Федеральное государственное учреждение
«Федеральный исследовательский центр
«Информатика и управление»
Российской академии наук»
(ФИЦ ИУ РАН)

Россия, 119333, г. Москва, ул. Вавилова, д .44, корп.
Тел. 8 (499) 135-62-60, факс 8 (495) 930-45-05
E-mail: ipiran@ipiran.ru http://www.ipiran.ru

От _____ № _____

На № _____

УТВЕРЖДАЮ
Директор ФИЦ ИУ РАН

И.А. Соколов
мая 2017 г.

ведущей организации на диссертацию Маркина Юрия Витальевича «Методы и средства углубленного анализа сетевого трафика», представленную на соискание ученой степени кандидата технических наук по специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей»

В современных информационных системах большинство анализаторов трафика предназначено для решения какой-то одной практической задачи. При этом в каждом инструменте анализа используются собственные разборщики протоколов для выделения полей в заголовках пакетов. Добавление поддержки нового протокола во все инструменты предполагает создание разборщика для каждого из них, что, в свою очередь, требует времени и ресурсов, а также повышает риск появления эксплуатируемой ошибки в коде, обрабатывающем трафик. Для компрометации всей подсети в таком случае будет достаточно успешной атаки на одну из используемых программных систем анализа трафика. В задачах современной практики целесообразно организовать разбор пакетов в рамках одной системы, предоставляющей доступ к результатам разбора в качестве услуги для других систем, решающих прикладные задачи, такие как, управление трафиком и обеспечение информационной безопасности. В связи с этим, тема диссертации Ю.В.

Маркина, посвященной разработке архитектуры подобной программной системы, является ***актуальной***.

В диссертации рассматривается подход к построению системы анализа сетевого трафика, состоящий в том, что разбор трафика осуществляется одним инструментом, а результаты разбора предоставляются всем остальным анализаторам в качестве входных данных. К основным результатам диссертации можно отнести следующие:

1) разработана модель представления разобранных сетевых пакетов, в которой введены сущности и методы для обработки ситуаций потери/переупорядочивания пакетов, описания зашифрованных данных, разбора заголовков нефиксированного стека протоколов;

2) разработан алгоритм восстановления потоков, устойчивый к переупорядочиванию пакетов;

3) разработана архитектура системы углубленного анализа сетевого трафика, позволяющая создавать и отлаживать разборщики заголовков протоколов на предварительно сохраненном трафике и впоследствии использовать их при анализе трафика на потоке;

4) разработаны и реализованы программные инструменты для проведения углубленного анализа сетевого трафика в online и offline режимах.

В главе два в рамках разработанной автором модели представления данных сетевое взаимодействие двух приложений впервые интерпретируется посредством набора логических соединений, по одному на каждом уровне стека протоколов. В анализируемом трафике, как правило, представлено множество различных взаимодействий, организованных посредством разных протоколов, поэтому при проведении разбора пакеты необходимо группировать по протоколам и логическим соединениям. Для этого в модели введены

понятия контекста и ключевой группы. С помощью контекстов реализуется разделение пакетов по протоколам. Далее, уже в рамках зафиксированного протокола на основе служебных полей заголовка этого протокола формируются ключи, отличающие пакеты разных логических соединений друг от друга. Важно, что в разработанной модели ключ формируется из служебных полей пакетов того протокола, к которому относится логическое соединение, тогда как в рассмотренных автором анализаторах трафика для группировки пакетов применяются служебные поля из заголовков нижележащих протоколов, что делает разборщики заголовков зависимыми между собой. В рамках логических соединений для проведения дальнейшего разбора пакетов может потребоваться объединить данные, передаваемые в разных пакетах. В частности, необходимость в объединении возникает при разборе TCP-пакетов, поскольку границы сообщений, передаваемых поверх TCP, в общем случае не совпадают с границами TCP-пакетов, а перед тем, как разбирать сообщения, нужно представить их в том виде, в котором они были отправлены. Объединение данных из разных пакетов в разработанной модели описывается с помощью потоков. В разработанном алгоритме восстановления потоков для каждого пакета, пришедшего в неправильном порядке, создается отдельный поток. По мере получения «запоздавших» пакетов осуществляется склейка частично-восстановленных потоков, что в итоге приводит к формированию лишь одного потока, восстановленного полностью. В широко распространенном анализаторе Wireshark применяется другая стратегия обработки переупорядоченных TCP-пакетов, состоящая из 1) определения размера извлекаемого пакета и 2) объединения полезной нагрузки пакетов, формирующих пакет вышележащего протокола. Если размер извлекаемого пакета вышележащего протокола неизвестен на момент обработки TCP-пакета, пришедшего в неправильном порядке,

полезная нагрузка этого TCP-пакета будет обрабатываться независимо от данных, извлеченных из других TCP-пакетов, что является ошибкой. Такая ситуация может возникнуть при порционной передаче полезной нагрузки HTTP-пакета, когда размер пакета определяется последней порцией данных. В настоящее время переупорядочивание пакетов при передаче происходит регулярно, и при этом принимающая сторона извлекает данные вышележащих пакетов корректно без каких-либо дополнительных запросов о повторной передаче.

В главах три и четыре автором впервые установлено, что контекст служит хранилищем для логических соединений, относящихся к заданному протоколу. В свою очередь в рамках одного логического соединения могут восстанавливаться потоки данных и создаваться контексты, соответствующие протоколам следующего уровня стека. В результате восстановления логических соединений для всех взаимодействий в сетевом трафике будет построено дерево контекстов, ключевых групп и восстановленных потоков с вершинами трех типов: вершина-контекст задает протокол, вершина-группа характеризуется ключом и описывает логическое соединение в рамках протокола контекста-родителя, вершина-поток описывает восстановленный поток данных протокола вышележащего уровня. Посредством дерева осуществляется доступ к пакетам любого логического соединения – для этого необходимо задать путь от корня к вершине, описывающей нужное логическое соединение. Кроме того, следует подчеркнуть, что разработанная модель представления данных реализована в виде ядра системы, которое компилируется в динамическую библиотеку. На базе API ядра построена работа модулей разбора и распознавания данных. Ядро также предоставляет доступ к результатам проводимого разбора, благодаря чему оно может использоваться при построении инструментов анализа, требующего разбирать сетевой трафик.

Рассмотренные автором анализаторы трафика не предполагают такой функциональности.

В главах четыре и пять в работе на базе ядра разработано два инструмента для проведения анализа в online и offline режимах. Инструмент online-анализа предназначен для извлечения данных из передаваемого трафика. Предполагается, что он будет использоваться сторонними системами, в которых необходимо проводить разбор пакетов. В отличие от online-анализатора, обладающего минимальным графическим интерфейсом, offline-анализатор предоставляет пользователю графические компоненты, с помощью которых можно проследить за тем, из каких пакетов были восстановлены потоки данных, какие пакеты пришли в неправильном порядке, по каким протоколам осуществлялись взаимодействия, какие ошибки в процессе разбора произошли. Основное предназначение инструмента – разработка и отладка разборщиков. На вход offline-анализатор получает файл с предварительно сохраненным сетевым трафиком. Оба инструмента используют один и тот же комплект исходных кодов разборщиков – это достигается благодаря различной для двух режимов анализа реализации в ядре функций API, применяемого для разбора. Таким образом, разрабатываемый посредством offline-анализатора разборщик будет использован при проведении анализа в online-режиме. Кроме того, для увеличения числа поддерживаемых протоколов был разработан инструмент портирования разборщиков из анализатора Wireshark. Исходный код разборщика Wireshark транслируется в абстрактное синтаксическое дерево, над которым проводятся преобразования, направленные на замену функций используемого API, после чего генерируется код разборщика для разработанной системы. В диссертации приведены примеры практического применения разработанных инструментов. Offline-анализатор был применен к

набору файлов с трафиком, содержащим переупорядоченные TCP-пакеты. В результате все TCP-потоки были корректно восстановлены, и из них были извлечены данные вышележащих протоколов, в частности, html-страницы. Для сравнения, Wireshark не смог корректно восстановить TCP-потоки при анализе этих файлов с трафиком. Показана возможность работы с зашифрованным трафиком. Описана последовательность действий для разработки разборщика закрытого протокола, используемого ботнетом.

Представленная диссертация позволяет сделать вывод, что автором проделан большой объем новых исследований в области анализа сетевого трафика и связанных с ним практических задач. Предложенная автором модель представления данных учитывает существующие особенности передачи сетевого трафика, а также позволяет единообразно выделять и описывать множества пакетов для последующего восстановления потоков. Разработанный автором новый алгоритм восстановления потоков данных обладает устойчивостью к потере отдельных пакетов и их переупорядочиванию, что увеличивает глубину проводимого разбора. Разработанные инструменты offline и online анализа могут применяться как для решения практических задач, так и в области научных исследований. Процесс разработки и последующей отладки разборщиков в offline-режиме приводит к повышению качества разбора, проводимого инструментом online-анализа, уменьшая количество неточностей при выделении полей и последующем объединении данных в потоки. Вынос функциональности разбора сетевого трафика в отдельный инструмент, предоставляющий доступ к пакетам всех логических соединений и восстановленным потокам данных, обеспечивает возможность построения на базе такого инструмента анализаторов, решающих различные прикладные задачи.

Достоверность и обоснованность научных результатов обусловлена, тем, что автором применен формализованный подход, предполагающий использование математических объектов для представления сетевых данных. Математическую основу исследования составляют теория множеств, теория графов, теория алгоритмов, математическая логика, теория формальных языков и теория автоматов.

Работа и автореферат тщательно оформлены. Автореферат хорошо отражает содержание диссертации. В целом работа завершена как в методическом, так и в прикладном плане.

Результаты диссертации достаточно опубликованы: семь публикаций и 3 доклада на конференциях. Работа поддержана грантом РФФИ.

По тексту диссертации могут быть сделаны следующие замечания:

1. в описании online-анализатора не сказано о скорости проведения разбора сетевых пакетов;

2. применение разработанного инструмента, автоматизирующего портирование разборщиков системы Wireshark, показано всего на одном примере, не приводится оценка полученных результатов.

Перечисленные недостатки не влияют на положительную оценку диссертации и не ставят под сомнение полученные в ней результаты, которые представляют несомненный интерес, как в научном, так и в практическом плане.

Практические результаты диссертанта внедрены в НИОКР ФИЦ ИУ РАН для систем высокой доступности.

В качестве рекомендации можно предложить диссертанту опубликовать цикл научно-технических статей по применению результатов автора в задачах проектирования систем высокой

доступности (журнал «Системы высокой доступности», издательства «Радиотехника»).

Разработанный диссертантом инструмент для проведения анализа в on-line режиме рассчитан на использование в сторонних информационных системах. Инструмент для проведения отложенного анализа используется при решении задач, связанных с обратной инженерной или отладкой сетевых протоколов, в области научных исследований и учебных курсах ВМК МГУ и ФУМП МФТИ.

Диссертационная работа соответствует паспорту специальности 05.13.11 «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей» по областям исследований:

– модели, методы и алгоритмы проектирования и анализа программ и программных систем, их эквивалентных преобразований, верификации и тестирования;

– модели, методы, алгоритмы, языки и программные инструменты для организации взаимодействия программ и программных систем.

Таким образом, диссертация Маркина Юрия Витальевича является научно-квалифицированной работой, в которой содержится разработка методического и инструментального программного обеспечения углубленного анализа сетевого графика в помехоустойчивых перспективных информационных системах высокой доступности, позволяющих автоматизировать расширение их функциональности, что соответствует требованиям п. 7 «Положения о порядке присуждения ученых степеней», утвержденного Постановлением Правительства РФ от 30.01.2002 г. №74 (с изменениями, внесенными Постановлением Правительства РФ от 20.06.2011 г. №475, предъявляемым к диссертациям на соискание ученой степени кандидата наук по специальности 05.13.11 «Математическое и программное обеспечение

вычислительных машин, комплексов и компьютерных сетей». Ее автор, Маркин Юрий Витальевич заслуживает присуждения ему ученой искомой ученой степени кандидата технических наук.

Отзыв обсужден и утвержден на заседании секции Ученого совета ИПИ РАН Федерального исследовательского центра «Информатика и управление» Российской академии наук, протокол № 3 от 24 апреля 2017 г.

Ученый секретарь ФИЦ ИУ РАН,
доктор технических наук

В.Н. Захаров
__ 2017 г.

Заведующий отделом ФИЦ ИУ РАН,
член-корреспондент РАН,
доктор физ.-мат. наук, профессор

Серябряков
__ 2017 г.

Главный научный сотрудник ФИЦ ИУ РАН
Заслуженный деятель науки РФ,
доктор технических наук, профессор

И.Н. Синицин
____ 2017 г.