3 Разработка алгоритма обнаружения низкоинтенсивного сканирования сети в АС ВН.

В данной главе на основании предыдущих, нами будет проведена разработка алгоритма обнаружения низкоинтенсивного сканирования сети в АС ВН, позволяющий решить задачу данной работы.

3.1 Алгоритм обнаружения низкоинтенсивного сканирования сети в АС ВН.

Исходя из определенных нами границ исследования, выбранных программных средств, анализов работы средств защиты, используемых в ВС РФ, начинаем выстраивать пошагово блок-схему алгоритма.

Проводим сбор трафика. Собираем все соединения, имеющиеся с данной ЭВМ.

После сбора, данные соединения выгружаются в файл (список А), хранящийся в директории для трафика, подлежащего проверке,   
(директория А), данный файл получает имя, соответствующее дате сбора и диапазона времени.

Таким образом, система мониторинга не нагружена задачей постоянного хранения трафика. Трафик «нарезается» и отправляется в независимую от работы в реальном времени системы мониторинга программу.

Так как период сбора трафика равен одному часу, то возникает вероятность пропуска процесса сканирования: начало сканирования может находиться в предыдущем часовом диапазоне, а заканчиваться в текущем. Таким образом, процесс сканирования может анализироваться в разных временных интервалах по частям и может быть не обнаружен. Для решения этой проблемы предпринимаем следующее действие.

Так как мы уже определились с максимально эффективным размером «временного окна» сканирования – Т0, что равно пяти минутам, то с целью ликвидации вышеуказанной проблемы со снятого часового трафика все соединения, что находятся в диапазоне от последнего за час и до крайнего предыдущего включительно, попадающего во временной диапазон Т0 копируются отдельным файлом (список В) в специальную директорию (директория В).

Если анализируемый файл (список А) является первым за рабочую сессию, то он отправляется на обработку только после формирования и сохранения списка В на его основе. Иначе, список В с предыдущего трафика обработки добавляется к началу текущего, таким образом формируется новый файл на основе суммирования списков А и В - список С, далее использованный список В при создании списка С удаляется из директории. На основе списка С создается и сохраняется новый список В для добавления к последующим часовым дампам трафика.

Следующим шагом приступаем к анализу трафика.

Вначале мы уберем из списка соединений те из них, в которых   
IP-адрес отправителя совпадает с IP-адресом интерфейсов используемой ЭВМ.

Далее отбираем соединения по следующим критериям:

1. Количество пакетов не более 5 на соединение

Так как злоумышленник не имеет целью захватить данные, то соединение будет с наименьшей вероятностью установлено (при использовании SYN-сканирования), а при установлении, будет мгновенно разорвано.

1. Общая длина поля данных в потоке не более \_\_\_\_;

Фактор, указанный в пункте 1, имеет непосредственное отношение к данному критерию, соответственно объем информации обмена не должен быть более указанного значения, иначе данное соединение имеет информационный обмен и не может быть сканирующим.

1. Средняя длина поля данных в потоке не более \_\_\_\_;
2. Наличие пакетов с флагом RST в потоке;

Данный фактор не является достаточным для идентификации низкоинтенсивного сканирования, однако не может быть игнорирован.

1. Тип протокола прикладного и транспортного уровня (SMTP, HTTP и др.).
2. Номер порта отправителя.
3. Номер порта получателя.

Соединения будут отбираться по первому критерию, таким образом, при удовлетворении условия количества пакетов в сессии, соединение помечается подозрительным и отправляется на дальнейшее рассмотрение. Остальные параметры принимаются во внимание для принятия решения оператором об отнесении соединений к сканирующим, однако в первичной фильтрации участия не принимают.

Далее подвергаем обработке данные соединения, сортируя их по параметру «Адрес отправителя», таким образом, мы получаем множество подсписков (подсписки D) соединений от каждого конкретного адреса. Располагаем соединения в подсписках по времени возникновения в порядке возрастания данного параметра, вычисляем разность во времени между соединением N и N-1, где N = 2, …. , последнее соединение с данного   
IP–адреса. Данный параметр (Т) – период между соединениями вносим дополнительным в характеристику соединения. Возникает проблема определения вышеуказанного параметра для первого соединения в любом нарезанном участке трафика. В данном случае мы вычислим среднее значение периода между соединениями и присвоим это значение первому соединению фрагмента (Т’). Далее для всех соединений, имеющих параметр Т и T’ вычисляем параметр, равный Т0 – T (T’) соответственно, данному параметру присваиваем наименование T”. Назначение данного параметра описано ниже при рассмотрении оси Z.

После приведения списка в данный вид, формируем пространство данных для работы нейронного газа. Данное пространство состоит из трех осей координат, по принципу ХYZ.

X – количество пакетов в потоке.

Y – длина поля данных в потоке.

Z – максимальная временная задержка (5 минут) – временной интервал между данным соединением и предыдущим.

Далее выберем пороговые значения для каждого параметра, создавая трехмерную «красную плоскость». Все соединения, относящиеся к кластеру, который касается «красной плоскости» или находится между ней и точкой начала отсчета для каждой оси координат рассматриваются как подозрительные.

Рассмотрим расположение плоскости на осях координат (XYZ).

Ось Х:

В соответствии с первичным параметром отбора соединений, на оси Х «красная плоскость» находится в точке со значением 5.

Ось Y:

Длина поля данных, необходимая для идентификации сканирующего соединения должна быть не более \_\_\_\_\_, именно в таком значении находится точка расположения «красной линии» на оси Y.

Ось Z:

Для «красной плоскости» точка ее координаты на оси Z равняется:   
Тк – Тп, где Тк = 5 минут, а Тп равен пределу обнаружения активного сканирования штатными средствами защиты (\_\_\_\_\_\_) Данная точка имеет   
значение \_\_\_\_\_\_. Именно для этого и был введен параметр T”, так как низкоинтенсивное сканирование подразумевает большой временной интервал между соединениями, то с целью расположения подозрительных соединений между точкой отсчета и «красной плоскостью», мы и вычитали от Т0 Т и T’ соответственно.

В пространство данных загружаются соединения согласно подспискам D. Каждое соединение занимает свое положение в пространстве данных согласно параметров своей оси. Нейронный газ после обработки выдает нам трехмерное изображение параметров соединений, разбитое по кластерам. Рассчитываем среднеквадратичное отклонение по каждому параметру осей координат по формуле

,

где  – значение элемента выборки,  – среднее арифметическое,  – количество элементов выборки.

Все соединения, параметры которых будут меньше , будут считаться сканированиями. В случае обнаружения таковых соединений, они выгружаются отдельным файлом (списком E) на основе подсписка D отдельную директорию для подозрительного трафика (директорию Е) с трехмерной картинкой всех соединений. Обработанный трафик (список А или С) сохраняется в состоянии предшествующего анализу нейронным газом и выгружается в специальную для его хранения директорию обработанного трафика (директория F). Оператор получает оповещение о пополнении директории С, изучает поступивший файл и изображение, принимает решение о противодействии атаке или ложном срабатывании.

Пошаговый вид алгоритма:

1. Начало.
2. Сбор АС ВН соединений в течение 1 часа.
3. Вывод собранных соединений отдельным файлом (список А) в директорию (директория А) для хранения трафика, подлежащего обработке.
4. Проверка выполнения условия первого сбора трафика за рабочую сессию. При выполнении переход к п. 5, иначе к п. 6.
5. Формирование списка соединений (список В) отдельным файлом для дальнейших соединений, сохранение в отдельной директории (директория В). Формируется на основе списка А, путем отбора из него соединений во временном диапазоне Т1 в направлении от последнего по списку до крайнего предыдущего, попадающего в диапазон Т1.
6. Так как текущий список А не первый в сессии, то для полноты анализа необходимо добавить в него соединения из списка В, таким образом получаем список С
7. Очистка директории В от уже добавленного списка В в п.6.
8. Формируем и сохраняем в директории В новый список В для последующего файла трафика (списка А) на основании новообразованного списка С.
9. Удаление всех соединений, где IP–адрес отправителя равен   
   IP–адресу используемой ЭВМ из списка С (списка А для первого списка за сессию).
10. Анализ параметров первого (очередного для пп. 10.2, 13) соединения.
11. Проверка выполнения первичного условия для отбора   
    соединения - «количество пакетов». Пи выполнении – переход к п. 12, иначе к п. 11.1.
    1. При выполнении данного условия обработка файла с соединениями прекращается, переход к п. 11.3.
    2. Удаление данного соединения, как не прошедшего отбор из списка, начало рассмотрения очередного соединения в п. 4.
    3. При выполнении данного условия первичная фильтрация прошла успешно, прошедшие её соединения обрабатывается далее с п. 15, иначе нет соединений, удовлетворяющих ключевым условиям, и дальнейшая обработка не происходит, переход к п. 24.
12. Сохранение соединения в списке как прошедшего отбор.
13. При выполнении данного условия приступаем к рассмотрению очередного соединения в п. 10., иначе переход к п. 15.
14. Отобранные соединения сортируются по IP-адресу отправителя и располагаются в хронологическом порядке.
15. Формирование подмножеств (подсписков D) на основе IP-адреса отправителя.
16. Присваивание в каждом подмножестве соединениям параметра «Т»   
    (период между соседними соединениями), а первому соединению подмножества – «Т’» (среднее значение за подмножество), далее выполняя преобразования с ними присвоить значения T”.
17. Каждое подмножество по очереди выгружается в пространство XYZ, где:

X – количество пакетов в потоке.

Y – общая длина поля данных в потоке.

Z – временной интервал между данным соединением и предыдущим. Проводится кластеризация соединения расширяющимся нейронным газом.

1. Получение трехмерного изображения соединений и кластеров.
2. Сопоставление изображения соединений и кластеров пороговой плоскости.
3. При невыполнении данного условия работа с подмножеством прекращается, т.к. сканирующих соединений не выявлено, иначе п. 22.
4. Сохранение подмножества отдельным файлом (список D) и изображения в директорию для подозрительного трафика   
   (директория D).
5. При выполнении данного условия прекращается работа с текущим подмножеством, а очередное подмножество проходит анализ с п. 17, иначе п. 24.
6. Сохранение первичного файла соединений (п.3), прошедших п. 11, по состоянию на п. 16 (на п. 9. для не прошедших п. 11.) включительно, и перенос из директории для предварительной обработки (директория А) в директорию с обработанным трафиком (директория В).
7. Анализ файлов (для директории D) и принятие решения об отнесении данных соединений к сканирующим.
8. Конец.