УДК 004.051

ПРИНЦИПЫ ПОСТРОЕНИЯ МОДЕЛИ СИСТЕМ *HONEYPOT*

Ю.В. Алейнов

Технология динамических систем *Honeypot* является одной из перспективных технологий для защиты сетей предприятий от атак. В данной статье предлагается концепция модели для исследования компьютерной сети, включающей ложные информационные системы.

Ключевые слова: сетевые атаки, ложные информационные системы, системы *Honeypot*, модель сети.

Введение

В настоящее время всё большую актуальность приобретает идея использования ложных информационных систем (также называемых обманными системами или системами *Honeypot*) для защиты компьютерных сетей от атак. Система *Honeypot* — это система, основное предназначение которой — подвергнуться атаке. Ложная информационная система (ЛИС), будучи включена в вычислительную сеть, не выполняет никаких производственных функций. Её задача — отвлечь внимание злоумышленника от реальных систем, дезинформировать его, а в случае атаки на неё, собрать информацию о действиях атакующего для последующего анализа [1].

Среди недостатков, присущих подобным системам, можно выделить тот, что конкретные параметры конфигурации ЛИС (количество и свойства отдельных ловушек, их расположение) сильно зависят от свойств рассматриваемой сети. Для того чтобы спроектировать и построить систему *Honeypot*, нужно затратить большое количество ресурсов. А поскольку параметры сети с течением времени меняются (подключаются и отключаются устройства, меняется топология), необходимо постоянно поддерживать систему ловушек в актуальном состоянии. В производственной сети затраты, связанные с этим могут быть слишком велики. Решением может стать концепция *динамической системы Honeypot*, предложенная Лэнсом Спицнером (*Lance Spitzner*) [2]. Согласно этой концепции, динамическая система *Honeypot* — это система, способная при подключении к сети автоматически определять параметры конфигурации ЛИС и развертывать их в соответствующем адресном пространстве. В зависимости от изменения состояния сети со временем такая система адаптирует свою конфигурацию. В разных работах, посвящённых конструированию систем *Honeypot*, задача нахождения оптимальной конфигурации ЛИС неявно решается разными способами [3,4]. Формальная же постановка данной задачи и её решение в общем виде отсутствуют. В статье предложена схема модели сети, содержащей ЛИС, позволяющая ответить на данный вопрос.

Условия оптимальности конфигурации обманных систем.

Безусловно, конфигурация ЛИС в сети зависит и от преследуемых целей её применения, и от способов взаимодействия системы *Honeypot* с другими подсистемами защиты. Тем не менее, можно выделить следующие основные требования к обманным системам в сети:

- факт наличия ЛИС должен быть незаметным для злоумышленника
- ЛИС должны быть настроены так, чтобы обеспечить максимальную привлекательность ловушек для злоумышленника.

Если атакующему удалось узнать, что система, с которой он взаимодействует — ложная, он может использовать это в своих целях. Например, исключить её из списка своих целей или провести на неё отвлекающую атаку с целью дезинформировать защищающую сторону. Таким образом, необходимо создать для атакующего «условия априорной неопределённости» [5].

Вопрос о возможности различения реальной системы и системы-ловушки атакующим является комплексным. Он включает в себя, как минимум, две стороны.

Во-первых, это проблема имитации. Если ЛИС представляет собой сценарий, имитирующий какой-либо сервис, то очевидно, существуют различия в поведении имитатора и реального сервиса, благодаря которым можно обнаружить использование ЛИС.

Во-вторых, имеет значение распределение систем в сети. Первоначальный выбор цели осуществляется на основании ряда параметров, таких как тип используемой опера-

ционной системы и набор поддерживаемых сервисов. Распределение этих параметров по хостам и подсетям является характерным для данной конкретной сети. Большие отклонения в настройках ЛИС от этого распределения могут вызвать подозрения у атакующей стороны. Кроме того, существуют и вовсе маловероятные комбинации параметров, такие как наличие сервиса, имитирующего веб-сервер *IIS* на хосте под управлением операционной системы *Cisco IOS*. Далее будем говорить о задаче маскировки именно в контексте распределения параметров систем в сети.

Привлекательность обманной системы для атакующего — это понятие, которое трудно формализовать. С другой стороны, смысл условия привлекательности ЛИС для злоумышленника состоит в том, чтобы он предпочёл её реальной системе при выборе цели для атаки. Формально говоря, в каждый момент времени вероятность атаки на ЛИС должна быть максимально возможной. Говоря о привлекательности, мы полагаем, что злоумышленник имеет выбор. То есть существует вероятность атаки как реальной, так и ложной системы. Во многих работах предполагается схема включения ложных систем, в которой сетевой трафик перенаправляется на ловушку в случае срабатывания системы обнаружения вторжений. В этом случае атакующий не может непосредственно взаимодействовать с ложными системами, которые отражают реальные объекты сети.

Таким образом, представим задачи построения оптимальной конфигурации ЛИС в следующем виде:

- распределение параметров обманных систем должно повторять распределение этих же параметров среди реальных систем;
- в каждый момент времени вероятность атаки на одну из ЛИС должна быть максимально возможной.

Исходя из этого, будем строить формальную модель системы *Honeypot*.

Принципы формальной модели.

Прежде всего, определим понятие сети. Сеть — это множество хостов. При этом каждому хосту можно однозначным образом поставить в соответствие некоторый набор параметров:

$$N = \{N_i\}_{i=1..d} \ , \qquad (1)$$

$$f : N \to \Gamma, \Gamma = \{(\gamma_1, \gamma_2, ..., \gamma_l)\} \ , \qquad (2)$$

причём $\forall n_i, n_j = N$

$$f(n_i) \neq f(n_j) \Rightarrow n_i \neq n_j \ , \qquad (3)$$

где N – сеть;

n_i - хосты сети;

Γ - множество всевозможных наборов параметров одного хоста;

f - отображение, задающее соответствие между хостами сети и множеством комбинаций их параметров.

Пространство наборов Γ зависит от выбора значимых параметров хоста при построении модели. Так, если в рамках модели какие-либо различия между хостами отсутствуют, то имеет место вырожденный случай, и можно положить $\Gamma = .Ø$ Выбор множества Γ влияет на степень детализации при имитации распределения параметров. Кроме этого, включение в рассмотрение нового значимого параметра означает, что он должен присутствовать как у реального хоста, так и у ЛИС, а значит, выбор Γ влияет и на глубину имитации ловушкой реальной системы. Можно сказать, что задав пространство Γ, мы, фактически, задаём степень реалистичности модели.

Кроме того, сеть обладает определённой структурой. В пределах широковещательного домена имеется гораздо больше возможностей для атаки, чем вне его (например, легко становится организовать атаку «человек посередине» с помощью подмены служебного трафика *ARP*). Поэтому целесообразно ввести отношение принадлежности одной подсети на *N*. Обозначим это следующим образом:

Для любых $\forall n_1, n_2 = N$ запись $n_1 S n_2$ будет означать, что хосты n_1 и n_2 принадлежат одной подсети.

На рисунке 1 хост n_1 находится в отношении принадлежности одной подсети с хостами n_2 и n_3, последний из которых представляет собой маршрутизатор *R1* (под хостом будем понимать операционную среду, а не интерфейс с адресом на третьем уровне модели *OSI*).

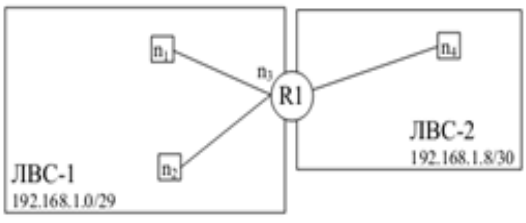

Рисунок 1 - Пример разбиения сети на подсети.

На рисунке 2 показана схема отношений между хостами сети.

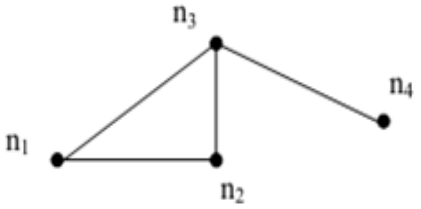

Рисунок 2 - Отношение S между хостами сети.

Отношение *S* задаёт покрытие множества *N* подмножествами:

$$N = S_1 \bigcup S_2 \bigcup ... \bigcup S_k \ , \qquad (4)$$

причём эти подмножества либо не пересекаются, либо их пересечение содержит единственный элемент (который представляет маршрутизатор).

Множество всех подсетей в сети обозначим через *D*. Вырожденный случай, когда сеть не разделена на подсети, записывается в этом случае так:

$$D = \{N\} \ , \qquad (5)$$

Таким образом, сеть с точки зрения топологии может быть представлена в виде неориентированного графа, в вершинах которого находятся хосты, а дуги задаются отношением *S*. Если существует путь из одной вершины графа в другую, то возможна последовательная атака одного хоста через другой.

Так как выше рассматривалась модель сети независимо от наличия или отсутствия в ней ЛИС, то систему *Honeypot* можно представить как подмножество (подграф) сети *N*:

$$H \subset N \ , \qquad (6)$$

Атаку можно представить в виде повторяющейся последовательности действий:

1. выбор цели;
2. попытка эксплуатации уязвимости.

При этом выбор цели происходит случайно, а вероятность выбора того или иного хоста зависит от некоторых условий. Состав этих условий, как и Γ, определяет глубину детализации модели. В простейшем случае всякие условия отсутствуют, и выбор того или иного хоста равновероятен. В этом случае, очевидно, для того, чтобы обеспечить максимальную вероятность атаки на ЛИС, нужно обеспечить максимальное количество ЛИС в сети. Однако ясно, что на вероятность выбора цели для атаки влияет множество факторов, среди которых:

- особенности топологии;
- наличие подконтрольных атакующему систем в сети;
- прочие факторы, например, условие невозможности атаки реальной системы с использованием ЛИС в качестве промежуточного звена.

С учётом введённых понятий, можно описать процесс последовательных атак на сеть, включающую в себя обманные системы наряду с реальными (целевыми). Для этого необходимо построить следующие модели:

- модель сети (в виде графа, задав N, Γ, разбиение на подсети S_i);
- модель системы ловушек (подграф $H \subset N$);
- модель атак (задав систему условий на вероятность выбора того или иного хоста сети);
- модель атакующего (задав субъективные параметры, такие как алгоритм выбора цели, вероятность успешной эксплуатации уязвимости, частоту попыток атаки, и т. д.).

Критерий оптимальности для конфигурации *Honeypot* выражается следующими условиями:

- распределение наборов параметров из Γ на N должно быть одинаково с распределением на $N \setminus H$;
- в каждый момент времени вероятность выбора ЛИС в качестве следующей цели для атаки должна быть максимально возможной.

Используя данную схему, можно поставить задачу нахождения оптимальных параметров множества *H* как подмножества *N* при фиксированных параметрах сети.

Предложенная модель может служить обобщением для других, более простых моделей, описывающих процесс последовательных атак на сеть с ловушками.

В качестве примера рассмотрим модель, предложенную в работе [6]. В этой модели сеть представлена множеством хостов, среди которых есть ЛИС. На сеть совершаются последовательные атаки. Выбор цели при атаке осуществляется с равной вероятностью среди всех хостов сети.

Данная модель была построена для того, чтобы дать ответ на вопрос об оптимальном количестве ЛИС в сети, состоящей из $|N|$ хостов.

Таким образом, можно записать:

$$\begin{cases} N = \{n_i\}_{i=1..d}, \\ D = \{N\}, \\ P_k(n_i) = P_k(n_j) \forall i, j, k \end{cases} \quad (7)$$

где $P_k(n_j)$ - вероятность того, что на *k*-том шаге вероятность будет выбрана для атаки цель $n_i \in N$.

Единственным параметром сети в рамках данной простой модели является количество хостов в ней. Соответственно, единственный параметр системы ловушек — это также количество ЛИС.

В силу отсутствия разбиения на подсети, а также условия $\Gamma = .\varnothing$, первая часть критерия оптимальности конфигурации ЛИС выполнена для любого *H*. Вторая часть критерия оптимальности в рамках действующей модели сводится к тому, что количество ЛИС в сети должно быть максимально возможным. В условиях равной вероятности выбора целей для атаки, мощность множества *H* должна быть максимальной.

Формально, оптимальную конфигурацию *Honeypot* в рамках данной модели можно записать так:

$$|H| = \max \ , \quad (8)$$

где максимум определяется количеством доступных компании адресов и материальными возможностями для заполнения адресного пространства ложными системами.

Данная конфигурация понимается в работе [6] как «идеальная», а результаты, полученные в ней отличаются от приведённых выше в силу того, что дополнительно накладывалось условие минимальности затрат на поддержку ЛИС.

Выводы

В данной статье описаны принципы построения модели систем *Honeypot*. Данные принципы были введены на основании общепринятых требований к обманным системам (факторы маскировки и привлекательности). В данной работе автором были использованы элементы методов теории множеств и теории вероятности для формального описания основных понятий системы *Honeypot*.

В работе показано, что формальная модель, основанная на описанных принципах, расширяет предложенную ранее модель и даёт предсказуемые результаты.

СПИСОК ЛИТЕРАТУРЫ

1. Dunnigan. J.F. Victory and Deceipt: Dirty Tricks at War [Текст] / J.F. Dunnigan, A.A. Nofi. – NY: William Morrow and Co., New York, 1995. – 350с.:ил.
2. Spitzner, L. Dynamic Honeypots [Электронный ресурс] / L. Spitzner. – режим доступа: http://www.securityfocus.com/infocus/1731.
3. Котенко И.В. Обманные системы для защиты информационных ресурсов в компьютерных сетях [Текст] / И.В. Котенко М. В. Степашкин. // Труды СПИИРАН, Вып. 2, т. 1. — СПб.: СПИИРАН, 2004. С. 211 – 230.
4. Hecker, C. Proceedings of the 10th Colloquium for Information Systems Security Education [Текст] / C. Hecker, K. Nance, B. Hay // ASSERT Center, University of Maryland, University College Adelphi, 2006.-29с.
5. Гладких А.А. Базовые принципы информационной безопасности вычислительных сетей [Текст] / А.А. Гладких, В.Е. Дементьев, УлГТУ, 2009, 168с.:ил.
6. Алейнов Ю.В. Применение динамических систем пассивной регистрации сетевых атак для обеспечения безопасности компьютерных сетей [Текст] / Алейнов Ю.В., Бондаренко В.В. // Сборник "Вычислительная техника и новые информационные технологии". Уфа: УГАТУ, 2011. С. 126-131.

*Аспирант **Алейнов Ю.В.** тел. 8-917-812-95-68, aleinov@gmail.com - кафедра безопасности информационных. систем Самарского государственного университета*