СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Некоторые разновидности вредоносных программ используют алгоритмы генерирования доменных имен для определения адресов управляющих серверов. Подобные алгоритмы позволяют защитить вредоносные сервера от однократного отключения или добавления адресов в черные списки. Чаще всего данные алгоритмы используются в круаных ботнетах.

Ботнеты - некоторая сеть, в том числе компьютерная, состоящая из устройств (ботов), со специально запущенным вредоносным программных обеспечением. Чаще всего боты инфицируются посредством вредоносного программного обеспечения, полученного из сети Интернет. Однако путём инфицирования может служить также локальная сеть или устройства ввода, например флэш накопители. Ботнеты являются наиболее распространенными средствами кибер атак. Они управляются его создателем при помощи специальных управляющих командных серверов (Command and Control Servers). Большинство из них используются для монетизации различными способами, такими как: распределенные атаки отказ в обслуживании (DDoS атаки), продажа Drive By Download, атаки на клиентов дистанционного банковского обслуживания, для спама и проведения фишинговых атак.

Для удержания контроля над ботами и их управления Ботнеты используют множество способов. Это может быть p2p сети, почтовые протоколы, социальные сети или анонимные сети, такие как TOR или i2p. Однако самым распространенным на данный момент является Алгоритмы Генерации Доменных Имен (Domain Generation Algotithms).

Они позволяют удерживать контроль над управляющими серверами. В основном подобные алгоритмы используются в крупных ботсетях. Например, одним из первых случаев был компьютерный червь Conficker в 2008 году. На сегодняшний день подобных вредоносных программ насчитываются десятки, каждая из которых представляет серьезную угрозу. Помимо этого, алгоритмы совершенствуются, их обнаружение становится сложнее. Например, осенью

2014 года была обнаружена новая версия ботнета Matsnu, в которой для генерации доменов используются существительные и глаголы из встроенного списка.

Целью данной работы является разработка модели на основе методов машинного обучения для распознавания и классификации вредоносных доменных имен, полученных при помощи анализа алгоритмов генерации доменных имен.

Идея использования методов машинного обучения освещена в работе [7]. Так, ряд известных компаний, занимающихся информационной безопасностью (Damballa, OpenDns, ClickSecurity и др.), применяют подобные решения для анализа и фильтрации сетевой активности вредоносных программ.

Последние исследования[6] показывают, что алгоритмы генерации доменных имен совершенствуются с целью обхода существующих способов обнаружения. Поэтому рассмотрение алгоритмов машинного обучения для предотвращения современных угроз является актуальной проблемой информационной безопасности.

1 АЛГОРИТМЫ ГЕНЕРАЦИИ ДОМЕННЫХ ИМЕН

Алгоритмы Генерации Доменных Имен (DGA) представляют собой алгоритмы, используемые вредоносным программным обеспечением (malware) для генерации большого количества псевдослучайных доменных имен, которые позволят им установить соединение с управляющим командным центром.

1.1 Cryptolocker

cryptolocker

1.2 Zeus

1.3 PushDo

1.4 Rovnix

1.5 Tinba

1.6 Conficker

1.7 Matsnu

1.8 Ramdo

2 АЛГОРИТМЫ МАШИННОГО ОБУЧЕНИЯ

2.1 Naive Bayes

3 НЕЙРОННЫЕ СЕТИ

3.1 Реккурентные нейронные сети

3.2 LSTM

4 ЭКСПЕРИМЕНТ

4.1 Обучающая выборка

4.2 Выделение признаков

4.3 Результаты алгоритмов

4.4 Результаты LSTM

4.5 Сравнительный анализ

ЗАКЛЮЧЕНИЕ

Выводы, значение полученных результатов Рекомендации по применению

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. T. Barabosch, A. Wichmann, F. Leder, and E. Gerhards-Padilla Automatic extraction of domain name generation algorithms from current malware.// STO-MP-IST-111 Information Assurance and Cyber Defence, 2012.

2. P. Barthakur, M. Dahal, and M. K. Ghose. An efficient machine learning based classification scheme for detecting distributed command & control traffic of p2p botnets.// 5(10):9, 2013.

3. Click Security. Exercise to detect algorithmically generated domain names.// 2014.

4. N. Davuth and S.-R. Kim. Classification of malicious domain names using support vector machine and bigram method.// 7(1):51–58, January 2013.

5. J. Jacobs. Building a dga classifier. [Электронный ресурс] // URL: http://datadrivensecurity.info/blog/posts/2014/Oct/dga-part3/, October 2014 (дата обращения:)

6. Raff. generation Dgas: A evolution. [Электронный ресурс] // URL: http://www.seculert.com/blog/2014/11/dgas-a-domain-generation-evolution, November 2014. (дата обращения:)

7. M. Stevanovic and J. Pedersen. Machine learning for identifying botnet network traffic //, 2013.

8. Z. Wei-wei, G. Jian, and L. Qian. Detecting machine generated domain names based on morpheme features.// pages 408–411, october 2013.

9. 9. S. Yadav, A. K. K. Reddy, A. L. N. Reddy, and S. Ranjan. Detecting algorithmically generated domain-flux attacks with dns traffic analysis.// 20(5):1663–1677, Oct. 2012.

10. Н. О. Гончаров. Современные угрозы ботсетей.// 10, октябрь 2014.

11. Machine Learning Text feature extraction (tf-idf) [Электронный ресурс] // URL: http://blog.christianperone.com/?p=1589 (дата обращения:)

12. Understanding LSTM Networks [Электронный ресурс] // URL: http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (дата обращения:)

13. Chunting Zhou, Chonglin Sun, Zhiyuan Liu, Francis C.M. Lau A C-LSTM Neural Network for Text Classification // 30 Nov 2015