Математическое моделирование

УДК 510.58:004.056.53

ОБ ОПТИМАЛЬНОЙ КОНФИГУРАЦИИ ОБМАННЫХ СИСТЕМ В КОМПЬЮТЕРНОЙ СЕТИ ПРЕДПРИЯТИЯ

Ю. В. Алейнов, И. Н. Саушкин

[1] Самарский государственный университет,
Россия, 443011, Самара, ул. Академика Павлова, 1.
[2] Самарский государственный технический университет,
Россия, 443100, Самара, ул. Молодогвардейская, 244.

E-mails: aleinov@gmail.com, saushkin@samgtu.ru

Рассмотрена проблема поиска оптимальной конфигурации обманных систем для защиты корпоративной сети. Авторами проанализирован стандартный подход к решению данной задачи. Предложена математическая модель компьютерной сети предприятия, содержащей ложные цели. Предложен критерий оптимизации конфигурации ложных целей, учитывающий динамику атакующего воздействия на сеть. Приведены постановка и решение соответствующей задачи оптимизации. Предложена процедура нахождения оптимальной конфигурации. Даны рекомендации по применению полученных результатов на практике.

Ключевые слова: *защита компьютерной сети, ложные системы, адаптивные системы управления, математическая модель, задача оптимизации.*

Введение. В настоящее время всё более актуальной становится проблема защиты компьютерной сети от атак извне. В числе прочих механизмов защиты можно выделить внедрение ложных (обманных) систем для отвлечения ресурсов атакующего от реальных систем и сбора информации об атаке [1]. Собранная такой обманной системой информация способна облегчить процесс генерации сигнатур для систем обнаружения (предотвращения) вторжений. Одна из главных особенностей ложных систем заключается в том, что весь трафик, регистрируемый такой системой, с высокой вероятностью является злонамеренным, что позволяет значительно уменьшить объём исходных данных для анализа [2]. Ложные системы широко и успешно применяются для целей исследования активности злоумышленников, используемых ими инструментов, а также тенденций возникновения вспышек вредоносной активности. Однако это применение обычно ограничено рамками специальных исследовательских лабораторий [2]. Вместе с тем идею размещения в сети ложных систем, очевидно, можно использовать и для защиты сети отдельной организации. Использование таких систем в связке с корпоративными системами обнаружения вторжений позволяет задействовать очень качественный детектор аномалий, повышая общий уровень защищенности сети организа-

Юрий Викторович Алейнов, аспирант, каф. безопасности информационных систем.
Иван Николаевич Саушкин (к.ф.-м.н.), начальник управления, управление информатизации и телекоммуникаций.

ции [3]. Любая активность ложной системы — аномалия, требующая внимательного рассмотрения и позволяющая своевременно принять меры против проводимого или планируемого вторжения.

Основная трудность, которая возникает при попытке использовать обманные системы для защиты инфраструктуры предприятия, заключается в том, что для их развертывания и поддержки нужны специалисты, обладающие специальными познаниями в области защиты от компьютерных атак. В настоящее время ряд исследований направлен на создание автоматизированной системы управления ложными объектами, способной анализировать параметры сети организации и в зависимости от них настраивать параметры ловушек [4–6]. В большинстве работ основное внимание авторы уделяют различным способам сбора информации о параметрах защищаемой сети, разворачивания системы ложных сетевых объектов и получения информации об атаках. Решение вопроса о нахождении оптимальной конфигурации обманных систем обычно сводится либо к дублированию объектов защищаемой сети, либо к воспроизведению её структуры в некотором масштабе (меняется общее количество узлов, а распределение их характеристик, таких как используемая операционная система и программное обеспечение, остаётся неизменным) [5,6].

Однако при расчёте конфигурации системы ложных объектов необходимо также учитывать особенности распространения атакующего воздействия по сети. Количество ложных целей, имитируемые ими операционные системы, поддерживаемые ими сетевые сервисы и прочие параметры должны выбираться таким образом, чтобы обеспечить максимальную вероятность выбора атакующей стороной ловушки в качестве цели для атаки при любых исходных параметрах сети. Если, имея некоторую модель действий злоумышленника, можно предсказать его поведение в следующий момент времени, то, владея этой информацией, можно динамически изменять конфигурацию ложных целей в сети с тем, чтобы повысить эффективность использования ложных объектов.

В данной статье предлагается способ определения оптимальной конфигурации ложных целей в зависимости от параметров защищаемой сети, а также от динамики атакующего воздействия, которому подвергается сеть.

1. Модель компьютерной сети, содержащей ложные системы. Прежде всего необходимо формализовать понятие конфигурации ложных систем. Традиционно под ней понимается количество ложных систем, их взаимное расположение в сети и значение некоторого вектора параметров, связанного с каждой из них [6]. Размерность этого вектора, а также конкретные параметры, определяемые им, задаются в процессе построения модели в зависимости от требуемой глубины детализации [7]. Под системой (реальной или ложной) при этом понимается, как правило, хост, работающий под управлением конкретной операционной системы, предоставляющий некоторое количество сетевых сервисов и имеющий сетевой адрес.

Как правило, сетевые атаки направлены на те или иные уязвимости, содержащиеся в системном и прикладном программном обеспечении (СПО и ППО). Очевидно, подавляющее большинство существующего кода относится к ППО, при этом данный код зачастую не проходит должного контроля и содержит в себе массу уязвимостей. Поэтому можно считать, что большин-

ство сетевых атак направлено на эксплуатацию уязвимостей, содержащихся в различных прикладных программах [8].

Будем рассматривать защищаемую сеть как множество целей для атаки. При этом любое обращение к ложному объекту будем считать атакой. Поток событий, состоящих в обращении к ложным объектам защищаемой сети, будем называть потоком атак. Для средств атакующей стороны характерна специализация на конкретном наборе уязвимостей. Как правило, средство для осуществления атаки (эксплойт), написанный для эксплуатации уязвимости в определённом приложении, не подходит для другого приложения. Это означает, что в каждом конкретном случае атаке будет подвержен ограниченный набор узлов в сети — только те, на которых запущено приложение с интересующими злоумышленника уязвимостями. С учётом того, что большинство сетевых атак направлено на эксплуатацию уязвимостей в ППО, целью для атаки является экземпляр сетевого сервиса (приложения), работающего на каком-либо хосте в защищаемой сети. Поскольку для атакующего важен лишь факт наличия определённой уязвимости в целевом приложении, можно считать, что цели, являющиеся экземплярами приложений одного типа и одной версии, неразличимы для злоумышленника (при условии, что они не были успешно атакованы).

Пусть S — множество всевозможных типов и версий сетевых приложений, работающих в сети. Тогда сеть в соответствии с вышеизложенным можно задать следующим образом:

$$N = \{(s, x_s)\}_{s \in S}.$$

Здесь x_s — количество целей типа s.

Поскольку S конечно, можно перенумеровать все элементы $s \in S$. Тогда конфигурация сети может быть задана вектором $\vec{X} = (x_1, x_2, \ldots, x_L)$, где его размерность L определяется мощностью множества S и равна ему.

Ложные системы отличаются от настоящих только тем, что они не выполняют никакой полезной работы, а нужны лишь для того, чтобы злоумышленник попытался совершить атаку на них, поэтому множество ложных целей F, очевидно, является подмножеством множества целей сети N: $F \subseteq N$. Так как каждая цель из множества F также имеет определённый тип $s \in S$, множество ложных целей можно также задать вектором $\vec{H} = (h_1, h_2, \ldots, h_L)$. Соответственно, вектор $\vec{R} = (r_1, r_2, \ldots, r_L)$, где $r_i = x_i - h_i$, $i \in [1, L]$, может характеризовать множество реальных целей. Рассмотрим ограничения, накладываемые на h_i.

Обычно в любой производственной сети имеется фиксированное адресное пространство, частично занятое различными системами. Поскольку мы рассматриваем сеть как множество целей (сервисов), мы должны использовать двумерное адресное пространство для их размещения, где адрес состоит из IP-адреса и номера порта приложения. Если организация обладает пространством, состоящим из N_{IP} IP-адресов для размещения своих систем, то формально адресное пространство целей состоит из $65535 \times N_{\text{IP}}$ адресов. Таким образом, если в сети имеется $R_0 = \sum_{i=1}^{L} r_i$ реальных целей, то для размещения в этой же сети ложных целей может быть использовано $65535 \times N_{\text{IP}} - R_0$ адресов. Однако если обратиться к наиболее часто используемой (особенно автоматизированными средствами) методике проведения сетевой разведки,

то окажется, что, хотя адресное пространство третьего уровня (IP-адреса) обычно сканируется злоумышленником полностью в поисках уязвимых систем, диапазон сканируемых портов для каждого IP-адреса намного уже и ограничивается, как правило, стандартным набором портов для конкретного приложения [8]. Исходя из этого, а также из общепринятой практики, в соответствии с которой необходимо настраивать ложные системы таким образом, чтобы они как можно более точно повторяли конфигурацию реальных [7], определим допустимое адресное пространство для размещения ложных систем. Пусть P — множество портов, используемых всеми реальными системами в сети для работы реальных сервисов. Пусть T — множество свободных (не занятых реальными системами) IP-адресов. Тогда адресное пространство для размещения ложных целей в сети можно определить как декартово произведение этих множеств:

$$A_H = P \times T.$$

Пусть $|A_H| = H_0$, тогда по отношению к координатам вектора $\vec{H}$ справедливо

$$H_0 = \sum_{i=1}^{L} h_i. \tag{1}$$

Таким образом, формально задача поиска оптимальной конфигурации множества ложных целей означает поиск оптимального значения вектора $\vec{H}$ относительно некоторого критерия при условии (1).

Модель компьютерной сети, содержащей ложные системы и подвергающейся атакам, может быть описана следующим образом:

- имеется множество целей, разбитое на классы по их типам;
- внешняя среда постоянно генерирует поток атак на эти цели;
- интенсивность потока атак γ меняется со временем;
- интенсивность общего потока атак равна сумме интенсивностей потоков атак, направленных на различные классы целей ($\gamma = \gamma_1 + \gamma_2 + \cdots + \gamma_L$);
- в каждом классе целей существует разбиение на реальные и ложные цели;
- атака на ложную цель в классе S_k повышает вероятность генерации сигнатуры атаки в данном классе p^k_{sign} на неизвестную величину Δp^k_{sign};
- если сгенерирована сигнатура атаки, то соответствующий поток может быть остановлен другими средствами защиты.

2. Постановка задачи оптимизации. Выбор критерия. Сформулировать задачу оптимизации — значит задать некоторую целевую функцию $F(X, H)$, зависящую в нашем случае от конфигурации сети и от конфигурации системы ловушек, а также критерий её оптимизации. При выборе целевой функции необходимо учитывать следующие факторы:

- главная цель размещения обманных систем — определение мотивов злоумышленника, используемых им средств, методов и инструментов, с тем чтобы впоследствии обеспечить генерацию сигнатур зафиксированных атак и более успешное противодействие им;
- заранее нельзя сказать, сколько времени потребуется для производства сигнатуры зафиксированной атаки и данные о каком количестве атак необходимо иметь для этого;

– с другой стороны, ясно, что для генерации пригодной к использованию сигнатуры атаки необходимо иметь некоторый (заранее неизвестный) объем данных об этой атаке.

В имеющемся случае отсутствует возможность определения целевой функции как функции выигрыша или убытков, поскольку как возможный выигрыш от атаки на обманную систему, так и убытки от атаки на реальную систему зависят от очень большого количества факторов. Среди них — особенности самой атаки, распределение приложений по хостам сети, субъективная ценность данных, которые могут быть получены злоумышленником при успешной атаке, действующие политики безопасности, и так далее. В условиях такой неопределённости естественным будет считать, что оптимальная конфигурация ложных систем должна обеспечивать равный поток атакующих воздействий со стороны внешней среды на ложные системы в разных классах. Таким образом, целевой функцией для задачи оптимизации может являться векторная функция, где каждая координата L-мерного вектора задаётся значением интенсивности потока атакующих воздействий на ложные цели соответствующего класса. Критерий оптимизации — равенство всех потоков.

Вычислим значение интенсивности потока атакующих воздействий на ложные системы класса S_k. При этом будем полагать, что реализация ложных объектов такова, что атакующий не может отличить их от реальных хостов. Как уже отмечалось ранее, такое требование является общепринятым при создании систем ложных объектов в сети. Выполнение данного предположения означает, что атаки на реальные и ложные цели распределены равномерно. Значит, вероятность того, что атакована ложная цель класса S_k при условии, что атака произошла на какую-то цель из этого класса, составляет величину

$$P_H^k = \frac{h_k}{h_k + r_k},$$

где h_k — количество ложных целей в классе S_k, а r_k — количество реальных целей в этом же классе.

Пусть γ_k — интенсивность потока атакующих воздействий на цели класса S_k. Равномерное распределение атак на цели из класса S_k значит, что средняя интенсивность атак на ложные цели этого класса может быть записана так:

$$\gamma'_k = \gamma_k \frac{h_k}{h_k + r_k}.$$

Таким образом, целевая функция может быть определена как

$$F(X, H) = (\gamma'_1, \gamma'_2, \ldots, \gamma'_L) = \left(\gamma_1 \frac{h_1}{h_1 + r_1}, \gamma_2 \frac{h_2}{h_2 + r_2}, \ldots, \gamma_L \frac{h_L}{h_L + r_L}\right). \qquad (2)$$

При этом критерий оптимальности может быть записан в виде

$$\gamma'_i = \gamma'_j, \quad i, j \in [1, L]. \qquad (3)$$

Таким образом, задача оптимизации полностью определяется выражениями (2) и (3).

3. Решение задачи оптимизации. Итак, для того чтобы найти вектор $\vec{H} = (h_1, h_2, \ldots, h_L)$, соответствующий оптимальному решению, определяемому целевой функцией (2) и критерием (3), нужно решить систему уравнений относительно переменных h_i, $i \in [1, L]$:

$$\gamma_1 \frac{h_1}{h_1 + r_1} = \gamma_2 \frac{h_2}{h_2 + r_2} = \cdots = \gamma_L \frac{h_L}{h_L + r_L}. \tag{4}$$

Проведём замену переменных (замена и все преобразования проведены с учётом h_i, r_i, $\gamma_i > 0$):

$$\theta_i = r_i / h_i.$$

С учётом этого и после преобразований система (4) запишется в виде

$$\gamma_i(1 + \theta_i) = \gamma_j(1 + \theta_j), \quad i, j \in [1, L].$$

Получившаяся система — система линейных уравнений относительно переменных θ_1, θ_2, …, θ_L. Её ранг равен $L - 1$, а значит, она имеет бесконечное множество решений. Её общее решение можно записать в виде

$$\begin{pmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_L \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ \vdots \\ -1 \end{pmatrix} + C \begin{pmatrix} \gamma_1/\gamma_L \\ \gamma_2/\gamma_L \\ \vdots \\ 1 \end{pmatrix}, \tag{5}$$

где C — некоторая константа. Подставляя выражение для θ и объединяя (5) и (1), получаем уравнение относительно C:

$$\sum_{i=1}^{L} \frac{r_i}{C\gamma_i/\gamma_L - 1} = H_0. \tag{6}$$

Уравнение (6) не может быть разрешено аналитически для больших L. Однако можно получить его численное решение. Действительные решения данного уравнения, очевидно, существуют, причём в силу ограничений на коэффициенты все они неотрицательны. Решив данное уравнение, из (5) можно найти значение вектора $\vec{H}$, удовлетворяющее критерию оптимальности. Обозначим через C_{opt} некоторое решение уравнения (6). Поскольку практическую ценность имеют только неотрицательные значения h_i, нас будут интересовать только такие C_{opt}, которые удовлетворяют следующему условию:

$$C_{\text{opt}} \geqslant \max_{1 \leqslant i \leqslant L} \{\gamma_i/\gamma_L\}. \tag{7}$$

Такие C_{opt} найдутся в случае наличия среди γ_i таких, что $\gamma_i \leqslant \gamma_L$, а это, очевидно, достижимо в силу произвольности выбора γ_L. Однако и при этом условии сразу использовать полученное частное решение системы (4), вообще говоря, нельзя, так как смысл имеют только целые неотрицательные h_i. Пусть $\vec{H}_{\text{opt}}$ — некоторое решение системы (4), для которого все $h_i \geqslant 0$. Необходимо найти такой вектор $\vec{H}$ с натуральными значениями всех его координат,

что $\vec{H}$ окажется наиболее близок в смысле используемого критерия (3) к полученному решению $\vec{H}_{\text{opt}}$. В качестве соответствующей метрики может быть использован максимум абсолютного отклонения компонент вектора решения от соответствующих координат $\vec{H}_{\text{opt}}$:

$$\rho(\vec{H}, \vec{H}_{\text{opt}}) = \max_{1 \leqslant i \leqslant L} \left\{ |h_i - h_i^0| \right\},$$

где h_i^0, $i \in [1, L]$ — координаты вектора $\vec{H}_{\text{opt}}$.

Таким образом, процедура нахождения оптимальной конфигурации ложных объектов в соответствии с предложенным подходом может быть описана следующим образом:

- решая уравнение (6), получаем набор коэффициентов C_{opt};
- отбрасываем те из них, которые не удовлетворяют условию (7);
- оставшиеся коэффициенты используем для нахождения соответствующих им значений $\vec{H}_{\text{opt}}$;
- выбираем $\vec{H}_{\text{opt}}$, наиболее близкое к текущему распределению ложных объектов в сети;
- находим ближайший к $\vec{H}_{\text{opt}}$ в смысле предложенной выше метрики вектор $\vec{H}$ с натуральными коэффициентами.

Значения интенсивностей γ_i, $i \in [1, L]$ на практике могут быть получены с использованием некоторой вероятностной модели, обученной на статистических данных.

Заключение. Данная статья посвящена решению задачи поиска оптимальной конфигурации ложных систем, внедренных в компьютерную сеть предприятия. Описана модель сети, содержащей ложные системы и подвергающейся потоку атакующих воздействий со стороны внешней среды. Приведена формальная постановка задачи оптимизации с учётом изменяющейся со временем интенсивности потока атак на сеть, предлагается целевая функция и критерий оптимальности. Получено уравнение, решение которого необходимо для получения оптимального распределения ложных целей согласно предложенному критерию, и даны практические рекомендации по использованию полученных результатов на практике.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. *А. В. Лукацкий*, Обнаружение атак. СПб.: БХВ–Петербург, 2001. 624 с. [*A. V. Lukatcky*, Attack detection. St. Petersburg: BHV—Petersburg, 2001. 624 pp.]
2. *L. Spitzner*, Honeypot. Tracking Hackers. Boston: Addison-Wesley, 2003. 429 pp.
3. *И. В. Котенко, М. В. Степашкин*, "Обманные системы для защиты информационных ресурсов в компьютерных сетях" // *Тр. СПИИРАН*, 2004. Т. 2, №1. С. 211–230. [*I. V. Kotenko, M. V. Stepashkin*, "Deception systems for protection of information resources in computer networks" // *Tr. SPIIRAN*, 2004. Vol. 2, no. 1. Pp. 211–230].
4. *C. Döring*, Improving network security with honeypots. Honeypot project: Master's thesis. Darmstadt: University of Applied Sciences Darmstadt, 2005. 123 pp.
5. *C. Hecker, K. L. Nance, B. Hay*, "Dynamic Honeypot Construction" / In: *Proc. of the 10th Colloquium for Information Systems Security Education*. Adelphi, MD: University of Maryland, University College, 2006. Pp. 95–102.
6. *I. Kuwatly, M. Sraj, Z. A. Masri, H. Artail*, "A Dynamic Honeypot Design for Intrusion Detection" / In: *2004 IEEE/ACS International Conference on Pervasive Services (ICPS'04)*, 2004. Pp. 95–104.

7. *Ю. В. Алейнов*, "Принципы построения модели систем HONEYPOT" // *Ползуновский вестник*, 2012. №3/2. С. 36–39. [*Yu. V. Aleinov*, "The principles of construction of model Honeypot systems" // *Polzunovskiy vestnik*, 2012. no. 3/2. Pp. 36–39].
8. *И. Д. Медведовский, П. В. Семьянов, Д. Г. Леонов, А. В. Лукацкий*, Атака из Internet. М.: СОЛОН-Р, 2002. 368 с. [*I. D. Medvedkovsky, P. V. Sem'yanov, D. G. Leonov, A. V. Lukatsky*, Attack from Internet. Moscow: SOLON-R, 2002. 368 pp.]

Поступила в редакцию 26/IX/2013;
в окончательном варианте — 1/XI/2013.

MSC: 90B15, 90B80

OPTIMAL HONEYNET CONFIGURATION IN ENTERPRISE COMPUTER NETWORKS

Yu. V. Aleinov, I. N. Saushkin

[1] Samara State University,
1, Academician Pavlov st., Samara, 443011, Russia.
[2] Samara State Technical University,
244, Molodogvardeyskaya st., Samara, 443100, Russia.

E-mails: aleinov@gmail.com, saushkin@samgtu.ru

The article is devoted to the optimal configuration of honeypots in the enterprise network. It describes the mathematical model of an enterprise computer network with honeypots. The authors analyze the standard ways of setting up honeypot parameters and propose the optimization criterion with regard to the dynamics of the external environment. The optimization problem of configuring decoys is reviewed and the solution of the problem is discussed. Also a procedure of searching optimal honeypot configuration is proposed and recommendations concerning practical appliance of the obtained results are given.

Keywords: *computer network security, honeypots, adaptive systems, mathematical model, optimization.*

Original article submitted 26/IX/2013;
revision submitted 1/XI/2013.

Yuriy V. Aleinov, Postgraduate Student, Dept. of Computer Security.
Ivan N. Saushkin (Ph. D. Phys. & Math.), Chief of Dept., Dept. of Informatization & Telecommunications.