Применение алгоритма Холланда к задаче обнаружения атак

Линь Данил

Студент

Филиал МГУ имени М. В. Ломоносова в г. Севастополе, Севастополь, Россия

E-mail: linadnil@mail.ru

В 80-е годы XX века появилось и начало активно развиваться направление информационной безопасности, связанное с обнаружением атак на информационную систему (ИС), в качестве эффективного временного решения, позволяющего закрывать «бреши» в безопасности систем до их исправления. Существует два основных подхода к обнаружению сетевых атак: определение злоупотреблений и выявление аномальной активности. Первый метод заключается в том, что сетевые пакеты сравниваются с шаблонами, содержащими признаки атаки. Соответствие образца известной атаке называется сигнатурой. В работе рассматривается применение алгоритма, предложенного Холландом и называемом Simple Genetic Algorithm (SGA), для синтеза сигнтур атак на ИС.

К основным проблемам, которые должны быть решены при настройке алгоритма, можно отнести следующие: кодирование структуры сигнатуры, выбор механизмов скрещивания и мутации, настройка критерия выживаемости. Рассмотрим настройку алгоритма на решение задачи обнаружения атак.

Сигнатура кодируется целочисленным массивом, содержащим информацию о семи признаках сетевого соединения. Каждый признак описывается несколькими компонентами(генами) и имеет свой индекс. Так, например, признак, характеризующий продолжительность соединения, имеет 3 гена. В таблице 1 приводятся индексы,типы используемых данных и количество генов каждого признака.

Каждая сигнатура представляет собой условный оператор (if A then B), где A есть логическое объединение первых шести признаков посредством логического оператора AND, а B – признак “тип атаки”, являющийся результатом логического следования. В сигнатуре также возможен групповой символ “-1”, обозначающий любое возможное значение гена. Далее приводится пример сигнатуры классифицирующей сетевое соединение как DoS-атаку именуемую “neptune”.

Индекс	Признак	Формат	Количество генов
1	Продолжительность	h:m:s	3
2	Протокол	int	1
3	Порт сервера	int	1
4	Порт клиента	int	1
5	IP адрес сервера	a.b.c.d	4
6	IP адрес клиента	a.b.c.d	4
7	Тип атаки	int	1

Таблица 1: Кодирование данных

```
If (duration=”0:0:1” and protocol=“finger” and source_port=
18982 and destination_port=79 and source_ip=”9.9.9.9” and
destination_ip=”172.16.112.50”) then(attack_name=”neptune”)
```

Таким образом, данная сигнатура будет кодироваться следующим обазом:

[0, 0, 1, 2, 18982, 79, 9, 9, 9, 9, 172, 16, 112, 50, 1]

В алгоритме используются стандартные операции скрещивания и мутации SGA. В качестве критерия выживаемости (fitness-функции) используется формула основаная на оценке ассоциативных правил:

$$F = \frac{|AandB|}{N} + \frac{|AandB|}{|A|} \tag{1}$$

где |A and B| - число соединений для которых выполняется правило if A then B

N – общее число сетевых соединений

|A| - число соединений, для которых выполняется условие A.

Для оценки, полученных в ходе работы алгоритма сигнатур используется эталонный набор данных предоставляемый организацией DARPA. Данный набор содержит записи журналов событий реальной компьтерной сети.

Предложенный алгоритм может быть использован для обнаружения следующих типов атак: сканирование активных портов, удаленный запуск командного интерпритатора, попытка подбора пароля.

Литература

1. Ren Hui Gong, Mohammad Zulkernine, Purang Abolmaesumi. A Software Implementation of a Genetic Algorithm Based Approach to Network Intrusion Detection. 2005
2. Костенко В. А. Алгоритмы оптимизации, основанные на методе проб и ошибок. 2013 С. 57–60