УДК 519.248:004.056

Вероятностные модели метаданных

А. А. Грушо*, **Е. Е. Тимонина***, **С. Я. Шоргин***

* *Институт проблем информатики,*
Федеральный исследовательский центр "Информатика и управление"
Российской академии наук,
ул. Вавилова, д.44, стр. 2, Москва, Россия, 119333

Аннотация. Для обеспечения информационной безопасности сетевых взаимодействий ранее было предложено управлять сетевыми соединениями с помощью метаданных. Метаданные содержат информацию о допустимых взаимодействиях задач и расположениях приложений для их решения в распределенной сети. В работе рассмотрена атака на сеть из захваченного хоста, использующая вероятностную модель функционирования метаданных.

Ключевые слова: вероятностные модели, сетевые метаданные, информационная безопасность.

1. Введение

Возможности математического моделирования вычислительных процессов позволяют проводить глубокое тестирование этих процессов, и тем самым гарантировать их правильность в решении задач [1]. Однако пользоваться такими моделями для контроля вычислительных процессов в распределенных информационно-вычислительных системах (РИВС) практически невозможно, т.к. такой контроль является трудоемким процессом. Отсюда возникает идея использования части информации (метаданные), заложенной в математических моделях для быстрого управления соединениями в сети и снижения угроз информационной безопасности. Однако возможны новые угрозы.

В работе рассмотрена атака на сеть из захваченного хоста, использующая вероятностную модель функционирования метаданных.

2. Модель метаданных

Рассмотрим модель РИВС в виде следующей иерархической декомпозиции. Работа, выполняемая РИВС, сводится к решению задач, реализуемых приложениями. Решение задач состоит из трех процессов:

— сбор информации, для решения задачи (исходные данные);
— обработка информации на компьютерах с помощью программного обеспечения (приложений);
— распределение результатов обработки информации.

Информационные технологии можно представить в виде множества задач. Будем объединять информационные технологии и задачи в один верхний уровень иерархической декомпозиции. На нижнем уровне иерархической декомпозиции находятся компьютеры и сеть. В компьютерах находятся информационные ресурсы и решаются задачи.

Компьютер, как узел сети, называется хостом. Различные задачи можно решать на различных хостах сети. Тогда сеть позволяет собирать исходные данные для задач, и распределять результат обработки.

Политика безопасности РИВС требует контроля взаимодействий хостов в сети, который сводится к мониторингу взаимодействий и управлению соединениями. Контроль взаимодействий хостов в сети позволяет снизить угрозы внедрения и распространения вредоносного кода через сетевое оборудование и каналы связи. В работах [2, 3] управление взаимодействиями хостов в сети предлагается реализовать с помощью метаданных.

Предположим, что для РИВС создана математическая модель, определяющая все действия системы для выполнения требуемых вычислений или работ. Например, это может быть диаграмма UML (Unified Modeling Language) [4] коммуникаций, несущая информацию о "линии жизни" [4]. Аналогичная полная информация о выполнении работы может быть представлена диаграммами PERT (Program (Project) Evaluation and Review Technique) [5], а также сетями Петри [6] и структурами со многими связями [7]. От модели требуется, чтобы во всех указанных случаях была возможность вычислить, какой блок O_l, $l = 1, ..., t$, задач и с какими исходными данными надо решать на очередном этапе вычислительного процесса, а также переходы от блока к блоку. Задачу отслеживания переходов от блока к блоку назовем $\mathcal{C}$. Она отслеживает старт каждого блока, обеспеченность блока исходными данными, окончание работы блока и формирование исходных данных для следующего блока.

Обозначим множество задач блока $O_l, l = 1, ..., t$, выполняемой работы через $\Omega_l, l = 1, ..., t$, и рассмотрим на нем множество поименованных подзадач блока, которое можно представить в виде дерева. Корнем дерева является задача всего блока O_l. Введем бинарное отношение порождения (A_i, A_j), где на первом месте в паре фигурирует задача, которая определяет запуск задачи, стоящей на втором месте. Ясно, что это бинарное отношение однозначно определяется моделями блоков. Полученное дерево можно рассматривать как полурешетку, в которой исходные данные могут передаваться через соответствующие верхние грани.

Многие задачи решаются (обеспечены ПО и информационными ресурсами) на тех же хостах, где расположены задачи, их порождающие. Однако часть задач может быть решена только на других хостах сети.

Определим операцию сжатия дерева следующим образом. Если в порождении (A_i, A_j) задача A_j решается на том же компьютере, что и задача A_i, то эти вершины объединяются под именем A_i, а ребро

ликвидируется. Разумеется, задача A_i имеет в памяти название задачи A_j и последующие задачи, попавшие в нее при сжатии.

Лемма. *В результате последовательного применения сжатия к исходному дереву остается сжатое дерево, смежные вершины которого определяют задачи, решаемые на разных хостах.*

Доказательство. Пусть задача A в сжатом дереве решается на хосте $H(A)$. Если смежная с A вершина, находящаяся ниже в сжатом дереве, решается на хосте $H(A)$, то это ребро подлежит сжатию. Поэтому эта смежная вершина не может находиться в сжатом дереве, или она определяет задачу на другом хосте. Лемма доказана.

Множество задач $\Omega_l, l = 1, ..., t$, разбивается на непересекающиеся классы таким образом, что задачи одного класса решаются на одном компьютере. Множество ребер в сжатом дереве будем обозначать $\mathcal{B}_l, l = 1, ..., t$, и называть *метаданными* блока $O_l, l = 1, ..., t$. Задача из $\mathcal{C}$ определяет переход в управлении сети к очередному множеству метаданных.

Для задач, входящих в отношение $\mathcal{B}_l$, определим три дополнительные задачи $\mathcal{M}$, $\mathcal{N}$, $\mathcal{R}$, которые управляют взаимодействиями в сети на основе метаданных $\mathcal{B}_l$. Задача $\mathcal{M}$ распределяет приложения для решения задач между хостами, для простоты будем говорить о распределении задач на хостах. Задача $\mathcal{M}$ определяет бинарное отношение $H(A)$, означающее, что на хосте H может вычисляться задача A. Результаты задачи $\mathcal{M}$ используются задачей $\mathcal{N}$. Задача $\mathcal{N}$ поддерживает связь с каждым хостом, и отвечает за разрешение и предоставление хостам информации по запросу о взаимодействии задач на разных хостах. Разрешение основывается на метаданных $\mathcal{B}_l$. Задача $\mathcal{R}$ строит основной и резервный маршруты по заданию задачи $\mathcal{N}$. Например, $\mathcal{R}$ находится в контроллере сети SDN (Software-Defined Network).

Пусть на хосте $H(A)$ легально запущена задача A. Задача A_1 находится в отношении порождения (A, A_1) и расположена на другом хосте. На каждом хосте H есть агент с криптографическими средствами и ключом $k(H)$ для связи с хостом $H(\mathcal{N})$. Причем для каждого H соединение с $H(\mathcal{N})$ поддерживается постоянно.

Для обращения к задаче A_1 задача A через агента хоста $H(A)$ связывается с задачей $\mathcal{N}$, которая определяет наличие порождения (A, A_1). Тогда на хост $H(A_1)$ через агента этого хоста направляется информация о необходимости соединения с $H(A)$, ключ $k((A, A_1))$ для защиты этого соединения, идентификатор, порт, время. Аналогичная информация направляется на хост $H(A)$. После выполнения задачи A_1 задача A получает результаты от A_1, а соединение $H(A)$ с $H(A_1)$ разрывается.

Алгоритм управления сетью с помощью метаданных однозначно определяет порядок решения задач в каждом блоке, т.к. переход на другие хосты однозначно определен решаемыми задачами, которые в свою очередь определяются выполнением задач всего блока.

Рассмотрим вопросы информационной безопасности в данной схеме. Ранее в [2, 3] отмечалось, что с помощью метаданных удается избегать ряда существенных атак. Однако угроза захвата хоста была рассмотрена не полностью.

Предположим, что противник захватил хост, содержащий задачи из множества Ω_l. Рассмотрим атаку на сеть, управляемую метаданными $\mathcal{B}_l$. Пусть в дереве задач Ω_l задача A находилась на захваченном хосте $H(A)$, и до сжатия порождала затем поглощенную задачу A_1. Теперь противник при необходимости перехода от A к A_1 инициирует запрос к задаче $\mathcal{N}$ о сетевом соединении с задачей A_1. Естественно, в метаданных $\mathcal{B}_l$ нет разрешения на такую связь через сеть. Тогда хост $H(A)$ получает отказ в инициации соединения, и вычислительный процесс останавливается.

Допустим, что возможен повторный запрос, который отклоняется по той же причине. Выход из тупика возможен только при обращении к задаче $\mathcal{M}$, которая через задачу $\mathcal{N}$ сообщает задаче A о необходимости решать задачу A_1 на хосте $H(A)$. Если задача A опять обращается к задаче $\mathcal{N}$ за разрешением на соединение с хостом $H(A_1)$, то задача $\mathcal{M}$ меняет статус задачи A_1 и ее соединений таким образом, чтобы разместить ее решение на другом хосте. В этом случае метаданные $\mathcal{B}_l$ меняются, и запрос может быть удовлетворен.

Однако это единственное решение выхода из тупика останавливает работу сети на реконфигурацию и построение новых маршрутов. Если для простоты предположить, что все временные затраты на запросы и ответы распределены экспоненциально с параметром λ, а время реконфигурации распределено экспоненциально с параметром $\lambda_1 \ll \lambda$, то среднее время задержки вычислений равно $\frac{8}{\lambda} + \frac{1}{\lambda_1}$. Здесь число 8 определяется четырехкратным повторением протокола пересылки данных между захваченным хостом и хостом $H(\mathcal{N})$ (отдельно туда и отдельно обратно).

Атака может быть повторена с другими задачами, и только анализ компьютерного мониторинга на хосте $H(A)$ может определить причину сбоя в этом хосте.

3. Заключение

Управление соединениями хостов с помощью метаданных [2,3] дает возможность решения многих проблем информационной безопасности. Однако это не исчерпывает множества угроз для РИВС.

В представленной работе рассмотрена возможность захвата хоста противником, и рассмотрена вероятностная модель атаки на сеть в этом случае. Хотя в статье приведен метод противодействия атаке на сеть, управляемую метаданными, вероятность больших временных задержек остается высокой.

Благодарности

Работа поддержана грантом РНФ № 16-11-10227.

Литература

1. *Самуйлов К. Е., Чукарин А. В., Яркина Н. В.* Бизнес-процессы и инфофрмационные технологии в управлении телекоммуникационными компаниями. – М.: Альпина Паблишерз, 2009.
2. *Grusho A. A., Timonina E. E., Shorgin S. Ya.* Modelling for Ensuring Information Security of the Distributed Information Systems // Proceedings of 31th European Conference on Modelling and Simulation ECMS 2017, May 23rd–29th, 2015, Budapest, Hungary. – Digitaldruck Pirrot GmbHP, Germany, 2017. — P. 656–660.
3. *Grusho A., Grusho N., Zabezhailo M., Piskovski V.,Timonina E.* Information Security of SDN on the Basis of Meta Simulation // Proceedings of 7th International Conference on Mathematical Methods, Models, and Architectures for Computer Network Security (MMM-ACNS 2017), Warsaw, Poland, 2017 (to be published).
4. *Буч Г., Якобсон А., Рамбо Дж.* UML. – Классика CS.– 2-е изд. / Пер. с англ.: Под общей редакцией проф. С. Орлова. – СПб.: Питер, 2006.
5. *Танаев В. С., Шкурба В. В.* Введение в теорию расписаний. – М.: Наука, 1975.
6. *Питерсон Дж.* Теория сетей Петри и моделирование систем / Пер. с англ. – М.: Мир, 1984.
7. *Кнут Д.* Искусство программирования для ЭВМ. Т.1: Основные алгоритмы / Пер. с англ. – М.: Мир, 1976.

UDC 519.248:004.056

Probabilistic Models of Meta Data

A. A. Grusho*, E. E. Timonina*, S. Ya. Shorgin*

** Institute of Informatics Problems,*
Federal Research Center "Computer Science and Control"
of the Russian Academy of Sciences,
Vavilova str. 44, Moscow, 119333, Russia

For support of information security of network interactions earlier a control of network connections by means of meta data was offered. Meta data contain information on admissible interactions of tasks and positions of applications for their decision in a distributed network. In the paper the attack to a network from a captured host using a probability model of meta data is considered.

Keywords: probability models; network security; control meta data.