электронный журнал
МОЛОДЕЖНЫЙ НАУЧНО-ТЕХНИЧЕСКИЙ ВЕСТНИК
Издатель ФГБОУ ВПО "МГТУ им. Н.Э. Баумана". Эл No. ФС77-51038.

УДК 004.732.056

Аспекты защиты информации в облачных технологиях

Богомолов И.В., *студент*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Защиты информации»

Маликов А.Ю., *аспирант*
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Защиты информации»

Научный руководитель: Богомолова Н.Е., к.т н.
Россия, 105005, г. Москва, МГТУ им. Н.Э. Баумана,
кафедра «Защиты информации»
runc@bmstu.ru

1. Введение

В облачных технологиях могут применяться средства защиты информации традиционных IT-инфраструктур [1], например, системы обнаружения вторжений, но они обладают недостаточной гибкостью, так как архитектура «облака» часто меняется, добавляются новые узлы, изменяется конфигурация. Поэтому систему обнаружения вторжений нужно постоянно перенастраивать, что приводит к временны́м и денежным затратам, а также рано или поздно может привести к ошибкам или несвоевременному обнаружению вредоносных воздействий. Наиболее полный обзор литературы по облачной безопасности произведен в [2]. Размещение и перемещение данных в «облаках» можно сравнить в логистическими задачами, решаемыми интеллектуальными управляющими системами, при транспортных перевозках грузов для составления расписаний [3], здесь применяются технологии мультиагентных систем с последующей оптимизацией. В данной работе рассмотрен алгоритм поиска элементов «облаков», в которых возникают различные виды атак на облачные сервисы. Описаны методы выявления различных угроз с помощью метода группового стохастического поллинга.

2. Постановка задачи

В настоящее время во всем мире и, в том числе в России, все большее распространение получают технологии облачных вычислений. Облачные вычисления – это информационно-технологическая концепция, подразумевающая обеспечение

повсеместного и удобного сетевого доступа по требованию к общему пулу конфигурируемых вычислительных ресурсов (например, сетям передачи данных, серверам, устройствам хранения данных, приложениям и сервисам — как вместе, так и по отдельности), которые могут быть оперативно предоставлены и освобождены с минимальными эксплуатационными затратами. Основным препятствием для их повсеместного применения являются проблемы безопасности, при этом многие проблемы облачной безопасности являются традиционными проблемами IT-инфраструктур. Архитектура системы защиты облачной инфраструктуры требует сбора информации с различных ключевых точек для выявления аномалий в поведении облачной среды. Этими ключевыми точками являются: запущенные виртуальные машины пользователей, хостовая операционная система для виртуальных машин, хранилища данных, а также такие узлы сети, как программные и аппаратные свитчи, межсетевые экраны и, в особенности, система управления «облаком».

Многие средства защиты, применяемые для традиционных IT-инфраструктур, с успехом применяются и в «облаках». В частности, широко применяются системы обнаружения вторжений. Принцип работы таких систем основан на сборе событий с сенсоров, распределенных по ключевым точкам защищаемой системы. Каждый сенсор имеет собственную политику безопасности, которая определяется общим блоком генерации политик безопасности. Этот блок состоит из редактора политик безопасности и средств конфигурирования безопасности виртуальных машин. Например, правило может быть следующим: «При обнаружении попыток изменения файлов в указанной директории, отказать в доступе и уведомить администратора». Сенсор также может содержать алгоритмы обучения, позволяющие более эффективно отслеживать внештатные ситуации. На основе обработки большого количества простых событий из сенсоров создаются высокоуровневые события, которые передаются на обработку сервису обработки событий. Этот сервис накапливает информацию и отслеживает аномалии в поведении на более высоком уровне абстракции. После обработки события сохраняются в архиве прошедших событий.

Обнаружение простых событий (например, вторжений) в системах защиты может осуществляться с помощью методов упорядоченного опроса сенсоров, получившего название «поллинга», который может быть как циклическим, так и динамически адаптивным. Время опроса при циклическом поллинге зависит от числа сенсоров, применение динамически адаптивных алгоритмов может снизить его не более чем в два раза, что все равно является значительным при возникновении критических ситуаций.

Поэтому для быстрого опроса сенсоров необходимо применять групповые методы, в данной работе предложено применить метод группового стохастического поллинга.

3. Математическая модель

Имеется распределенная облачная архитектура, в которую для обнаружения вторжений внедрены различные типы сенсоров, передающие информацию о состоянии безопасности элементов «облака». Всего имеется t сенсоров, требуется разработать стратегию их опроса для выявления s из них, имеющих информацию для передачи о внештатном состоянии элементов «облака» за наименьшее число шагов. Особенностью рассматриваемой модели является, что $s << t$, это соответствует случаю, когда общее число вторжений в облачную архитектуру невелико. То есть имеется t сенсоров, состояние которых описывается переменными $x1, x2, \ldots, xt$, которые могут принимать значения 0 или 1, соответственно при отсутствии угрозы безопасности и при возникновении угрозы.

Групповой опрос заключается в том, что одновременно опрашиваются несколько сенсоров. Задается случайный вектор опроса $a=(a_1, a_{2,\ldots}, a_t)$, в котором a_i принимает значение равное 1, если i- ый сенсор участвует в опросе и принимает значение равное 0, если не участвует. Всего проводится *N* опросов, таким образом, возникает булева матрица опросов *A*. Если в группе опрашиваемых сенсоров (строке матрицы) имеется хотя бы один активный, то передается сигнал, который интерпретируется как 1. Если в опрашиваемой группе нет ни одного активного сенсора, то передается сигнал, который интерпретируется как 0. Таким образом, в качестве ответа сенсоров $j-$ ой группы формируется результат:

$y_1 = \left(a_1^j \wedge x_1\right) \vee \ldots \vee \left(a_t^j \wedge x_t\right)$, где $\wedge$ - булево произведение, а $\vee$- булева сумма.

Необходимо получить алгоритмы построения матрицы опроса A и определить множество активных сенсоров на основании наблюдений $y_1, \ldots, y_N$.

В работе [4] показано, что асимптотически оптимальный план опроса *A* можно получить из случайной матрицы, в которой значения единиц выбираются независимо друг от друга с вероятностью $p_0 = 1 - \sqrt[s_0]{\frac{1}{2}}$, где s_0 – предполагаемое число активных сенсоров. Решение об активности конкретного сенсора принимается на основе пофакторного анализа с использованием метода максимального правдоподобия. Это отношение задается формулой правдоподобия $L(i) = a_{10}x_{10}(i) + a_{11}x_{11}(i)$, где

$$a_{10} = \log_2 \frac{\beta_1}{1-\beta_0-p^*(1-\beta_0-\beta_1)},$$

$$a_{11} = \log_2 \frac{1-\beta_1}{\beta_0+p^*(1-\beta_0-\beta_1)},$$

где $p^* = 1 - (1 - p_0)^{s_0}$.

Если значение отношения правдоподобия превышает порог $L(i) > L_0$, то считается, что сенсор является активным. Порог $L_0 > 0$ выбирается таким образом, что бы влияние случайных факторов было минимальным. При высоком значении этого порога вероятность неправильной идентификации сенсора меньше, но возрастает вероятность пропуска активного сенсора.

Как правило, на одном элементе «облака» устанавливается несколько сенсоров и в случае возникновения внештатных ситуаций угроз безопасности они все, или несколько из них могут сработать одновременно. Поэтому необходимо провести еще дополнительное исследование с учетом взаимного влияния сенсоров обнаружения угроз безопасности при взаимозависимых событиях. Очевидно, что при этом вероятность одновременного срабатывания нескольких сенсоров, как реакция на одно явление, существенно больше рассмотренной выше вероятности активности сенсоров, равной $\frac{s_0}{t}$. Поэтому для этого случая необходимо изменить модель и учитывать возможность одновременного срабатывания нескольких сенсоров.

Здесь, как и в случае независимого срабатывания сенсоров предполагается, что их число - B велико, а максимальное число одновременно срабатывающих сенсоров при возникновении угрозы безопасности $n^* = \max_{j=1,\dots,B} n_j$ относительно невелико. Таким образом, имеется $t = \sum_{j=1}^{B} n_j$ сенсоров и требуется разработать стратегию их опроса с целью скорейшего выявления s из них, имеющих данные о нарушении безопасности.

В данной задаче предполагается, что вероятность возникновения внештатных ситуаций в сети p_d не известна, но известно, что она достаточно мала: $p_d \le p_d^0$, где p_d^0 - некоторая заданная вероятность, аналог величины $\frac{s_0}{t}$, используемой при независимом срабатывании сенсоров. Вместе с тем, использование в алгоритме управления опросами верхней границы этой вероятности для расчета допустимого

числа активных датчиков $s_0 = n * Bp_d^0$ не эффективно, поскольку, используемый метод группового поллинга [4] чувствителен к предполагаемому количеству активных сенсоров. Это связано с тем, что при неправильном задании предполагаемого числа активных сенсоров значительно падает информативность данных опроса. Такая ситуация является наиболее типичной для облачных архитектур, которые распределены по большим территориям, а вероятность внештатных ситуаций на локальном участке «облака» мала. При возникновении внештатных ситуаций на j-ом элементе «облака» одновременно срабатывает z_j сенсоров, где z_j - дискретная случайная величина с распределением на множестве $\{1,\ldots,n_j\}$.

Далее возможны различные постановки задачи, в зависимости от априорных предположений о распределении величины z_j. Понятно, что распределение величины z_j зависит от структуры «облака», причины возникновения внештатной ситуации обнаружения угрозы безопасности, расположения сенсоров и т. п. Поэтому далее рассматривается предположение, что вероятность срабатывания всех сенсоров на элементе «облака» достаточно велика.

В остальном приняты все те же предположения: групповой опрос, как и ранее, задается с помощью вектора $a = (a_1,\ldots,a_t)$. Если N – число опросов, то все опросы задаются булевой матрицей опросов $A = (a^1,\ldots,a^N)^T$, где $a^j = (a_1^j,\ldots,a_t^j)$ - вектор j-го опроса. Отличие от рассмотренной выше модели состоит в том, что теперь величины $x_1,\ldots,x_t$ не являются независимыми.

В работе рассмотрен простейший случай $n_j \equiv n^*$, т. е. количество сенсоров на всех элементах одинаково. В этом случае сначала выбираются элементы «облака», на которых будут опрашиваться сенсоры, а затем на этих элементах выбирается опрашиваемый сенсор. Следовательно, в каждом опросе на одном элементе «облака» может участвовать не более чем 1 сенсор. Для этого сначала строится матрица опросов элементов «облака», как и в предыдущем случае (т.е. при $t = B$). Далее нужно на каждом элементе «облака» необходимо выбрать опрашиваемый сенсор. Делается это следующим образом: если объект участвует в

опросе, т. е. $a_i^j = 1$, где j – номер опроса, а i– номер объекта, то среди сенсоров i– объекта опрашиваемый сенсор выбирается случайным образом независимо от участия в предыдущих опросах.

Понятно, что здесь имеется некоторая избыточность, однако, она приводит лишь к добавлению небольшого числа дополнительных опросов. Для принятия решения о том, является ли элемент «облака» активным или нет, используются следующие данные: $x_{00}(i)$ - количество наблюдений, когда на i - ом элементе «облака» не опрашивался ни один сенсор и результат опроса оказался g=0, $x_{10}(i)$ - количество наблюдений, когда на i- ом элементе «облака» опрашивался хотя бы один сенсор, а результат опроса g=0, $x_{01}(i)$ - количество наблюдений, когда на i - ом элементе «облака» не опрашивался ни один сенсор, а результат опроса g=1 и $x_{11}(i)$ - количество наблюдений, когда на i - ом элементе «облака» опрашивался хотя бы один сенсор и результат опроса g=1. Поскольку результаты наблюдений предполагаются независимыми, то $x_{00}(i)$, $x_{10}(i)$, $x_{01}(i)$ и $x_{11}(i)$ образуют достаточную статистику.

В предположении, что на элементе «облака» срабатывают одновременно все сенсоры, получим для отношения правдоподобия поиска активных сенсоров те же результаты, что и в предыдущем случае. То есть, после проведения N опросов для каждого из t сенсоров вычисляются величины $x_{00}(i)$, $x_{10}(i)$, $x_{01}(i)$ и $x_{11}(i)$, а на их основании отношения правдоподобия

$$L(i) = a_{00}x_{00}(i) + a_{10}x_{10}(i) + a_{01}x_{01}(i) + a_{11}x_{11}(i) \quad . \qquad (4.2)$$

Далее решение о наличии угрозы безопасности на элементе «облака» принимается в том случае, когда хотя бы один сенсор признан активным.

4. Заключение

Рассмотренный в работе алгоритм обнаружения активных сенсоров при организации системы безопасности облачной инфраструктуры на основе группового стохастического поллинга позволит значительно снизить время обнаружения вторжений и других ситуаций, угрожающих безопасности облачной архитектуры. При этом работа алгоритма не зависит от структуры «облака» и изменения его конфигурации, поэтому система обнаружения вторжений не требует перенастройки при измени структуры

«облака». В дальнейшей работе планируется разработать имитационную модель для оценки и настройки параметров разработанной системы.

Список литературы

1. Куприянов А.И., Петренко П.Б., Сычев М.П. Теоретические основы радиоэлектронной разведки: учеб. пособие. М.: Изд-во МГТУ им. Н.Э. Баумана, 2010. 381 с.
2. Маликов А.Ю. Анализ угроз безопасности компонентам технологии облачных вычислений // Молодежный научно-технический вестник. Электрон. журн. 2014. № 1. Режим доступа: http://sntbul.bmstu.ru/doc/731369.html (дата обращения 17.03.2015).
3. Кузнецов Н.А., Пащенко Ф.Ф., Рябых Н.Г., Захарова Е.М., Минашина И.К. Алгоритмы оптимизации в задачах планирования на рельсовом транспорте // Информационные процессы. Т. 14, № 4. 2014. С. 307–318.
4. Малютов М.Б. Нижние границы для средней длины последовательного планирования экспериментов // Известия ВУЗов. Математика, 1983. Т. 27, № 11. С. 19-41.