METHOD AND SYSTEM FOR DETECTING AND PREVENTING MALWARE EXECUTION IN A TARGET SYSTEM

BACKGROUND

Field

[0001] This application generally relates to enterprise system security. In particular, this application describes a method and system for detecting malware on computer systems of the enterprise and for preventing execution of the malware.

Description of Related Art

[0002] Small and large computer systems may become infected with malware. Often times, the presence of the malware goes unnoticed until, for example, a ransom is received, computers become locked, etc. Such an attack may result in loss of intellectual property, lost productivity, lost revenue, and in some cases necessitate repair/replacement of components of the computer systems.

SUMMARY

[0003] A system for detecting and preventing execution of malware on a target system includes an interface for receiving training data. The training data includes domain names known to be legitimate and domain names known to be associated with malware. The system includes a processor in communication with the interface; and non-transitory computer readable media in communication with the processor that stores instruction code. The instruction code is executed by the processor and causes the processor to train a first model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names using a supervised learning methodology. The processor trains a second model to predict a correct domain name associated with domain names in the training data using an unsupervised learning methodology. The processor trains a third model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names based on an output of the first learning model and the second learning model. Subsequent to training of the first, second, and third models, the processor receive a new domain name and processes the new domain name through the trained models to determine whether the domain name is considered to be a legitimate domain name or a malware-associated domain name. The processor communicates a report to the target system based on the processing to thereby facilitate removal of the malware from the target system.

[0003a] According to a first aspect of the present invention, there is provided a system for detecting and preventing execution of malware on a target system, the system comprising: an interface for receiving training data that includes known legitimate domain names and known malware-associated domain names; a processor in communication with the interface; and non-transitory computer readable media in communication with the processor that stores instruction code, which when executed by the processor, causes the processor to: train a first model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names using a supervised learning methodology; train a second model to generate a predicted domain name associated with known legitimate domain names in the training data using an unsupervised learning methodology; train a third model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names based on an output of the first learning model and an output of the second learning model; subsequent to training of the first, second, and third models, receive a new domain name; process the new domain name through the trained models to determine that the domain name is a legitimate domain name or a malware-associated domain name; communicate with the target system based on the processing to thereby facilitate removal of the malware in response to the new domain name being determined as the malware-associated domain name by generating and transmitting information over a network or other communication path as instructions executable by the target system to remove, quarantine, restrict or otherwise disable detected malware in the target system.

[0003b] According to a second aspect of the present invention, there is provided a method for detecting and preventing execution of malware on a target system, the method comprising: receiving training data that includes known legitimate domain names and known malware-associated domain names; training a first model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names using a supervised learning methodology; training a second model to generate a predicted domain name associated with known legitimate domain names in the training data using an unsupervised learning methodology; training a third model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names based on an output of the first learning model and an output of the second learning model; subsequent to training of the first, second, and third models, receiving a new domain name; processing the new domain name through the trained models to determine that the domain name is a legitimate domain name or a malware-associated domain name; and communicating with the target system based on the processing to thereby facilitate removal of the malware from the target system by generating

and transmitting information over a network or other communication path as instructions executable by the system to remove, quarantine, restrict or otherwise disable detected malware in the target system.

[0003c] According to a third aspect of the present invention, there is provided a non-transitory computer readable media that stores instruction code for detecting and preventing execution of malware on a target system, the instruction code being executable by a machine for causing the machine to: receive training data that includes known legitimate domain names and known malware-associated domain names; train a first model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names using a supervised learning methodology; train a second model to generate a predicted domain name associated with known legitimate domain names in the training data using an unsupervised learning methodology; train a third model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names based on an output of the first learning model and an output of the second learning model; subsequent to training of the first, second, and third models, receive a new domain name; process the new domain name through the trained models to determine that the domain name is a legitimate domain name or a malware-associated domain name; and communicate with the target system based on the processing to thereby facilitate removal of the malware from the target system by generating and transmitting information over a network or other communication path as instructions executable by the system to remove, quarantine, restrict or otherwise disable detected malware in the target system.

CLAIMS

1. A system for detecting and preventing execution of malware on a target system, the system comprising:

an interface for receiving training data that includes known legitimate domain names and known malware-associated domain names;

a processor in communication with the interface; and

non-transitory computer readable media in communication with the processor that stores instruction code, which when executed by the processor, causes the processor to:

train a first model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names using a supervised learning methodology;

train a second model to generate a predicted domain name associated with known legitimate domain names in the training data using an unsupervised learning methodology;

train a third model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names based on an output of the first learning model and an output of the second learning model;

subsequent to training of the first, second, and third models, receive a new domain name;

process the new domain name through the trained models to determine that the domain name is a legitimate domain name or a malware-associated domain name;

communicate with the target system based on the processing to thereby facilitate removal of the malware in response to the new domain name being determined as the malware-associated domain name by generating and transmitting information over a network or other communication path as instructions executable by the target system to remove, quarantine, restrict or otherwise disable detected malware in the target system.

2. The system according to claim 1, wherein the domain names utilized to train the first model include both legitimate domain names and malware-associated domain names, and the domain names utilized to train the second model include legitimate domain names and not malware-associated domain names.

3. The system according to claim 2, wherein the domain names utilized to train the third model include both legitimate domain names and malware-associated domain names.

4. The system according to any previous claim, wherein the first model utilizes a long short-term memory configured as an encoder to generate as the output of the first model a binary output indicating a prediction that the new domain name is a predicted good domain or a predicted bad domain, and/or, wherein the second model implements an autoencoder that includes a first long short-term memory configured as an encoder to map the new domain name comprising a plurality of characters and a second-long short-term memory configured as a decoder to construct a predicted domain name comprising a plurality of characters, the predicted domain name provided as the output of the second model.

5. The system according to any previous claim, wherein the output of the first model is a binary output indicating a prediction that the new domain name is a predicted good domain or a predicted bad domain, and the output of the second model being a predicted domain name comprising a plurality of characters constructed by the second model based on the new domain name.

6. The system according to claim 5, wherein the third model is configured to classify the new domain name as a legitimate domain name or a malware-associated domain name based on the binary output of the first model and the predicted domain name output by the second model, and wherein the first model is configured to output a first confidence value indicative of a confidence level of accuracy of the binary output of the first model, the second model is configured to output a second confidence value indicative of a confidence level of accuracy of the predicted domain name output of the second model, and the third model (216) is configured to classify the new domain name as the legitimate domain name

or the malware-associated domain name using weighting based on the first and second confidence values.

7. A method for detecting and preventing execution of malware on a target system, the method comprising:

receiving training data that includes known legitimate domain names and known malware-associated domain names;

training a first model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names using a supervised learning methodology;

training a second model to generate a predicted domain name associated with known legitimate domain names in the training data using an unsupervised learning methodology;

training a third model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names based on an output of the first learning model and an output of the second learning model;

subsequent to training of the first, second, and third models, receiving a new domain name;

processing the new domain name through the trained models to determine that the domain name is a legitimate domain name or a malware-associated domain name; and

communicating with the target system based on the processing to thereby facilitate removal of the malware from the target system by generating and transmitting information over a network or other communication path as instructions executable by the system to remove, quarantine, restrict or otherwise disable detected malware in the target system.

8. The method according to claim 7, wherein training the first model to classify the domain names in the training data comprises providing as inputs to the first model both legitimate domain names and malware-associated domain names as the training data, and wherein training the second model to predict the predicted domain name comprises providing as inputs to the second model as the training datalegitimate domain names and omitting provision of the malware-associated domain names, and

wherein training the third model to classify the domain names in the training data comprises providing as inputs to the third model both legitimate domain names and malware-associated domain names.

9. The method according to claim 7 or claim 8, wherein the first model utilizes a long short-term memory configured as an encoder, and wherein processing the new domain name through the trained models comprises generating as the output of the long short-term memory configured as an encoder a binary output indicating a prediction that the new domain name is a predicted good domain or a predicted bad domain.

10. The method according to claim 9, wherein the second model implements an autoencoder that includes a first long short-term memory configured as an encoder and a second-long short-term memory configured as a decoder and wherein processing the new domain name through the trained models comprises mapping, with the first long short-term memory configured as an encoder, the new domain name comprising a plurality of characters and constructing, with the second-long short-term memory configured as a decoder, a predicted domain name comprising a plurality of characters, the predicted domain name provided as the output of the second model.

11. The method according to any of claims 7 to 10, wherein processing the new domain name through the trained models comprises:

outputting, by the first model, a binary output indicating a prediction that the new domain name is a predicted good domain or a predicted bad domain and a first confidence value indicative of a confidence level of accuracy of the binary output;

outputting, by the second model, a predicted domain name comprising a plurality of characters constructed by the second model based on the new domain name and a second confidence value indicative of a confidence level of accuracy of the predicted domain name; and

classifying, with the third model, the domain name as the legitimate domain name or the malware-associated domain name using weighting based on the first confidence value and the second confidence value.

12. A non-transitory computer readable media that stores instruction code for detecting and preventing execution of malware on a target system, the instruction code being executable by a machine for causing the machine to:

receive training data that includes known legitimate domain names and known malware-associated domain names;

train a first model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names using a supervised learning methodology;

train a second model to generate a predicted domain name associated with known legitimate domain names in the training data using an unsupervised learning methodology;

train a third model to classify the domain names in the training data as being legitimate domain names or malware-associated domain names based on an output of the first learning model and an output of the second learning model;

subsequent to training of the first, second, and third models, receive a new domain name;

process the new domain name through the trained models to determine that the domain name is a legitimate domain name or a malware-associated domain name; and

communicate with the target system based on the processing to thereby facilitate removal of the malware from the target system by generating and transmitting information over a network or other communication path as instructions executable by the system to remove, quarantine, restrict or otherwise disable detected malware in the target system.

13. The non-transitory computer readable media according to claim 12, wherein the domain names utilized to train the first model include both legitimate domain names and malware-associated domain names, and the domain names utilized to train the second model include legitimate domain names and not malware-associated domain names.

14. The non-transitory computer readable media according to claim 13, wherein the domain names utilized to train the third model include both legitimate domain names and malware-associated domain names.

15. The non-transitory computer readable media according to any of claims 12 to 14, wherein the first model utilizes a long short-term memory configured as an encoder; and/or wherein the second model implements an autoencoder that includes a first long short-term memory configured as an encoder and a second-long short-term memory configured as a decoder.