurality of event categories, the
4 proportion calculated out of a total count of events of the training corpus
5 occurring on the client system.
6

1 7. The computer system of claim 1, wherein assigning client systems to client clusters
2 comprises assigning client systems having similar event profiles to the same client
3 cluster.
4

1 8. The computer system of claim 1, wherein the event of the training corpus comprises a
2 launch of a selected process on a client system of the plurality of client systems.
3

1 9. A computer-implemented method comprising:
2 employing at least one hardware processor of a computer system to assign events of a
3 training corpus to a plurality of event categories, wherein the training corpus
4 comprises a collection of events occurring on a plurality of client systems;
5 in response to assigning events to event categories, employing at least one hardware
6 processor of the computer system to assign client systems of the plurality of client

7 systems to a plurality of client clusters according to the plurality of event
8 categories; and

9 in response to assigning client systems to client clusters, employing at least one hardware
10 processor of the computer system to transmit a client cluster membership
11 indicator to an anomaly detector configured to determine whether a target event
12 occurring on a target client system is indicative of a computer security threat;

13 wherein assigning events to event categories comprises:

14 selecting a plurality of events from the training corpus, the plurality of events
15 having occurred on a client system of the plurality of client systems,
16 arranging the plurality of events according to a time of occurrence to form an
17 event sequence, and

18 in response, assigning a selected event of the event sequence to a selected event
19 category according to a first event preceding the selected event and further
20 according to second event following the selected event within the event
21 sequence;

22 wherein assigning client systems to client clusters comprises assigning the client system
23 to a selected client cluster according to an event profile determined according to a
24 count of events occurring on the client system and belonging to the selected event
25 category; and

26 wherein the anomaly detector is configured to determine whether the target event is
27 indicative of the computer security threat according to a behavior model trained
28 on a client-cluster-specific sub-corpus of events, the client-cluster-specific sub-
29 corpus selected from the training corpus to include only events that have occurred
30 on a plurality of members of a target cluster of the plurality of client clusters.

31
1 10. The method of claim 9, wherein assigning events to event categories comprises:

2 determining a position of the selected event in a multidimensional event
3 embedding space according to the first and second events; and

4 assigning the selected event to the selected event category according to the
5 position of the selected event in the event embedding space.

6
1 11. The method of claim 10, further comprising determining the selected event
2 category according to a cluster of events occupying similar positions in the
3 event embedding space.

4
1 12. The method of claim 10, wherein determining the position of the selected
2 event in the event embedding space comprises training an event encoder to
3 produce a set of coordinates indicating the position, and wherein training
4 the event encoder comprises:

5 coupling the event encoder to an event decoder configured to receive the
6 set of coordinates and to output a prediction indicator indicative of
7 a likelihood that the event sequence comprises the first and second
8 events; and

9 adjusting a set of parameters of the event encoder according to the
10 prediction indicator.

11
1 13. The method of claim 10, wherein determining the position of the selected
2 event in the event embedding space comprises training an event encoder to
3 produce a set of coordinates indicating the position, and wherein training
4 the event encoder comprises:

5 coupling the event encoder to an event decoder configured to receive the
6 set of coordinates and to output a predicted context of the selected
7 event, the predicted context comprising a first predicted event and
8 a second predicted event;

9 comparing the first predicted event to the first event and the second
10 predicted event to the second event; and

adjusting a set of parameters of the event encoder according to a result of the comparison.

14. The method of claim 9, wherein the event profile comprises a plurality of components, each component determined according to a proportion of events belonging to each event category of the plurality of event categories, the proportion calculated out of a total count of events of the training corpus occurring on the client system.

15. The method of claim 9, wherein assigning the plurality of client systems to client clusters comprises assigning client systems having similar event profiles to the same client cluster.

16. The method of claim 9, wherein the event of the training corpus comprises a launch of a selected process on a client system of the plurality of client systems.

17. A non-transitory computer-readable medium storing instructions which, when executed by at least one hardware processor of a computer system, cause the computer system to: assign events of a training corpus to a plurality of event categories, the training corpus comprising a collection of events having occurred on a plurality of client systems; in response to assigning events to event categories, assign client systems of the plurality of client systems to a plurality of client clusters according to the plurality of event categories; and in response to assigning the client systems to client clusters, transmit a client cluster membership indicator to an anomaly detector configured to determine whether a target event occurring on a target client system is indicative of a computer security threat; wherein assigning events to event categories comprises:

13 selecting a plurality of events from the training corpus, the plurality of events
14 having occurred on a client system of the plurality of client systems,
15 arranging the plurality of events according to a time of occurrence to form an
16 event sequence, and

17 in response, assigning a selected event of the event sequence to a selected event
18 category according to a first event preceding the selected event and further
19 according to second event following the selected event within the event
20 sequence;

21 wherein assigning client systems to client clusters comprises assigning the client system
22 to a selected client cluster according to an event profile determined according to a
23 count of events occurring on the client system and belonging to the selected event
24 category; and

25 wherein the anomaly detector is configured to determine whether the target event is
26 indicative of the computer security threat according to a behavior model trained
27 on a client-cluster-specific sub-corpus of events, the client-cluster-specific sub-
28 corpus selected from the training corpus to include only events that have occurred
29 on a plurality of members of a target cluster of the plurality of client clusters.
30

CLAIMS

1. A server computer system comprising at least one hardware processor configured to execute an event encoder, an event clustering engine and a client clustering engine, wherein:

the event encoder is trained to determine a position of each event of a training corpus in an event embedding space, the training corpus comprising a collection of events having occurred on a plurality of client systems;

the event clustering engine is configured to distribute the training corpus among a plurality of event clusters according to the position of each event in the event embedding space; and

the client clustering engine is configured to:

distribute the plurality of client systems among a plurality of client clusters according to a content of each event cluster of the plurality of event clusters, and

in response to distributing client systems to client clusters, transmit a client cluster membership indicator to an anomaly detector configured to determine whether a target event occurring on a target client system is indicative of a computer security threat according to a subset of the training corpus consisting of events having occurred on members of a target client cluster selected from the plurality of client clusters to include the target client system;

wherein the event encoder is configured to determine a position of a selected event according to an event sequence comprising the selected event, a first event, and a second event of the training corpus, the first event having occurred earlier than the selected event, and the second event having occurred later than the selected event; and

wherein training the event encoder comprises:

coupling the event encoder to an event decoder configured to receive an indicator of the position of the selected event and to output a prediction indicator indicative of a likelihood that the event sequence comprises the first and second events, and

adjusting a set of parameters of the event encoder according to the prediction indicator.

2. The computer system of claim 1, wherein training the event encoder further comprises:

employing the event decoder to output a first predicted event and a second predicted event according to the position of the selected event;

comparing the first predicted event to the first event and the second predicted event to the second event; and

adjusting the set of parameters of the event encoder according to a result of the comparison.

3. The computer system of claim 1, wherein the client clustering engine is configured to assign a selected client system to a selected client cluster according to an event profile of the selected client system, the event profile comprising a plurality of components, each distinct component determined according to a count of events having occurred on the selected client system and belonging to a distinct event cluster of the plurality of event clusters.

4. The computer system of claim 3, wherein distributing the plurality of client systems among client clusters comprises assigning client systems having similar event profiles to the same client cluster.

5. The computer system of claim 1, wherein an event of the training corpus comprises a launch of a selected process on a client system of the plurality of client systems.

6. A computer-implemented method comprising employing at least one hardware processor of a computer system to:

execute an event encoder trained to determine a position of each event in an event embedding space, the training corpus comprising a collection of events having occurred on a plurality of client systems;

distribute the training corpus among a plurality of event clusters according to the position of each event in the event embedding space;

in response to distributing events to event clusters, distribute the plurality of client systems among a plurality of client clusters according to a content of each event cluster of the plurality of event clusters; and

in response to distributing client systems to client clusters, transmit a client cluster membership indicator to an anomaly detector configured to determine whether a target event occurring on a target client system is indicative of a computer security threat according to a subset of the training corpus consisting of events having occurred on members of a target client cluster selected from the plurality of client clusters to include the target client system;

wherein the event encoder is configured to determine a position of a selected event according to an event sequence comprising the selected event, a first event, and a second event of the training corpus, the first event having occurred earlier than the selected event, and the second event having occurred later than the selected event; and

wherein training the event encoder comprises:

coupling the event encoder to an event decoder configured to receive an indicator of the position of the selected event and to output a prediction indicator indicative of a likelihood that the event sequence comprises the first and second events, and

adjusting a set of parameters of the event encoder according to the prediction indicator.

7. The method of claim 6, wherein training the event encoder further comprises:

employing the event decoder to output a first predicted event and a second predicted event according to the position of the selected event;

comparing the first predicted event to the first event and the second predicted event to the second event; and

adjusting the set of parameters of the event encoder according to a result of the comparison.

8. The method of claim 6, comprising assigning a selected client system to the selected client cluster according to an event profile of the client system, the event profile comprising a plurality of components, each distinct component determined according to a count of events having occurred on the selected client system and belonging to a distinct event cluster of the plurality of event clusters.

9. The method of claim 8, wherein distributing the plurality of client systems among client clusters comprises assigning client systems having similar event profiles to the same client cluster.

10. The method of claim 6, wherein an event of the training corpus comprises a launch of a selected process on a client system of the plurality of client systems.

11. A non-transitory computer-readable medium storing instructions which, when executed by at least one hardware processor of a computer system, cause the computer system to execute an event encoder, an event clustering engine and a client clustering engine, wherein:

the event encoder is trained to determine a position of each event of a training corpus in an event embedding space, the training corpus comprising a collection of events having occurred on a plurality of client systems;

the event clustering engine is configured to distribute the training corpus among a plurality of event clusters according to the position of each event in the event embedding space; and

the client clustering engine is configured to:

distribute the plurality of client systems among a plurality of client clusters according to a content of each event cluster of the plurality of event clusters; and

in response to distributing client systems to client clusters, transmit a client cluster membership indicator to an anomaly detector configured to determine whether a target event occurring on a target client system is indicative of a computer security threat according to a subset of the training corpus consisting of events having occurred on members of a target client cluster selected from the plurality of client clusters to include the target client system;

wherein the event encoder is configured to determine a position of a selected event according to an event sequence comprising the selected event, a first event, and a second event of the training corpus, the first event having occurred earlier than the selected event, and the second event having occurred later than the selected event; and

wherein training the event encoder comprises:

coupling the event encoder to an event decoder configured to receive an indicator of the position of the selected event and to output a prediction indicator indicative of a likelihood that the event sequence comprises the first and second events, and

adjusting a set of parameters of the event encoder according to the prediction indicator.