

Claims

1. A method for detecting a malicious file, comprising:

obtaining a to-be-detected file;

parsing the to-be-detected file, to determine binary values corresponding to the to-be-detected file and a high-level definition corresponding to each of the binary values, wherein the high-level definition is referred to an additional attribute of the binary value; and

inputting the binary values corresponding to the to-be-detected file and the high-level definition corresponding to each of the binary values into a neural network, and detecting the to-be-detected file by means of the neural network, wherein the neural network is trained with a malicious file collection in advance;

wherein the parsing the to-be-detected file, to determine the binary values corresponding to the to-be-detected file and the high-level definition corresponding to each of the binary values, comprises:

parsing the to-be-detected file, to determine the binary values corresponding to the to-be-detected file as a first layer and to determine the high-level definition corresponding to the binary value as a second layer; and

combining the first layer and the second layer to generate a multi-layer image of the to-be-detected file, wherein the second layer comprises at least one layer, and each layer corresponds to each type of high-level definition.

2. The method for detecting a malicious file according to claim 1, wherein the high-level definition corresponding to each of the binary values comprises at least one of the following high-level definitions:

a high-level definition for representing a source of a binary value;

a high-level definition for representing a specific meaning of a binary value in an instruction; and

a high-level definition for representing a relationship of the binary value in the instruction in an execution flow.

~~3. The method for detecting a malicious file according to claim 1, wherein the parsing the to-be-detected file, to determine binary values corresponding to the to-be-detected file and a high-level definition corresponding to each of the binary values, comprises:~~

~~parsing the to-be-detected file, to determine the binary values corresponding to the to-be-detected file as a first layer and to determine the high-level definition corresponding to the binary value as a second layer; and~~

~~combining the first layer and the second layer to generate a multi-layer image of the to-be-detected file, wherein the second layer comprises at least one layer, and each layer corresponds to each type of high-level definition.~~

43. The method for detecting a malicious file according to claim 31, wherein the second layer comprises at least one of the following layers:

a layer for representing a source of a byte;

a layer for representing a specific meaning of a byte in an instruction;

a layer for representing a relationship of the byte in the instruction in an execution flow; and

a layer for representing a read-write permission of the byte.

54. The method for detecting a malicious file according to claim 43, wherein the layer for representing the source of the byte comprises:

an instruction layer, a data layer, a PE header structure layer, and a resource section layer.

65. The method for detecting a malicious file according to claim 43, wherein the layer for representing the specific meaning of the byte in the instruction comprises:

an operation code layer, an operand layer and a shift length layer.

76. The method for detecting a malicious file according to claim 43, wherein the layer for representing the relationship of the byte in the instruction in the execution flow comprises:

a target instruction address layer, a function range layer, and a node range layer of a control flow chart.

87. The method for detecting a malicious file according to claim 31, wherein the inputting the binary values corresponding to the to-be-detected file and the high-level definition corresponding to each of the binary values into a neural network, and detecting the to-be-detected file by means of the neural network comprises:

reducing a dimension of the multi-layer image, to generate dimension-reduced data; and

inputting the dimension-reduced data into the neural network, and detecting the to-be-detected file by means of the neural network.

98. The method for detecting a malicious file according to claim 87, wherein the reducing a dimension of the multi-layer image to generate dimension-reduced data comprises:

reducing the dimension of the multi-layer image by using an unsupervised learning network and/or a supervised learning network, to generate the dimension-reduced data.

109. An apparatus for detecting a malicious file comprising:

an acquisition module, configured to obtain a to-be-detected file;

a determination module, configured to parse the to-be-detected file, to determine binary values corresponding to the to-be-detected file and a high-level definition corresponding to each of the binary values, wherein the high-level definition is referred to an additional attribute of the binary value; and

a detection module, configured to input the binary values corresponding to the to-be-detected file and the high-level definition corresponding to each of the binary values into a neural network, and detect the to-be-detected file by means of the neural network, wherein the neural network is trained with a malicious file collection in advance;

wherein the determination module comprises:

a layer determination unit, configured to parse the to-be-detected file, to determine the binary values corresponding to the to-be-detected file as a first layer and to determine the high-level definition corresponding to the binary value as a second layer, wherein the second layer comprises at least one layer, and each layer corresponds to each type of high-level definition; and

a multi-layer image determination unit, configured to combine the first layer and the second layer, to generate a multi-layer image of the to-be-detected file.

4410. The apparatus for detecting a malicious file according to claim 409, wherein the high-level definition comprises at least one of the following high-level definitions:

a high-level definition for representing a source of a binary value;

a high-level definition for representing a specific meaning of a binary value in an instruction; and

a high-level definition for representing a relationship of the binary value in the instruction in an execution flow.

~~12. The apparatus for detecting a malicious file according to claim 10, wherein the determination module comprises:~~

~~a layer determination unit, configured to parse the to-be-detected file, to determine the binary values corresponding to the to-be-detected file as a first layer and to determine the high-level definition corresponding to the binary values as a second layer, wherein the second layer comprises at least one layer, and each layer corresponds to each type of high-level definition; and~~

~~a multi-layer image determination unit, configured to combine the first layer and the second layer, to generate a multi-layer image of the to-be-detected file.~~

~~1311.~~ The apparatus for detecting a malicious file according to claim ~~129~~, wherein the second layer comprises at least one of the following layers:

a layer for representing a source of a byte;

a layer for representing a specific meaning of a byte in an instruction;

a layer for representing a relationship of the byte in the instruction in an execution flow; and

a layer for a read-write permission of the byte.

~~1412.~~ The apparatus for detecting a malicious file according to claim ~~1311~~, wherein the layer for representing the source of the byte comprises: an instruction layer, a data layer, a PE header structure layer, and a resource section layer.

~~1513.~~ The apparatus for detecting a malicious file according to claim ~~1311~~, wherein the layer for representing the specific meaning of the byte in the instruction comprises: an operation code layer, an operand layer, and a shift length layer.

~~1614.~~ The apparatus for detecting a malicious file according to claim ~~1311~~, wherein the layer for representing the relationship of the byte in the instruction in the execution flow comprises: a target instruction address layer, a function range layer, and a node range layer of a control flow chart.

~~1715.~~ The apparatus for detecting a malicious file according to claim ~~129~~, wherein the detection module comprises:

a data dimensionality reduction unit, configured to reduce a dimension of the multi-layer image to generate dimension-reduced data; and

a file detection unit, configured to input the dimension-reduced data into the

neural network, and detect the to-be-detected file by means of the neural network.

~~18~~16. The apparatus for detecting a malicious file according to claim ~~17~~15, wherein the data dimensionality reduction unit is configured to reduce the dimension of the multi-layer image by using an unsupervised learning network and/or a supervised learning network, to generate the dimension-reduced data.

~~19~~17. A device for detecting a malicious file, comprising:
a memory, configured to store a computer program; and
a processor, configured to, when executing the computer program, implement steps of the method for detecting a malicious file according to any one of claims 1 to ~~9~~8.

~~20~~18. A computer-readable storage medium storing a computer program, wherein the computer program, when executed by a processor, implements steps of the method for detecting a malicious file according to any one of claims 1 to ~~9~~8.