А. В. Вешторт, Д. В. Царегородцев

НЕЙРОСЕТЕВАЯ КЛАССИФИКАЦИЯ ВРЕДОНОСНОГО ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ НА ОСНОВЕ АНАЛИЗА ПОВЕДЕНИЯ

Статья описывает решение проблемы анализа больших объемов вредоносного программного обеспечения путем классификации вредоносных образцов в автоматически сформированные классы, а также саму реализацию такого классификатора.

I. Введение

Ежедневно в мире создаётся большое количество новых экземпляров вредоносных программ. Очевидно, что подробное изучение каждого такого образца требует больших затрат человеческих и временных ресурсов. Потому появляется необходимость создания такой автоматизированной системы, которая позволяла бы получить базовую информацию о каждом вредоносном образце.

Поэтому было решено разработать систему, разбивающую всю совокупность новых образцов вредоносного ПО на некоторое ограниченное количество хорошо изученных классов на основе набора поведенческих признаков.

II. Выделение признаков

Для решения задачи классификации было выделено множество бинарных признаков, которые являются, по сути, сигнатурами, основанными на определенном вредоносном поведении. Среди всех таких признаков был выделен 21 основной, чей вклад в общую дисперсию исследуемой выборки вредоносных экземпляров максимален.

III. Выделение классов

Одной из основных проблем классификации вредоносного програмного обеспечения является отсутствие четких классов вредоносных образцов. Каждый производитель антивирусного программного обеспечения использует собственную недокументированную либо очень размытую классификацию. В таких условиях, мы были вынуждены разработать свою собственную классификацию на основе выделенных нами признаков.

Для непосредственного формирования классов было решено использовать метод x-means++, являющийся комбинацией двух модификаций широко известного метода кластеризации k-means[1]: x-means[2] и k-means++[3].

В результате применения метода x-means++ было выделено 20 классов вредоносного ПО.

IV. Классификация

Для классификации новых образцов в полученные классы было решено использовать однослойный персептрон с сигмоидной функцией активации [4].

V. Заключение

В результате проделанной работы был разработан и внедрен в Cuckoo Sandbox прототип классификатора вредоносного программного обеспечения, который уже сейчас активно используется в антивирусной лаборатории ОДО «ВирусБлокАда». Дальнейшее развитие проекта будет направлено на общее улучшение качества кластеризации путем введения новых переменных, а также на оптимизацию реализованных алгоритмов.

1. Pelleg D. Accelerating Exact k-means Algorithm with Geometric Reasoning / D. Pelleg, A. Moore // Carnegie Mellon University, School of Computer Science, Pittsburgh, USA – 2000.
2. Pelleg D. X-means: Extending K-means with Efficient Estimation of Number of Clusters / D. Pelleg, A. Moore // Carnegie Mellon University, School of Computer Science, Pittsburgh, USA – 1999.
3. Arthur D. K-means++: The Advantages of Careful Seeding / D. Arthur, S. Vassilvitskii // Stanford, USA – 2006.
4. Rosenblatt F. The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain / F. Rosenblatt//Cornell Aeronautical Laboratory, USA – 1958.

Вешторт Алесь Витальевич, студент кафедры интеллектуальных информационных технологий БГУИР, ales.veshtort@gmail.com.

Царегородцев Дмитрий Владимирович, студент кафедры интеллектуальных информационных технологий БГУИР, nakir24@gmail.com.

Научный руководитель: Вардеванян Левон Гарегинович, магистрант кафедры телекоммуникаций и информационных технологий БГУ, инженер-программист ОДО «ВирусБлокАда», lv@vba.com.by.