УДК 004.738 (043.3)

Данилина Г.В., канд. техн. наук
Милокум Я.В.

КОМБИНИРОВАННЫЙ МЕТОД ОБНАРУЖЕНИЯ СЕТЕВЫХ АТАК НА ОСНОВЕ АНАЛИЗА И РАСПОЗНАВАНИЯ АНОМАЛЬНЫХ СОСТОЯНИЙ СЕТИ

Институт компьютерных технологий
Национального авиационного университета

Предложен комбинированный метод обнаружения сетевых атак на основе анализ сигнатур, протоколов и статических методов обнаружения и распознавания. Представлена методика совместного обнаружения сетевых атак сигнатурными и статистическими методами

Постановка задачи

Применяемые при обнаружении и предотвращении сетевых атак (СА) методы сводятся к анализу сигнатур или протоколов. Анализ сигнатур базируется на простом понятии совпадения последовательности с образцом. Во входящем пакете просматривается байт за байтом и сравнивается с сигнатурой (подписью) – характерной строкой программы, указывающей на характеристику вредного трафика. Такая подпись может содержать ключевую фразу или команду, которая связана с нападением. Если совпадение найдено, объявляется тревога. В противном случае в пакете отыскивается следующая подпись. Как только все подписи проверены, в память записывается следующий пакет, и процесс начинается снова.

Второй метод анализа состоит в рассмотрении строго форматированных данных трафика сети, известных как протоколы. Каждый пакет сопровождается различными протоколами. Каждый протокол имеет несколько полей с ожидаемыми или нормальными значениями. Если что-нибудь нарушает эти стандарты, то вероятна злонамеренность. Анализ протокола использует детальное знание об ожидаемых или нормальных значениях в полях пакета, для того чтобы обнаружить вредоносный трафик.

Системы анализа сигнатур обладают высокой скоростью, в отличие от систем анализа протоколов. С другой стороны, со временем скорость анализа будет снижаться с ростом числа проверяемых сигнатур. Это – существенная проблема, поскольку число проверяемых сигнатур может расти очень быстро. Фактически, список проверяемых сигнатур увеличивается при реализации каждой атаки нового типа.

Учитывая вышеизложенное, можно сделать вывод, что для вынесения предположения о цели и последствиях атаки необходима разработка системы обнаружения вторжений, профилирующей комбинации сигнатур и сетевых аномалий на основании задаваемых признаков. Инструментом для эффективного группирования и обработки может стать методика упорядочивания данных, которые не могут быть классифицированы по непосредственным признакам. Наиболее перспективным направлением решения данной задачи является теория распознавания образов [1].

Сигнатурный анализ и контроль профилей (протоколов) включает в себя анализ заданных заранее как самих анализируемых данных, так и последовательностей действий.

Недостатками рассматриваемых моделей являются: для моделей, использующих статистические методики, – большое количество ложных тревог и ошибок второго рода, для моделей, использующих сигнатурные методики, – невозможность самостоятельного обнаружения новых атак и постоянная необходимость обновления базы сигнатур.

При обнаружении и предотвращении сетевых атак уже используются методики, имеющие возможность именно

предотвращения СА и включающие в себя только лишь такие действия, как блокировка приёма/передачи тех сетевых пакетов, которые идентифицируются как пакеты, содержащиеся в атаке.

В настоящее время невозможно применение формализованных методик и моделей обнаружения сетевых атак из-за их отсутствия, что в свою очередь не даёт правильного и точного обнаружения и предотвращения СА как в самой работе, так и в разработке СОА [2,3].

Вероятностная модель процесса обнаружения сетевых атак

При обнаружении и предотвращении СА рассматривается следующая математическая модель процесса обмена информацией между узлами сети [4].

1. Сетевой трафик представляется как совокупность дискретных сообщений $S_{k,1}^{n_{Uk,1}}$, где $n_{Uk,1}$ – номер сообщения от последнего по порядку источника сообщений U_k к первому по порядку источнику сообщений U_1; k - количество узлов в информационной системе (ИС).

2. После приема сообщения с номером $n_{Uk-1,1}$ могут иметь место следующие события:

– с вероятностью p прием сообщения с номером $n_{Uk,1}$;

– с вероятностью q потеря сообщения с номером $n_{Uk,1}$.

Вероятность приёма следующего сообщения после приёма/передачи предыдущего сообщения, обозначается как $P_{k,1}^{n_{Uk,1}}$ - переходная вероятность приёма сообщения с порядковым номером $n_{Uk,1}$ после приёма сообщения, отправленного от k-го узла к первому. Матрица переходных вероятностей имеет вид:

$$P = \begin{vmatrix} P_{2,1}^1, P_{2,1}^2, P_{2,1}^3, \ldots, P_{2,1}^{n_{U2,1}} \\ P_{1,2}^1, P_{1,2}^2, P_{1,2}^3, \ldots, P_{2,1}^{n_{U1,2}} \\ \cdots\cdots\cdots\cdots \\ P_{1,k}^1, P_{1,k}^2, P_{1,k}^3, \ldots, P_{1,k}^{n_{U1,k}} \\ P_{k,1}^1, P_{k,1}^2, P_{k,1}^3, \ldots, P_{k,1}^{n_{Uk,1}} \end{vmatrix}.$$

3. Состояния узлов ИС обозначаются как нулевое (исходное) состояние Q^0 и текущее состояние $Q_{k,1}^{n_{Uk,1}}$. Матрица состояний записывается в виде:

$$Q = \begin{vmatrix} Q_{2,1}^0, Q_{2,1}^1, P_{2,1}^2, \ldots, Q_{2,1}^{n_{U2,1}} \\ Q_{1,2}^0, Q_{1,2}^1, Q_{1,2}^2, \ldots, Q_{2,1}^{n_{U1,2}} \\ \cdots\cdots\cdots\cdots \\ Q_{1,k}^0, Q_{1,k}^1, Q_{1,k}^2, \ldots, Q_{1,k}^{n_{U1,k}} \\ Q_{k,1}^0, Q_{k,1}^1, Q_{k,1}^2, \ldots, Q_{k,1}^{n_{Uk,1}} \end{vmatrix}.$$

4. Совокупность статистических показателей сетевого трафика обозначается как $T = \{T_1, T_2, \ldots, T_h\}$, где h - число показателей. К таким показателям относятся:

– количество входящих *IP*-пакетов в единицу времени;

– количество исходящих *IP*-пакетов в единицу времени;

– количество входящих *TCP*-пакетов в единицу времени;

– количество исходящих *TCP*-пакетов в единицу времени;

– количество входящих *UDP*-пакетов в единицу времени;

– количество исходящих *UDP*-пакетов в единицу времени;

– время получения пакетов;

– время отправления пакетов;

– средняя продолжительность сеанса связи;

– вероятности P;

– состояния Q.

5. Сигнатуры представляют собой некоторую совокупность $M = \{M_1, M_2, ..., M_g\}$, где g – количество известных сигнатур атак. Например, для M_1 рассматривается следующие параметры:

- поле «адрес отправителя»;
- поле «адрес получателя»;
- поле «тип»;
- поле «данные»;
- поле «*CRC*»;
- непосредственно сами данные пакетов;
- время получения пакетов;
- время отправления пакетов;
- продолжительность сессии связи в сети.

Для M_2 могут рассматриваться такие параметры:

- поле «адрес отправителя»;
- поле «адрес получателя»;
- непосредственно данные пакетов и т.д.

Определим матрицы совокупностей сигнатур и статистических показателей. Матрицы совокупностей сигнатур M представляется в следующем виде:

	M_1	M_2	...	M_g
M_1	×	1	...	0
M_2	×	×	...	0
...	...	...	×	0
M_n	×	×	×	×

	M_1	M_2	...	M_g
M_1	×	1	...	1
M_2	×	×	...	1
...	...	...	×	1
M_g	×	×	×	×

, (1)

где элементы матриц вычисляются по правилу:

$$\begin{cases} M_{i,j} = 1, \text{ если } M \in \{M_1 \cup M_2, M_1 \cup M_3, M_1 \cup M_g, ..., M_1 \cup M_2 \cup ... \cup M_g\} \\ M_{i,j} = 0, \text{ если } M \notin \{M_1 \cup M_2, M_1 \cup M_3, M_1 \cup M_g, ..., M_1 \cup M_2 \cup ... \cup M_g\} \end{cases}$$

Матрицы статистических показателей T выглядят следующим образом:

	T_1	T_2	...	T_h
T_1	×	1	...	0
T_2	×	×	...	0
...	...	...	×	0
T_h	×	×	×	×

	T_1	T_2	...	T_h
T_1	×	1	...	1
T_2	×	×	...	1
...	...	...	×	1
T_h	×	×	×	×

, (2)

где элементы матриц вычисляются по правилу:

$$\begin{cases} T_{i,j} = 1, \text{ если } T \in \{T_1 \cup T_2, T_1 \cup T_3, T_1 \cup T_h, ..., T_1 \cup T_2 \cup ... \cup T_h\} \\ T_{i,j} = 0, \text{ если } T \notin \{T_1 \cup T_2, T_1 \cup T_3, T_1 \cup T_h, ..., T_1 \cup T_2 \cup ... \cup T_h\} \end{cases}$$

Алгоритм обнаружения и распознавания сетевых атак

Были исследованы свойства алгоритмов, которые могут быть применены для выявления аномального поведения информационно-вычислительной системы и распознавания конкретных атак на их объекты. Наиболее подходящими по критериям простоты настройки и быстродействия при обучении и эксплуатации являются методы распознавания образов [1]. При обнаружении СА рассматриваются матрицы сигнатур (1) и статистических показателей (2). В соответствии с принятой моделью марковского случайного блуждания [5] предельные (установившиеся или равновесные) вероятности состояний P_{sk} определяются из условия нормировки

$$\sum_{j=1}^{n_{U\,m,n}} P_{m,n}^{j} = 1, \; m = \overline{1,k}, \; n = \overline{1,k}, \quad (3)$$

согласно которому

$$P_{sk} = \frac{1 - p/q}{1 - (p/q)^{n-k+1}} \left(\frac{p}{q}\right)^{j-k} \quad j = k, \ldots, n. \quad (4)$$

Вероятности p и q в общем случае связаны с заданным качеством сервиса в сети.

Алгоритмическая схема процесса обнаружения СА, разработанная в соответствии с предложенной моделью, представлена на рис. 1.

На вход многоканальной параллельной системы подаются входные векторы из совокупностей T, M, P, Q. Начальные совокупности M_0 и T_0 формируются в процессе обучения системы на эталонном – специальном трафике, априори не содержащим СА.

Входные векторы формируются в блоке 2. В блоках 4.1 … 4.n формируются и анализируются сигнатуры и статистические показатели, включенные в состав (1) и (2). В блоке 5 выводы анализаторов формируются в совокупности итоговых выводов.

Совокупность итоговых выводов изначально формируется как вывод об отсутствии атаки при обучении анализаторов на эталонном трафике. Точнее входными векторами для блока 5 будут выходы блоков 4.1÷4.n, а все образцы выводов этих блоков при обучении на эталонном трафике будут формироваться как совокупности описанных сигнатур и статистических показателей, по которым выявляется отсутствие СА. Описанные итоговые выводы рассматриваются как множество предупреждений для формирования марковской модели [5]. В блоке 6 принимается решение о действии при обнаружении СА. В блоке 7 осуществляется предотвращение атаки путём блокировки необходимого сеанса связи между узлами в ИС или замена данных в сетевых пакетах в соответствии с ближайшей сигнатурой в кластере сигнатур, не содержащих атаки, сформированных на этапе обучения в совокупностях M.

Выводы

Рассмотренный метод обнаружения сетевых атак позволяет перейти на качественно новый уровень решения задачи обнаружения СА, с совместным применением для обнаружения сетевых атак марковской модели процесса приема данных, сигнатурных методик и анализатора протоколов. Для дальнейшего развития метода и доведения его до практического применения в дальнейшем планируется провести исследование асимптотической эффективности обнаружения сетевых атак с оцениванием ошибок первого и второго рода.

Список литературы

1. Вопросы статистической теории распознавания. / *Барабаш Ю.Л., Варский Б.В., Зиновьев В.Т., Кириченко В.С., Сапегин В.Ф.* – М.: Сов. радио, 1967. – 400 с.

2. *Amoroso, Edward, G.,* Intrusion Detection, 1st ed., Intrusion.Net Books, Sparta, New Jersey, USA, 1999.

3. *Denning, Dorothy.* (February, 1987). An Intrusion-Detection Model. IEEE Transactions on Software Engineering, Vol. SE-13, No. 2.

4. *Норткат С., Новак Дж.* Обнаружение безопасности в сетях.: Пер. с англ. – Изд. дом "Вильямс". – 2003. – 448 с.

5. *Тихонов В.И., Миронов М.А.* Марковские процессы. – М.: Сов. радио, 1977.– 488 с.

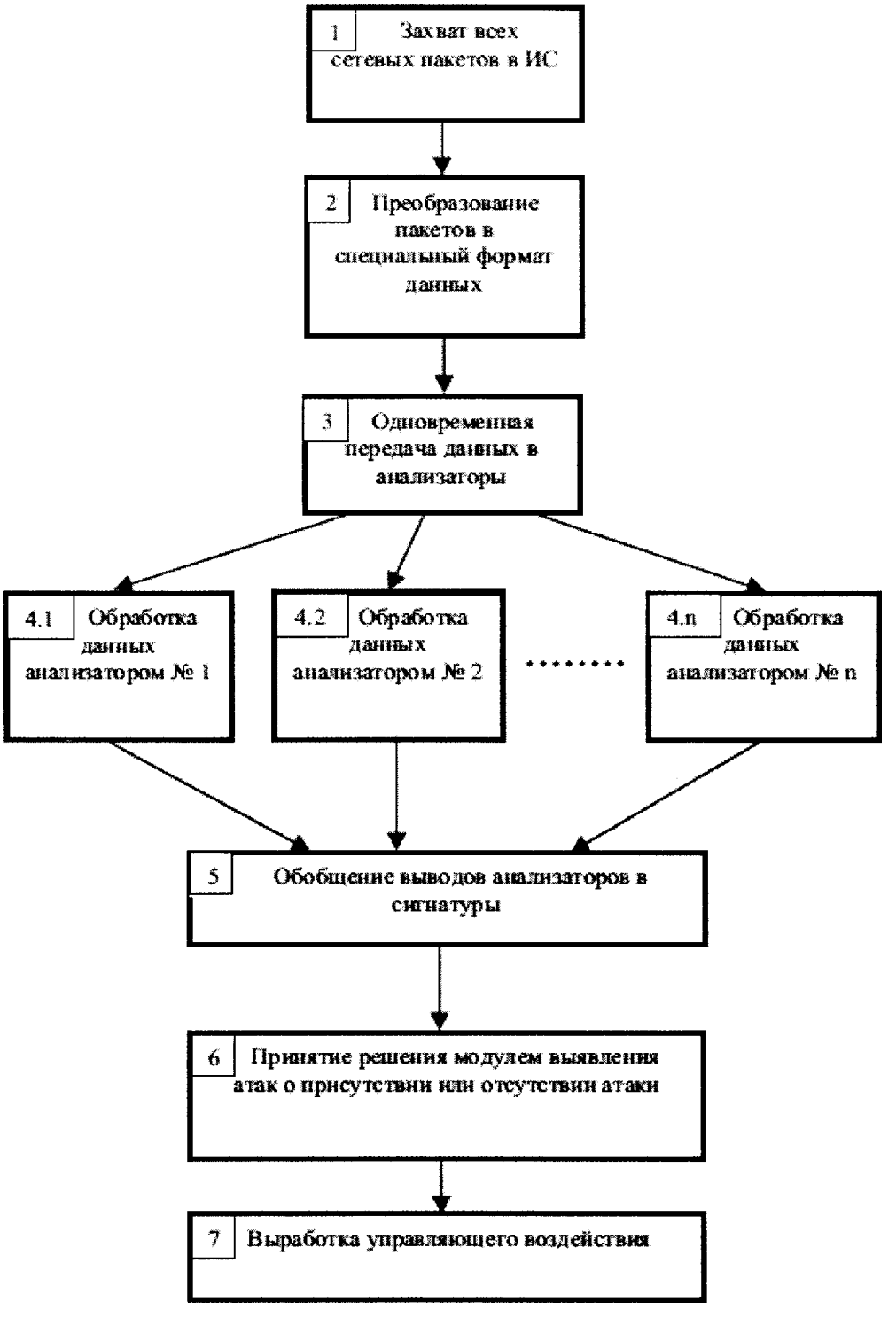

Рис. 1. Алгоритмическая схема процесса обнаружения СА