Информационные технологии

УДК 004.492.3

Распределённая система обнаружения и предотвращения сетевых атак на системы облачных вычислений

А. А. Кондратьев

Федеральное государственное бюджетное учреждение науки
Институт программных систем им. А.К. Айламазяна Российской академии наук
ул. Петра I, д. 4а, с. Веськово, Переславский район
Ярославская область, Россия 152021

Рассматривается проблема обнаружения сетевых атак на системы распределённых и облачных вычислений. Целью является обнаружение и предотвращение как классических, так и распределённых сетевых атак типа «отказ в обслуживании» (DoS, DDoS). В работе выделен ряд проблем различных популярных систем облачных вычислений, которые представляют опасность не только получения доступа к пользовательским данным, но и могут привести к нарушению целостности и работоспособности вычислительной системы. В качестве решения предлагается разработка системы обнаружения и предотвращения атак. Система состоит из нескольких компонентов, направленных на выполнение различных функций: обнаружение атак, их предотвращение, обеспечение взаимодействия компонентов, управление, хранение данных. Основные инструменты и алгоритмы распознавания основаны на методах искусственного интеллекта и теории вероятности. В разрабатываемом решении используются подходы к распознаванию атак, альтернативные сигнатурному. Представлены преимущества и недостатки используемого подхода. В работе описана архитектура и механизмы функционирования предлагаемого решения. Приведены описания задач и функционала разработанных компонентов. В заключение представлены результаты тестирования экспериментального образца на различных типах сетевых атак на специально подготовленном стенде.

Ключевые слова: облака, защита, сетевая атака, искусственный интеллект.

Введение

По данным исследований ассоциации производителей программного обеспечения (Business Software Alliance, BSA) от 22 февраля 2012 г. Россия занимает 16 место среди 24 стран в новом рейтинге государственного регулирования, влияющего на рост облачных вычислений. В современном мире происходит их активное развитие как основных средств работы и проведения «тяжёлых» вычислений. Стремительное развитие распределённых IT-систем сопровождается ростом количества сетевых атак, направленных как против отдельных компьютерных систем, так и вычислительных сетей в целом. Страны с самой развитой политикой в области распределённых вычислений, такие как Япония, Германия и США, заняты созданием фундаментальной правовой базы, позволяющей поддерживать рост облачных вычислений. В первую очередь производится установление стандартов в области понятий и структуры. Чёткого описания стандарта построения таких систем на данный момент не существует, и уже возникли вопросы обеспечения безопасности распределённых систем. Проблема защиты подобных систем до сих пор является открытой для ведущих игроков IT-рынка. В ходе проведённых исследований рынка существующих решений, предназначенных для построения систем облачных вычислений, был выделен ряд проблем, связанных с построением системы защиты:

Статья поступила в редакцию 13 сентября 2013 г.

Работа выполняется по программе фундаментальных исследований ОНИТ РАН «Архитектурно-программные решения и обеспечение безопасности суперкомпьютерных информационно-вычислительных комплексов новых поколений» в рамках проекта № 2.3 «Обнаружение и предотвращение распределённых сетевых атак на высокопроизводительные системы облачных вычислений на основе отечественных аппаратно-программных комплексов семейства “СКИФ”».

– необходимость обеспечения защиты всей сетевой инфраструктуры как между вычислительной сетью и клиентом, так и защиты самих элементов сети;
– наличие специфических уязвимостей, характерных для различных реализаций сервиса;
– отсутствие общепринятых стандартов построения облачных систем.

Данные проблемы связаны в основном со сложной архитектурой защищаемой системы. В отличие от защиты персонального компьютера, где в качестве объекта выступает отдельное устройство с определённой архитектурой, в данном случае объектом служит набор вычислительных узлов, связанных в многоуровневую сеть, сегменты который могут быть объединены посредством сети Internet. Вследствие этого злоумышленнику открывается гораздо больше целей и возможностей для проведения атаки. При этом существует опасность не только получения доступа к данным, но и встаёт вопрос о нарушении целостности и работоспособности вычислительной системы. Например, если система обеспечивает обработку большого количества клиентов с динамической балансировкой, то нарушение механизмов взаимодействия между узлами приведёт к снижению эффективности работы сервиса или его полной неработоспособности. Для подробного изучения вышеописанных проблем и поиска их решений в рамках настоящей работы проводится разработка программного комплекса для защиты распределённых вычислительных систем на основе методов искусственного интеллекта и теории вероятности как альтернатива сигнатурному подходу. Проект предполагает разработку:
1) сетевого анализатора для выявления атак;
2) механизма взаимодействия системы защиты в облачной архитектуре;
3) данного средства управления системой защиты.

1. Архитектура системы защиты

В ходе исследований разработан экспериментальный образец интеллектуальной программной системы защиты облаков («ЭО ИПС»). Общая архитектура «ЭО ИПС» представлена на рис. 1.

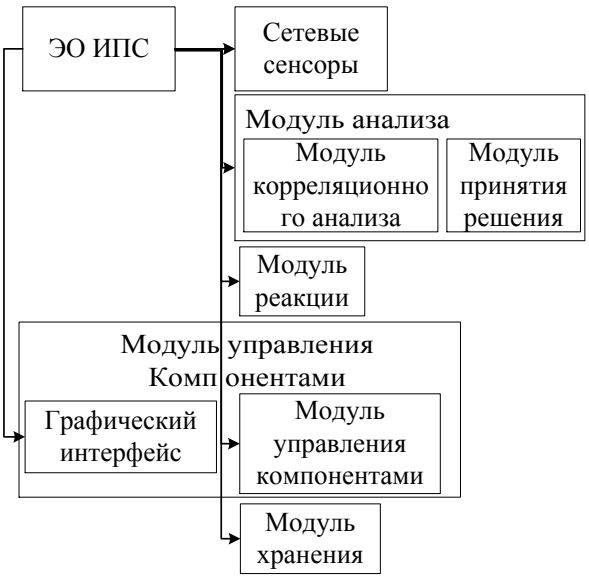

Рис. 1. Архитектура «ЭО ИПС»

Экземпляры системы размещаются во всех узлах распределённой конфигурации и анализируют трафик, взаимодействуя друг с другом. Система защиты разделена на несколько модулей, выполняющих задачи: анализа сетевого трафика, принятия решений, корреляции, реакции, хранения, управления компонентами.

Модуль анализа сетевого трафика построен на основе программных сетевых сенсоров. На каждую сетевую шину защищаемого объекта установлено по одному сенсору. Они выделяют информативные признаки, характеризующие состояние защищаемой системы и пакетов данных, циркулирующих в конкретной сети. Информативные признаки могут быть разделены на шесть групп:

– выделяемые из заголовка сетевого пакета;
– выделяемые из информационной части пакета на основании экспертных знаний либо как информация о состоянии контролируемого узла;
– признаки, значения которых вычисляются как статистика за прошедшие две секунды сетевой активности, для TCP-протокола;
– признаки, значения которых вычисляются как статистика за последние 100 соединений;
– выделяемые на основе анализа HTTP-запросов;
– признаки, значения которых вычисляются как статистика за прошедшие две секунды сетевой активности, для SNMP-протокола.

Первые четыре группы информативных признаков выявляются с помощью готового алгоритма анализа различных сетевых атак, представленного разработчиками базы KDD-99, находящейся в свободном доступе. Данный набор признаков описывает основные параметры сетевого обмена. Следующая группа признаков предназначена для выявления некоторых видов http-exploit-ов — на основе предложений Danforth M. (Models for Threat Assessment in Networks, 2006). Последняя группа отвечает за проверку нарушений стандарта SNMP-протокола.

Для выявления аномальной сетевой активности в модуле принятия решения используется механизм двухуровневой классификации сетевых записей (рис. 2).

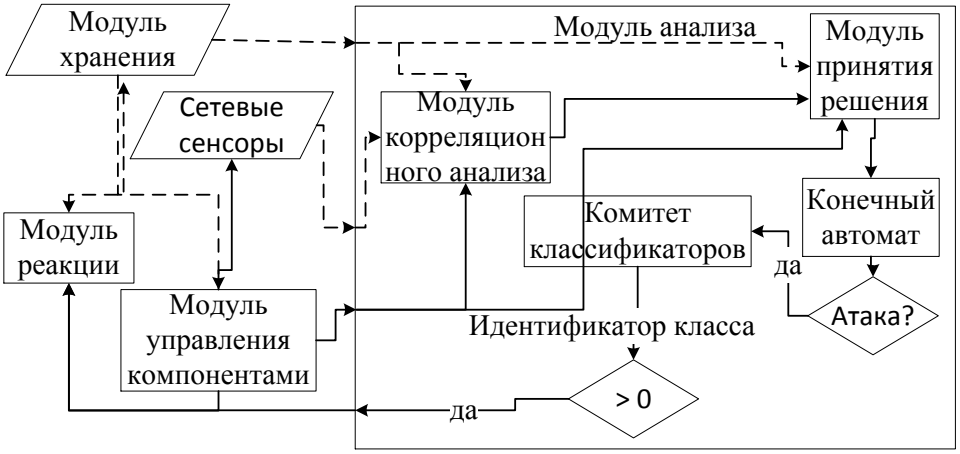

Рис. 2. Механизм обработки сетевых записей

В более ранних исследованиях проработаны вопросы интеллектуальной обработки сетевого трафика с помощью нейронных сетей [1–4]. Были проведены успешные испытания и достигнуты высокие показатели качества распознавания типовых сетевых атак, что позволило использовать данный подход в последующей работе [5]. Основным недостатком нейронных сетей является низкая скорость распознавания. Для повышения скорости распознавания был предложен алгоритм быстрого разделения сетевого трафика на «безопасный» и «подозрительный». Данный алгоритм основывается на использовании конечных автоматов, которые строятся с помощью генетически обучаемых алгоритмов [6]. Эти наработки послужили основой для создаваемого классифицирующего модуля системы.

На первом уровне находится классификатор на основе конечных автомата, разделяющего сетевые записи на два класса: «норму» и «подозрение на атаку» [6].

Второй уровень представляет собой комитет, включающий:

– классификатор на основе метода опорных векторов;

– нейросетевой классификатор на основе многослойной нейронной сети прямого распространения, обучаемой по методу Левенберга–Марквардта;
– классификатор на основе расстояния Евклида–Махаланобиса.

Каждый классификатор обучается для распознавания не только классов атак, но и «нормы». Необходимость в распознавании «нормы» возникает из-за недопустимости ложных срабатываний защитных механизмов. Комитет в данном случае — это набор классификаторов, который принимает общее решение на основе подсчёта количества голосов от каждого из классификаторов, поданных за тот или иной вид сетевой активности. Если количество голосов за «норму» совпадает с количеством голосов за какой-либо класс атаки, то сетевая запись относится к «норме». Если количество голосов совпадает для разных классов атак, то сетевая запись относится к классу неизвестных/смешанных атак. Применение автоматных моделей с парой выделенных состояний позволила надёжно решить задачу определения факта атаки, а использование двухуровневой классификации — существенно повысить скорость обработки сетевого трафика в нормальных условиях.

Отличительной особенностью использования нейросетевых классификаторов и генетически генерируемых автоматов является принцип их формирования. Все они требуют выполнения этапа обучения. Предложенный подход имеет положительные и отрицательные стороны. К отрицательным особенностям относится длительное время подготовки классификатора к работе. Для обучения требуется заранее подготовленная выборка, и качество распознавания напрямую зависит от её состава. Чем больше вариативность атак в выборке, тем большую точность классификации сможет обеспечить комитет. Однако в отличие от сигнатурного подхода, который требует наличие эксперта для определения сигнатуры атаки, данный метод предполагает автоматический механизм выделения информативных признаков из обучающей выборки. Подборку обучающей выборки может выполнить пользователь, не обладающий экспертными знаниями.

Модуль корреляции на основе методов теории вероятности производит ранжирование сетевых признаков по значимости, а также оценивает результаты, полученные модулями принятия решения из разных узлов облака [5]. Результаты ранжирования позволяют изменить порядок входного набора признаков для автомата. Данные действия связаны с различным уровнем значимости признаков для различных сетевых атак, что существенно ускоряет процесс распознавания факта атаки на первом уровне классификации трафика.

Модуль реакции на разные типы атак отключает узел, блокирует передачу и выдаёт предупреждающее сообщение с помощью графического интерфейса.

Основной задачей модуля хранения является предоставление информации, необходимой для обеспечения работоспособности системы защиты. К такой информации, например, относятся настройки модулей и классификаторов. Модуль предоставляет механизмы для сохранения признаков сетевого трафика и, в связи с ориентированностью на распределённую архитектуру, организует синхронизацию настроек и базы знаний экземпляров системы, расположенных на разных узлах. Важной функцией модуля является ведение журнала событий.

Модуль управления компонентами обеспечивает локальное взаимодействие модулей; взаимодействие экземпляров защиты, расположенных на разных узлах (рис. 3); синхронизацию системы защиты и предоставляет графический интерфейс администратору для управления и отображения информации. Он производит инициализацию модулей, передачу настроек, а также следит за целостностью системы, сообщая о неисправностях на этапе инициализации. Основной особенностью модуля является предоставление возможности централизованного управления системой защиты. При этом синхронизация и поддержка нескольких точек «входа» необходимы на случай выхода части узлов облачной системы из строя. Выход из строя одного из компонентов, не влияющего на функциональность облака в целом, не должен влиять на работоспособность системы защиты.

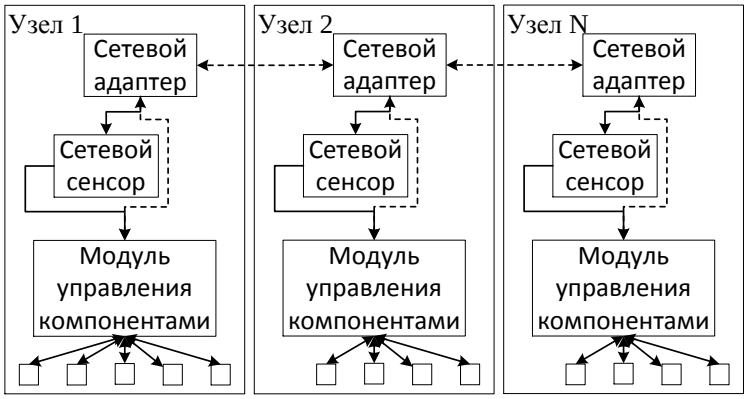

Рис. 3. Механизм обработки сетевых записей

2. Порядок функционирования системы защиты

После запуска системы проходит этап инициализации и проверки целостности. Следующим шагом происходит переход в режим захвата и обработки сетевого трафика. Сетевые сенсоры получают данные о сессии и поступающих сетевых пакетах. На основе полученных данных производится выделение информативных признаков, которые передаются на вход модулю управления компонентами. Полученные признаки передаются на обработку модулям анализа, а также, может произвести передачу признаков модулю хранения для записи в базу данных.

Модули анализа обрабатывают поступившие данные, и в случае выявления факта атаки производят оповещение модуля реакции, посредством модуля управления. Передаётся сообщение о факте атаки, её типе, а так же дополнительная информация, характерная для данного типа атаки. В соответствии с ранее определённым списком модуль реакции осуществляет ответные действия. При этом производится передача информации модулю хранения информации о типе атаки и применённой реакции. Также модуль контроля компонент производит оповещение всех экземпляров системы защиты.

3. Основные результаты и выводы

Созданный на данном этапе исследований экспериментальный образец позволил частично решить поставленные задачи. Разработанный механизм взаимодействия на основе модуля управления компонентами позволяет обеспечивать защиту облака, построенного на основе кластерного вычислителя. Система предоставляет единый графический интерфейс администратора, позволяющий централизованно контролировать все компоненты системы защиты. Механизм обнаружения сетевых атак при помощи комитета классификаторов характеризуется высокими показателями распознавания по сравнению с простыми нейросетевыми классификаторами, рассматриваемыми в более ранних исследованиях (табл. 1). Используемые классификаторы предоставляют возможность дообучения, что позволяет корректировать работу классификатора в случае появления новых типов атак и существенно сокращает время подготовки классификатора.

Экспериментальные исследования с использованием лабораторного стенда показали высокий уровень обнаружения атак на основе предложенных классификаторов (свыше 99% по полноте и точности обнаружения факта атаки распознающим автоматом, близкое к нулю число ошибок первого и второго рода у комитета классификаторов) (табл. 2).

Тестирование оценивалось по двум параметрам:

1) точность для класса A — делим число правильных определений записей класса A на число фактов указания на класс A;
2) полнота — делим число правильных определений записей класса A на реальное число записей класса A.

Таблица 1

Распознавание тестовой и обучающей выборок (определение нормы и классов атак)

Тип классификатора	**Распознавание, прец./сек.**
Автоматный классификатор	700000
Сеть прямого распространения, Л-М	53000
Полиномиальное расстояние Евклида–Махаланобиса	192000
Метод опорных векторов	59000

Таблица 2

Распознавание тестовой и обучающей выборок (определение нормы и классов атак)

Тестируемая система	**Результаты распознавания атак, точность, %/полнота, %**				
	«Норма»	«DoS»	«Nmap»	«HTTP-exploit»	«SNScan»
ЭО ИПС	100/100	100/100	100/99,8	100/100	100/99,4
IDS Snort	100/99,77	100/0,002	100/0,075	100/5	100/85,93
IDS Bro	100/99,97	100/40,68	100/0,006	94,91/100	100/0,05

Система защиты за счёт использования нейросетевого классификатора обеспечивает распознавание модифицированных сетевых атак и обнаружение закономерностей и аномалий в потоках данных.

В качестве дальнейшего развития работы предполагается разработка интеллектуального механизма управления распределённой структурой системы защиты, а также исследования в области автоматического переобучения комитета в ходе работы.

Литература

1. *Кондратьев А. А., Тищенко И. П., Фраленко В. П.* Разработка распределенной системы защиты облачных вычислений // Программные системы: теория и приложения: электрон. научн. журн. — 2011. — № 4. — С. 61–70. — URL: `http://psta.psiras.ru/read/psta2011_4_61-70.pdf`. [Kondratyev A. A., Tishchenko I. P., Fralenko V. P. Development of a Distributed System Security for Cloud Computing // Software Systems: Theory and Applications. Electronic scientific journal. — 2011. — No 4. — Pp. 61–70. — (in russian).]
2. Разработка нейросетевого модуля мониторинга аномальной сетевой активности / А. А. Талалаев, И. П. Тищенко, В. П. Фраленко, В. М. Хачумов // Нейрокомпьютеры: разработка и применение. — 2011. — № 7. — С. 32–38. [The Development of the Neural Network Module for Monitoring of Abnormal Network Activity / A. A. Talalaev, I. P. Tishchenko, V. P. Fralenko, V. M. Khachumov // Neurocomputers: Development and Application. — 2011. — No 7. — Pp. 32–38. — (in russian).]
3. *Фраленко В. П.* Обнаружение сетевых атак с помощью генетически создаваемых конечных автоматов // Вестник РУДН. Серия «Математика. Информатика. Физика». — 2012. — № 4. — С. 96–102. [Fralenko V. P. Intrusion Detection using Genetically Generated Finite Automata // Bulletin of Peoples' Friendship

University of Russia. Series “Mathematics. Information Sciences. Physics”. — 2012. — No 4. — Pp. 96–102. — (in russian).]

4. Нейросетевые алгоритмы сжатия и восстановления потоков данных, а также фильтрации сетевых атак комитетом классификаторов / В. П. Фраленко, В. М. Хачумов, Ю. М. Урличич и др. // Ракетно-космическое приборостроение и информационные технологии. 2011. Труды IV Всероссийской научно-технической конференции «Актуальные проблемы ракетно-космического приборостроения и информационных технологий» (15–17 июня 2011 г.). — М.: Радиотехника, 2012. — С. 197–205. [Neural Network Algorithms for Data Streams Compression and Decompression as well as Network Attacks Filtering by Classifiers Committee / V. P. Fralenko, V. M. Khachumov, Yu. M. Urlichich et al. // Rocket and space instrumentation and information technology, 2011. Proceedings of the IV All-Russian Scientific Conference “Actual problems of rocket-space instrumentation and information technology” (15–17 june 2011 year). — Moscow: Radiotekhnika, 2012. — Pp. 197–205. — (in russian).]
5. *Емельянова Ю. Г., П. Ф. В.* Анализ проблем и перспективы создания интеллектуальной системы обнаружения и предотвращения сетевых атак на облачные вычисления // Программные системы: теория и приложения: электрон. научн. журн. — 2011. — № 4. — С. 17–31. — URL: `http://psta.psiras.ru/read/psta2011_4_17-31.pdf`. [Emelyanova J. G., Fralenko V. P. Problems and Prospects Analysis for Cloud Computing Network Attacks Detection and Prevention Intelligent System Creation // Software Systems: Theory and Applications. Electronic scientific journal. — 2011. — No 4. — Pp. 17–31. — (in russian).]
6. *Емельянова Ю. Г., Мбайкоджи Э., Соченков И. В.* Современный уровень и тенденции развития средств обеспечения сетевой безопасности систем облачных вычислений // Вестник Российского университета дружбы народов. Серия «Математика. Информатика. Физика». — 2012. — № 2. — С. 116–126. [Emelyanova J. G., Mbaykodzhi E., Sochencov I. V. The Modern Level and Development Trends of Network Security for Cloud Computing System // Bulletin of Peoples’ Friendship University of Russia. Series “Mathematics. Information Sciences. Physics”. — 2012. — No 2. — Pp. 116–126. — (in russian).]

UDC 004.492.3

Distributed System for Detection and Prevention Network Attacks to Cloud Computing

A. A. Kondratyev

Organization of Russian Academy of Sciences Program System Institute of RAS
4a, Peter I str., Veskovo, Pereslavl-Zalessky, Yaroslavl Region, 152021

The paper describes the problem of detection of intrusion for distributed systems and cloud computing. The goal is to detect and prevent both classical and distributed network attacks such as Denial of Service (DoS, DDoS). The paper identifies a number of problems of various popular cloud computing systems that represent a danger not only obtaining access to the user data, but also it could compromise the integrity and efficiency of the computer system. As solution it is proposed to develop a system for detecting and preventing network attacks. The system consists of several modules designed to perform different functions: detection and prevention of attacks, interaction of system modules, data management and storage. Recognition algorithms are based on the methods of artificial intelligence and the theory of probability. The new solution uses some intellectual methods to recognize attacks as counter to signature-based approach. This paper describes the architecture and functioning of the proposed solutions. It presents the advantages and disadvantages of the described approach. The solution results were presented in the conclusion of the paper. Solution was tested with different types of network attacks on a specially prepared experimental stand.

Key words and phrases: cloud, security, network attack, artificial intelligence.