© 2012 г. И.С. МАКСИН,
В.А. МАЛЫШЕВ,
(ФГБОУ ВПО "Шуйский государственный педагогический университет")

РАСПРЕДЕЛЕННАЯ ДИНАМИЧЕСКАЯ HONEYNET (СЕТЬ-ЛОВУШКА) НА ОСНОВЕ СКРЫТОЙ МОДЕЛИ МАРКОВА

В статье рассмотрена и проанализирована технология распределенных многоагентных систем honeynet (сеть-ловушка), как одно из решений вопроса обеспечения информационной безопасности и поддержки комплекса интегрированной защиты сетевых ресурсов и информационных систем образовательных учреждений. Подробно рассмотрена архитектура и особенности сетей-ловушек, приведено описание событий информационной безопасности с использованием моделей Маркова и приведен механизм IDS для анализа и отслеживания инцидентов политики информационной безопасности.

DISTRIBUTED DYNAMYC HONEYNET BASED ON HIDDEN MARKOV MODEL I.S. Maksin (Shuya State Pedagogical University, Koperativnaya 24, Shuya 155908, Russia, E-mail: main@ivasha.ru), V.A.Malishev (Shuya State Pedagogical University, Koperativnaya 24, Shuya 155908, Russia, E-mail: vadim.a.malyshev@yandex.ru). In this paper the technology of distributed multi-agent systems (honeynet), as one of the solutions to information security and support of complex integrated security network resources and information systems of educational institutions is reviewd and analyzed. The architecture and features of honeynets are reviewd in detail, and description of information security events with using Markov models, and IDS mechanism to analyze and track incidents of information security policy is presented.

1. Введение

Управление инновационным образовательным учреждением (ИОУ) на основе комплексного использования информационных и коммуникационных технологий (ИКТ) требует системной постановки и согласованного решения ряда организационных и технологических задач с учетом экономической целесообразности. Автоматизация различных видов деятельности организации на концептуальной платформе корпоративной информационной среды (КИС) обеспечивает рациональную интеграцию информационных и вычислительных ресурсов и позволяет создать мощную сетевую инфраструктуру, отвечающую требованиям международных стандартов.

Одновременно с усилением зависимости от ИКТ, существенно возрастает уязвимость ИОУ по отношению к угрозам информационной безопасности (ИБ). Ослабление внимания к вопросам организации и совершенствования защиты сетевой

инфраструктуры, как показывает практика, неизбежно ведет к нарушению нормальной деятельности ОУ и к существенным экономическим потерям.

Анализ отечественного и зарубежного опыта позволяет сделать вывод, что попытки решить проблему информационной безопасности за счет прямого наращивания средств и организационных мер защиты во всех компонентах ИОУ являются малопродуктивными и, в принципе, не снимают обострившейся проблемы ИБ. Следует признать обоснованным переход к использованию адаптивных методов и средств защиты сетевых ресурсов. В этой связи научное исследование, направленное на создание гибких механизмов защиты информационных ресурсов ОУ, представляется актуальным и своевременным.

При общей тенденции развития сетевой инфраструктуры ОУ и внедрения средств ИКТ во все сферы деятельности по-прежнему наблюдаются относительно низкие темпы реализации инновационных методов и сервисов обеспечения ИБ и способов оперативного управления механизмами защиты (МЗ). [3]

Для решения проблемы информационной безопасности многие исследователи сосредоточили внимание на развитии системы обнаружения вторжений (IDS) и системы предотвращения вторжений (IPS). В настоящие время большинство IDS основано на изучении статистики вторжений, записи паттернов, сопоставлении сигнатур, анализе протокола. Чтобы создать эффективный механизм защиты ресурсов сети, нужно ясно представлять концептуальную модель вторжения и используемые злоумышленником методы. Описание вторжений с помощью детально разработанного механизма мониторинга – один из наиболее распространенных подходов, в котором точное и ясное понимание логики и действий злоумышленника – основание для создания более сильной и прочной IDS. Это также является ключом к обеспечению оптимальной информационной защиты и достижению целей создаваемой системы безопасности. В последнее время IDS включают в себя распределенные многоагентные системы honeynet (HN) для повышения производительности.

Хотя разработано множество IDS, методы сетевых атак развиваются очень быстро, поэтому требования к производительности IDS очень высоки. В особенности, выявление уязвимостей системы и решение проблем информационной безопасности в режиме реального времени является определяющим в реализации корпоративной политики безопасности. Чтобы улучшить производительность IDS и уменьшить информационные потери и риски, требуется идентифицировать процесс вторжения. В данный момент градация параметров, установленная в большинстве IDS, не сопровождается фактором динамической синхронизации и динамическим обновлением. Собранная в результате проверок информация, как правило, является неполной.

С развитием технологий непрерывно развиваются технические возможности злоумышленников. Принимая во внимание то, что вторжение может проходить в несколько этапов, а на каждом этапе используются различные техники, многие исследователи для решения проблемы сетевой безопасности предлагают многоуровневую IDS, основанную на скрытой модели Маркова. В такой модели обнаружение вторжений основано на оптимальном распределении детекторов на каждом уровне. Большинство атак обнаруживается по распределенным каналам или происходит по организованному сценарию. Хотя число детекторов, устанавливается и контролируется каждым агентом, данные о вторжении, собираемые и анализируемые одним агентом обнаружения, не могут быть объединены для совместного анализа. Поэтому перспективные IDS должны обладать возможностью больших систем и давать свидетельство вторжения на основе объединенной информации от распределенных агентов. [4]

2. Архитектура сетей-ловушек

Распределенная многоагентная система HN представляет собой своеобразную «приманку» на серверах сетей и ИС, обладающих повышенной уязвимостью, что и привлекает взломщиков. Цель развертывания HN – сбор информации о вторжениях в интересах выявления точной картины вторжения. Как только ловушка подверглась атаке, срабатывает сигнал о вторжении.

Традиционно HN делятся на два типа по уровню взаимодействия – низкий и высокий, а также проектируются как физические, так и виртуальные. Большинство агентов устанавливаются на таких сетевых узлах и устройствах как область между Интернетом и демилитаризированной зоной или просто устанавливаются на сети учреждения как показано на рисунке 1.

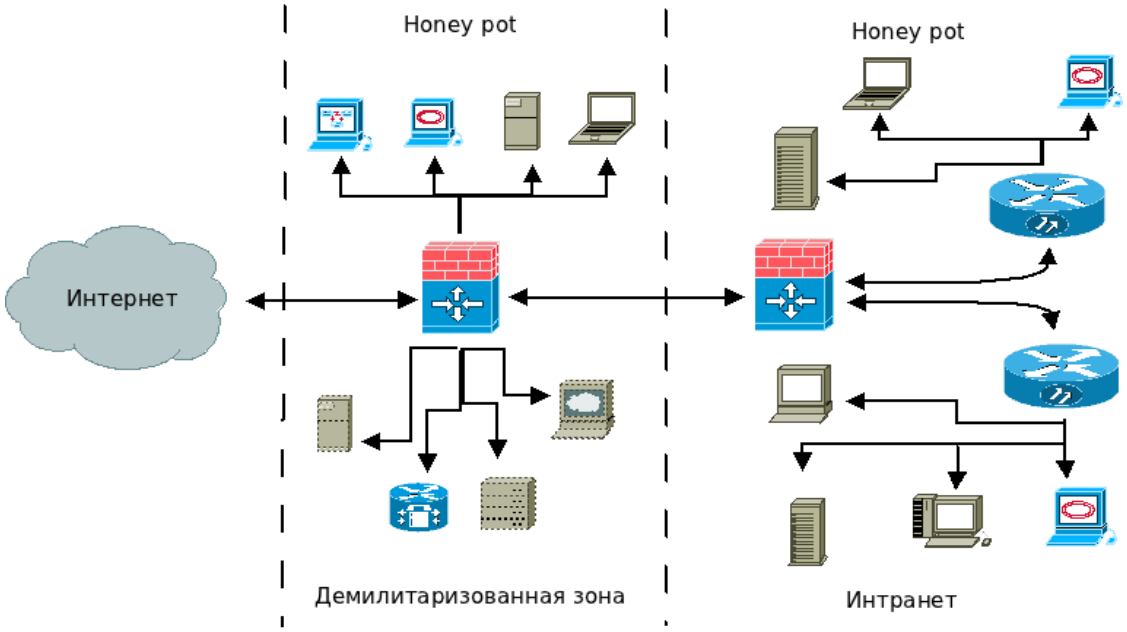

Рисунок 1. Пример расположения агентов в HN.

Перечислим некоторые элементы и базовые параметры сетей-ловушек:

1- архитектура операционной среды.

2 - операционные системы.

3 - открытые и закрытые порты связи.

4 - программы приманки.

5 - важные документы, конфиденциальные файлы, шифрованные файлы и т. д.

6 - привилегированные и общие файлы, каталоги, базы данных и т.д.;

7 - распределение серверов, включая веб-сервер, почтовый сервер, сервер баз данных, и т.д. и сеть всех возможных терминалов конечных пользователей;

8 - специально разработанный путь завлечения злоумышленников в ловушку.

После разработки HN каждому ее элементу присваивается показатель, характеризующий его ценность. Когда система введена в строй все попытки, как исследование коммуникационных портов, файлов, каталогов и баз данных, или попытка взлома пароля и несанкционированного доступа к привилегированным документам будут помечены различными маркерами (подобно таблице 1). Полный

путь каждого отдельного злоумышленника и сопутствующая информация, включая возможные IP адреса, ID пользователя, время и маршрут будут описаны различным количеством маркеров. То есть, большему числу попыток будет соответствовать больший уровень маркеров. С течением времени их уровень на каждом отдельном узле будет уменьшаться. Вся собранная информация исследуется и анализируется для отслеживания действий взлома и анализа поведения злоумышленников.

Уровень маркеров		Индекс	Важность информации	Примечание
Успешная попытка	PH(hc)	1	Строго конфиденциальная и секретная информация	Указывает на успешное посещение злоумышленником высокого уровня конфиденциальной информации
	PH(c)	2	Конфиденциальная и секретная информация	Указывает на успешное посещение злоумышленником уровня конфиденциальной информации
Неудачная попытка	F_PH(hc)	3	Строго конфиденциальная и секретная информация	Указывает на попытку посещения злоумышленником высокого уровня конфиденциальной информации
	F_PH(c)	4	Конфиденциальная и секретная информация	Указывает на попытку посещение злоумышленником уровня конфиденциальной информации
Попытка	PH(g)	5	Общая не конфиденциальная информация	Указывает на посещение в зоны без запроса и авторизации

Таблица 1. Показатели ценности информации.

С целью более эффективного анализа вторжения, особенно распределенной и организованной атаки, выделяются те цели, которые были подвержены большему числу атак. Кроме того, поскольку большинство взломщиков не использует повторно один и тот же путь к цели, уровень корреляции между различными путями вторжения подвергается интегрированному анализу. То есть, хоть пути и различаются, цель остается одна и та же. [1]

Хотя подобные «сети-ловушки» разрабатываются в настоящее время, их эффективность и частота обновлений все еще неудовлетворительны. Эти недостатки приводят к сложности повторного привлечения злоумышленника. Следовательно, для достижения цели хорошего контроля над вторжениями, необходимо применение к HN ниже перечисленных требований:

1 - каждому отдельному узлу сети присваивается значение, характеризующее уровень его важности по сравнению с остальными. Кроме того, данная градация зависит от числа попыток взлома этих узлов, как показано в таблице ниже.

2 - содержимое каждого узла должно обновляться немедленно после обнаружения с целью повторного привлечения злоумышленника.

3 - безопасность и сетевые параметры должны быть расширены и их класс безопасности должен быть повышен сразу после их обнаружения злоумышленником, чтобы привлечь его снова.

4 - когда установки безопасности возрастают и сетевые параметры изменяются, сопутствующее изменение градации важности узла изменяется синхронно.

Различного рода трудности к привилегированным или секретным данным призвана привлечь взломщиков и оценить их возможности при атаке. Кроме того, существует разный уровень доступа к данным секретной таблице, хранящейся в базе данных.

Уровень сложности взлома	Диапазон веса	Приложения	Примечание
Очень трудный	0.8-1.0	-приложения администрирования. -приложения руководства. -параметры IDS.	Класс высокой конфиденциальности.
Трудный	0.5-0.8	-базы данных. -привилегированные списки доступа. -важная информация.	Класс средней конфиденциальности.
Простой	0.2-0.5	-программы низким уровнем значимости. -открытый доступ с базовой аутентификацией.	Программы с аутентификацией по обычному паролю и простым ID...
Не защищенный	0 -.0.1	-общие файлы и директрии. -общедоступные сервисы.	Не конфиденциальный класс.

Таблица 2. Описание градации важности отдельных узлов сети

3. Многоуровневая скрытая модель Маркова (CMM)

Все цифровые системы могут быть описаны как совокупность N различных цифровых последовательностей {As1, As2,…, AsN}, соответствующих определенной временной выборке. Они также могут быть сопоставлены как модель переходного состояния.

Что касается вторжений, навязчивое поведение определенных узлов и файлов тоже может быть представлена как последовательность. Учитывая результат отслеживания вторжений, трансверсаль вторжения можно рассматривать в как переходное состояние и представить в форме {Asi}, где i – фактор времени, и он лежит в промежутке $0<i<N$. После этого, вся собранная информация, посылаемая агентом обнаружения и развернутыми в сети датчиками, может быть записана и представлена в дискретной модели. Наконец, отслеживание вторжения, то есть переходное состояние каждого

агента, можно рассматривать как набор дискретных последовательностей, описанных временным фактором.

Что касается активности вторжений, трансверсали, полученные после каждого вторжения, могут быть описаны как аналогичные изменения состояния при регулярной частоте дискретизации. Для получения пользы от точного анализа, каждое состояние активности вторжения, которое можно сопоставить с вероятностью (трансверсали), вероятность которой также связана с определенным состоянием. [1]

Что касается состояний и вероятностей, использование моментов времени, связанных с изменением каждого состояния и будет обозначаться как временная последовательность t =1, 2,… и фактическое состояние в момент времени t будет обозначаться как t(i). Полное вероятностное описание вторжения требует описания состояния во время t и всех предшествующих состояний. Поскольку этапы вторжения тесно связаны в таких условиях, описание CMM может быть упрощено до сохранения только предыдущего состояния, как показано в следующем уравнении. [5]

(1) $$P\left[i_t = As_j \middle| i_{t-1} = As_i, i_{t-2} = As_k, \ldots\right] = P\left[i_t = As_j \middle| i_{t-1} = As_i\right]$$

Вероятности временных состояний вторжения определяются следующим уравнением:

(2) $$p_{i,j} = P\left[i_t = As_j \middle| i_{t-1} = As_i\right], 1 \leqslant i,j \leqslant N$$

$$p_{i,j} \geqslant \geq 0$$

(3) $$\sum_{j=1}^{N} p_{i,j} = 1$$

Из приведенного выше уравнения, результатами процесса будут наборы состояний соответствующих моментов времени. Каждое состояние также связано с важным физическим событием информационной безопасности. Таким образом вышеописанный процесс можно считать моделью Маркова. Однако, не все события информационной безопасности могут быть описаны как Марковская модель, то есть использование многомерной Марковской модели было бы более уместно для анализа вторжений. Таким образом каждому агенту обнаружения в предложенном механизме будет присвоена соответствующая модель, описывающий его статус и результаты данных всех датчиков, собираемых под контролем определенного агента будут собираться вместе. После этого все CMM мульти-состояний будут объединены.

Для того чтобы точно выявлять связные или распределенные атаки для последующего анализа используется автокорреляция и перекрестная корреляция.

Измерения автокорреляции: для измерения автокорреляции служат данные, выявленные каждым агентом обнаружения. Входные данные представляют собой полную информацию, полученную от всех сканеров, контролируемых данным агентом обнаружения. Вычисление автокорреляции для каждого независимого агента обнаружения рассчитываются с помощью следующего уравнения:

(4) $$A_r\left(S_i, S_j\right) = \frac{E\left[\left(S_i - m_{S_i}\right)\left(S_j - m_{Sj}\right)\right]}{\sigma_{S_i}\sigma_{S_j}}$$

где Si – это i-й датчик, а msi – среднее значение, а σsi – дисперсия i-го датчика.

Измерение перекрестной корреляции: для измерения перекрестной корреляции используется выявление результатов, полученных каждым агентом обнаружения. Входные данные представляют собой полную информацию, полученную от всех

сканеров, контролируемых данным агентом обнаружения. Вычисление перекрестной корреляции для каждого независимого агента обнаружения рассчитываются с помощью следующего уравнения:

$$C_r\left(A_u\left(S_i\right),A_v\left(S_j\right)\right)=\frac{E\left[A_u\left(A_u\left(S_i\right)-m_{A_u\left(S_i\right)}\right)\left(A_v\left(S_j\right)-m_{sA_v\left(S_j\right)}\right)\right]}{\sigma_{A_u\left(S_i\right)}\sigma_{A_v\left(S_j\right)}} \quad (5)$$

где Au(Si) обозначает i-й датчик, u-го агента обнаружения, а mAu(Si) - среднее значение, σAu(Si) – дисперсия i-го датчика u-го агента обнаружения. [2]

Для точной работы предложенный механизм IDS должен быть скорректирован по ходу работы. Схема обработки данных показана на следующем рисунке:

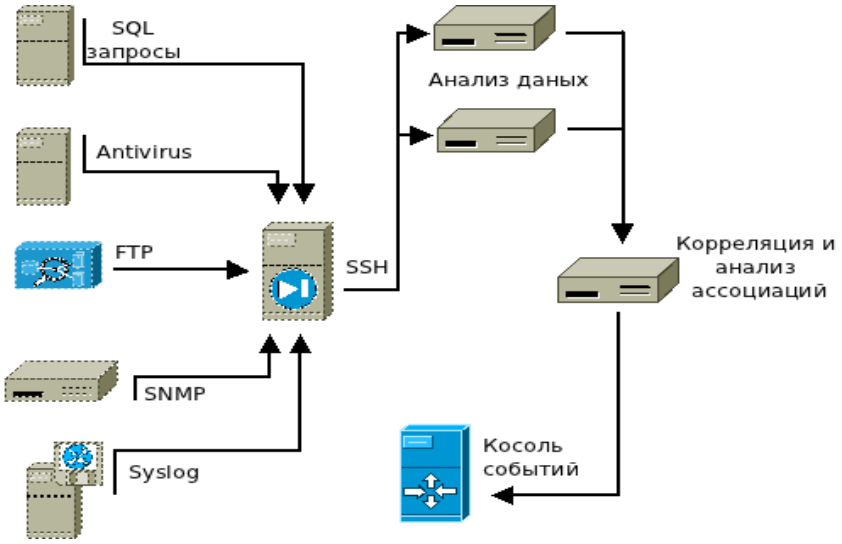

Рисунок 2. Схема обработки данных.

Здесь, вся информация, как журнал логов, данные SNMP (Simple Network Management Protocol) , запросы SQL, записи антивируса и фаервола и т.д. собираются и передаются в SSH режиме к блоку DA (анализ информации) для углубленного и комплексного анализа. Все результаты анализа представляются управляющему блоку для принятия соответствующих мер. [1]

При реализации вирусной атаки, новая сигнатура тотчас добавляется в базу данных для обновления экспертного отдела IDS в режиме реального времени для поддержания системы в оптимальном рабочем состоянии.

Таким образом, перспективными направлением следует считать разработку и внедрение распределенных многоагентных систем HN, входящих в состав комплекса интегрированной защиты сетевых ресурсов и ИС ИОУ. Внедрение HN несет в себе не только усиление защиты систем, но и возможность глубокого изучение инцидентов политики ИБ. Для разработки указанных средств необходимо создание комплекса имитационных моделей информационного взаимодействия злоумышленника и компонентов системы ИБ, позволяющих выявить уязвимости и обосновать оптимальные схемы защиты.

СПИСОК ЛИТЕРАТУРЫ

1. *Chang-Lung T., Min-Hsiung H.*, "Intrusive Behavior Analysis Based on Dynamic Honeynet and Multidimensional Hidden Markov Model" // journal of C.C.I.T., vol. 40, no. 1, may, 2011.
2. *Renuka, P. B., Abraham, A.*, "Hybrid Framework for Behavioral Prediction of Network Attacking Using Honeypot and Dynamic Rule Creation with ifferent Context for Dynamic Blacklisting" // Proceedings of 2010 Second nternational Conference on Communication Software and Networks, pp. 471-476, Singapore, Feb 26-28, 2010.
3. *Надеждин Е.Н., Шептуховский В.А., Максин И.С.* Проблемные вопросы создания защищённой корпоративной информационной образо-вательной среды // Электронный журнал «Информационная среда образования и науки».- М.: ИИО РАО, 2011.- Вып.
4. *Максин И.С., Малышев В.А.,* Концепция многоагентных систем автоматизированной поддержки интегрированной защиты сетевых ресурсов информационной образовательной среды региона. // Материалы V международной научно-методической конференции «Шуйская сессия студентов, аспирантов, молодых ученых». Шуя, 2012.
5. http://ru.wikibooks.org/wiki/скрытые_марковские_модели