3. Вентцель Е.С. Теория вероятностей: Учеб. для вузов. [Текст]/Е.С. Вентцель - М.: Высш. шк., 1999. - 537 с.

4. Вентцель Е.С. Исследование операций. [Текст]/ Е.С. Вентцель –М.: Советское радио, 1972 - 210 с.

5. Бахвалов Н.С., Жидков Н.П., Кобельков Г.М. Численные методы. [Текст]/Н.С. Бахвалов, Н.П. Жидков, Г.М. Кобельков - М.: Наука, 1987. – 636 с.

6. S. Kullback, R.A. Leibler. On Information and Sufficiency [Текст]/ S. Kullback, R.A. Leibler - Ann. Math. Statist №1, 1951 - p. 79-86.

А.Н. Ивкин, М.Е. Бурлаков

РЕАЛИЗАЦИЯ СИСТЕМЫ ОБНАРУЖЕНИЯ ВТОРЖЕНИЙ С ВНЕДРЕННЫМИ ПРАВИЛАМИ МАШИННОГО ОБУЧЕНИЯ

(Самарский университет)

Аннотация

Система обнаружения вторжений является одним из важнейших устройств для защиты вычислительных систем, она способна выявлять и исследовать пакеты сетевого трафика. СОВ Snort - это бесплатное программное обеспечение с открытым исходным кодом, используемое в качестве средства защиты сети. Инструмент Snort обнаруживает только подтвержденные атаки, используя заранее определенные сигнатуры. В целях обнаружения новых, ранее не известных сетевых атак в данной работе разработаны расширенные правила для Snort, полученные с помощью инструмента машинного обучения WEKA и алгоритма j48. В статье, для экспериментального исследования, используется набор данных KDDCUP99. Основная цель данного исследования – реализация СОВ с внедренными правилами инструмента машинного обучения. Основными этапами исследований являются подготовка данных, применение алгоритма машинного обучения, извлечение экспертных правил, реализация правил Snort, обнаружение атак. Предлагаемая система обеспечивает эффективные показатели обнаружения и успешно выявляет новые, не представленные в сигнатурах атаки.

Введение

Система обнаружения вторжений (СОВ), специализированное программно-аппаратное средство, предназначенное для выявления несанкционированного доступа к ресурсам системы. Snort это СОВ с открытым исходным кодом [1]. Как правило snort разворачивают на маршрутизаторе как сетевую СОВ. Snort обнаруживает атаки на основе правил, написанных в заданном формате и синтаксисе. Snort - это многовариантный инструмент исследования пакетов, работающий в нескольких режимах.

Основное преимущество Snort заключается в гибкости и простоте модификации правил по сравнению с другими коммерческими СОВ. Архитектура Snort показана на рисунке 1.

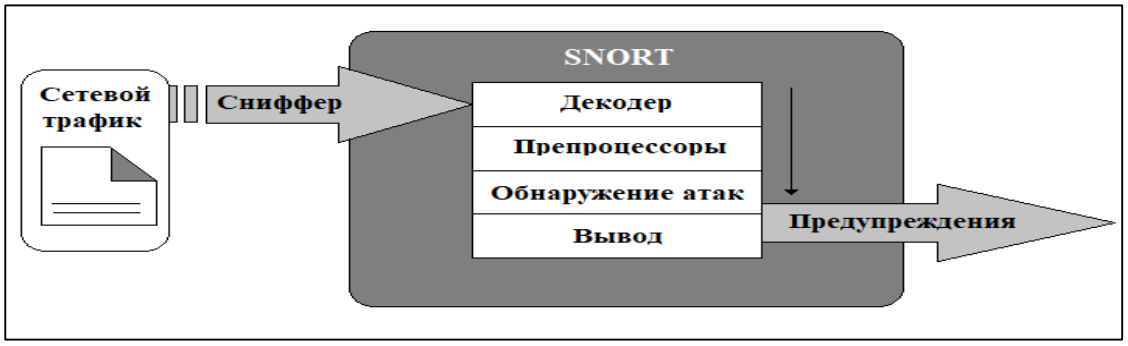

Рис. 1. Архитектура СОВ Snort

Правила Snort записываются в одной строке: заголовок правила и параметр. В заголовке правила содержатся действия правила, название протокола, IP-адреса, номера портов. Информация о проверяемой части пакета и предупреждающие сообщения включены в раздел параметров правил Snort. Пример простого правила Snort, определяющего наличие SYN флага, показан на рисунке 2.

```
alert tcp any any -> $MY_NET any (flags: S; msg: "SYN packet";)
```

Рис. 2. Пример правила Snort

Построение набора данных

В данной работе использовался инструмент машинного обучения WEKA – это программное обеспечение с открытым исходным кодом для машинного обучения [2]. Согласно результатам исследований [3], в статье используется дерево решений j48, более известное как алгоритм C4.5. Дерево решений *J48* - это алгоритм машинного обучения, определяющий значение примера на основе различных значений, наборов признаков (атрибутов), доступных данных.

Набор данных *KDDCup99* содержит множество отдельных атак, таких как *apache2*, *back*, *xterm* и т. д. [4]. Первоначально набор данных обрезается для удаления излишков. После удаления, отдельные атаки заменяются их категориями, как показано на рисунке 4.

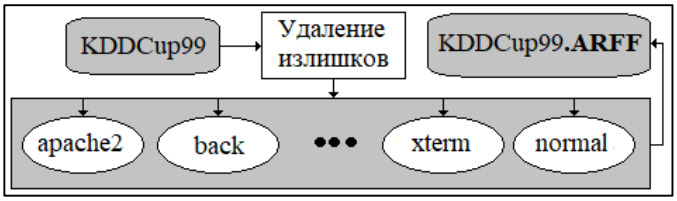

Рис.4. Процесс корректирования набора данных KDDCUP99

В этой работе мы использовали набор данных *KDDCUP99* (*corrected.zip*), который состоит из (311029 записей). В таблице 1 представлены все категории атак и количество образов.

Таблица 1 - Количество образцов и категории атак в *KDDCup99*

Категория атаки	Количество образов	Атаки в наборе данных *KDDCup99*
Normal	60589	*Normal*
Dos	229853	*apache2, back, land, mailbomb, processtable, smurf…*
R2L	16179	*ftp-write, guesspassword, imap, multihop, named, phf…*
U2R	228	*buffer_overflow, httprunnel, oadmidule, perl, rootkit…*
Probe	4165	*ipsweep, Mscan, Nmap, portsweep, saint, satan…*
Total	311014	Total

Выбор подобного набора данных обусловлен ориентацией на решение вопросов, связанных с обучением адаптивных алгоритмов [5-6].

Извлечение и применение экспертных правил

В статье используются наилучшие параметры входа, полученные экспериментальным путем для алгоритма j48, на рассмотренном наборе. В таблице 2 представлены результаты классификации откорректированного набора данных KDDCUP99.

Таблица 2 - Результаты классификации *KDDCup99.arff*

Алгоритм с набором входных параметров	Точность	Полнота	F-measure	AUC
*J48 (*C=0.60 M=2 MDL=on –S)	0.981	0.980	0.980	0.999

С помощью программы написанной на языке python из построенного дерева решений извлекаются экспертные правила для каждой из атак. Пример одного из извлеченных экспертных правил для атаки типа *Back* показан на рисунке 5.

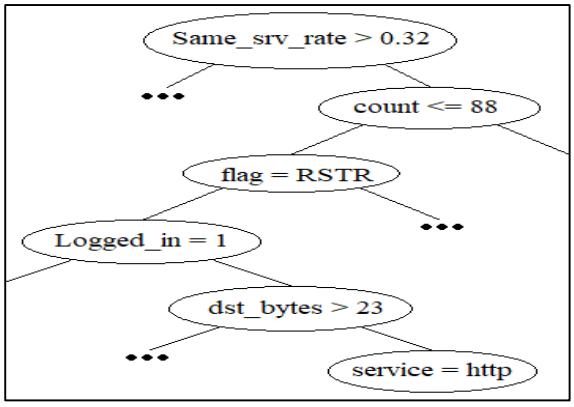

Рис. 5. Экспертное правило, извлеченное из WEKA

Где *Same_srv_rate* – процент подключений к одному и тому же сервису, *count* – количество подключений к одному и тому же хосту за последние 2 секунды, *flag* – нормальный или ошибочный статус соединения, *Logged_in* – по-

казатель входа в систему, dst_bytes – количество отправленных байтов источнику, service – сервис.

Далее извлеченные экспертные правила оформляются согласно синтаксису языка правил Snort, и записываются в файл Snort.rules. Snort запускается в режиме сетевой СОВ и обращается к файлу с правилами, как указано в snort.config. В таблице 3 показаны результаты работы реализованной СОВ на базе Snort с внедренными экспертными правилами.

Таблица 3 - Результаты работы предложенной СОВ

Данные	Количество	Процент	F-measure
TP + TN	304868	98.0192	0.98
FN + FP	6161	1.9808	

Заключение

В работе реализована СОВ, с внедренными экспертными правилами, полученными из инструмента машинного обучения WEKA с применением алгоритма j48. Было реализовано программное обеспечение для извлечения экспертных правил из WEKA. Подобранны оптимальные параметры входа для алгоритма j48 на наборе KDDCUP99. Из набора данных была удалена некорректная и устаревшая информация, а также набор был отформатирован в совместимый с инструментом машинного обучения формат. Все экспертные правила реализованы на языке правил Snort. Предлагаемая система обеспечивает эффективные показатели обнаружения (свыше 98 %) и успешно выявляет новые, не представленные в сигнатурах атаки. В дальнейшем набор данных будет заменен на более современный, а СОВ развернута на реальном сетевом трафике.

Литература

1. Sourcefire. Snort open source network intrusion prevention and detection system (ids/ips) [Электронный ресурс]. – Режим доступа: http://snort.org, свободный (Дата обращения 19.05.19).

2. Hall, Mark. The WEKA data mining software: an update / Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten. – ACM SIGKDD explorations newsletter_11, no. 1, 2009. –10-18 p.

3. Urvashi Modi. An improved method to detect intrusion using machine learning algorithms / Urvashi Modi, Anurag Jain. – Informatics Engineering, an International Journal (IEIJ), Vol.4, No.2, June 2016

4. Wu, Su-Yun. Data mining-based intrusion detectors / Wu, Su-Yun, Ester Yen. – Expert Systems with Applications_36, no. 3, 2009. – 605-612 p.

5. Stolfo, S.J. Anomalous Payload-based Network Intrusion Detection / Stolfo, S.J. – Heidelberg, vol. 3, 2004. – 190–240 p.

6. Chan, P.K. Learning Rules for Anomaly Detection of Hostile Network Traffic / Chan, P.K. – In Proc. of the 3rd IEEE Int. Conf. on Data Mining, 2003. – 50-76 p.