ОТЗЫВ

официального оппонента на диссертационную работу
Антоненко Виталия Александровича
«РАЗРАБОТКА И ИССЛЕДОВАНИЕ МОДЕЛИ
ФУНКЦИОНИРОВАНИЯ ГЛОБАЛЬНОЙ СЕТИ ДЛЯ АНАЛИЗА
ДИНАМИКИ РАСПРОСТРАНЕНИЯ ВРЕДОНОСНОГО
ПРОГРАММНОГО ОБЕСПЕЧЕНИЯ»,
представленную на соискание ученой степени кандидата физико-математических наук
по специальности 05.13.11 — математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей

Современные компьютерные сети очень масштабны содержат большое число узлов. Для обозначения таких сетей, с количеством узлов порядка сотен тысяч, используется термин «глобальная компьютерная сеть» (ГКС). Анализ потоков данных в таких сетях – весьма сложная и ресурсоёмкая задача. Тем не менее, во многих случаях требуется проведение экспериментов для оценки поведения сети в различных ситуациях, среди которых можно выделить:

- распространения вирусного ПО, для анализа динамики распространения;
- доставка контента от сервера к получателю, для прогнозирвания задержки;
- применение новых протоколов маршрутизации для оценки их скорости сходимости.

Учитывая размеры и сложность ГКС эксперименты без использования моделирования значительно затруднены и крайне дороги. Задача моделирования осложняется тем, что большинство существующих подходов либо немасштабируются на размеры ГКС, либо недостаточно точны для решения указанных задач.

Диссертационная работа Антоненко В.А., посвящёна **актуальной задаче** разработки и реализации системы моделирования информационных процессов, протекающих в ГКС.

Во **введении** обоснована научная новизна и актуальность работы, сформулирована её цель, перечислены задачи, решение которых необходимо для её достяжения и основные результаты. В общем виде сформулирована проблематика проводимого исследования и представлен краткий обзор предметной области.

Диссертация состоит из введения, четырех глав и заключения. Объем диссертации составляет 108 страниц, имеются 5 таблиц и 41 рисунок. Список литературы насчитывает 115 источников.

Первая глава содержит подробный обзор предметной области, а также методов и средств моделирования компьютерной сети, анализируются преимущества и недостатки различных подходов и инструментов. Данная глава состоит из двух основных частей. ***В первой части главы*** вводятся основные определения в области структуры и процессов в ГКС, а также формулируются требования, предъявляемые к моделированию функционирования ГКС, такие как: масштабируемость, ресурсоёмкость и точность. ***Вторая часть*** посвящена обзору состояния исследований в области математического аппарата, используемого при моделировании ГКС, а также средств реализации модели сети. Рассматриваются подходы к моделированию на основе теории вероятностей, сетей Петри, теории графов и автоматов. Среди средств реализации модели выделяются специализированные языки (SIMULA, GPSS) и системы иммитационного моделирования (OpenVSwitch, NS-3, OPNET). Отдельно рассматривается подход на основе high-fidelity (Hi-Fi) системы Mininet, как наиболее перспективный с точки зрения маштабирования и точности моделирования. По результатам обзора, автором делается вывод о невозможности выбора единственного математического аппарата и необходимости гибридного подхода. В качестве средства реализации модели предлагается разработка собственной системы имитационного

моделирования (СИМ) на основе Hi-Fi систем.

Во **второй главе** описывается математический аппарат, лежащий в основе разрабатываемой автором СИМ. Вначале описывается структура ГКС в виде графа, вершинами которого являются домены и полюса, а рёбрами – каналы. Также формализуется состояние домена, включающее число активных хостов, степень их защищённости от вирусного ПО, долю заражённых хостов и ресурсы домена. Далее описывается эпидемиологическая модель заражения отдельного домена, как функция доли заражённых хостов зависящая от времени. Сетевая активность домена описывается с использованием двух математических моделей – системы массового обслуживания и теории автоматов. Сетевые потоки через домен рассматривается как очереди, а ресурсы домена – как способность их обрабатывать. При этом интенсивность потоков существенно зависит от стратегии обработки. Автор выделяет пять основных стратегий и показывает как стратегия зависит от состояния сети и полноты информации о ней, а также даёт рекомендации по применению различных стратегий при проведении экспериментов для достижения более точных результатов. Сама обработка потока в домене описывается в терминах конечного автомата. При этом для того, чтобы автомат различал потоки разных типов (например потоки вирусного ПО или потоки потребляемые данным доменом), вводится система тегов, подмножества которых и составляют алфавит автомата. В конце главы содержится доказательство корректности построенной модели ГКС, то есть разрешимости, в рамках этой модели, задач оценки числа заражённых хостов и доли трафика вирусного ПО. Также рассматривается служба управления времени и алгоритм синхронизации часов, позволяющие распределять домены по разным узлам кластера с сохранением порядка пакетов в каналах.

Третья глава посвящена описанию разработанной автором системы иммитационного моделирования на основе Hi-Fi подхода системы Mininet. Вначале формулируются требования к реализуемой СИМ и перечисляются представления элементов ГКС, такие как каналы, хосты, сетевые устройства в реализуемой системе. Затем рассматривается расширяемая кластерная архитектура, лежащая в основе системы и функции отдельных узлов, каждый из которых является виртуальной или физической машиной с установленной ОС Linux. Важным преимуществом реализованной системы является возможность задания фоновой сетевой активности, помимо основной, исследуемой в конкретном эксперименте. Это достигается засчёт запуска на различных узлах сетевых приложений из реализованной библиотеки приложений, при этом для каждого приложения может быть определён профиль его сетевой активности. В конце главы описывается система легковесной виртуализации ОС Linux – LXC, которая позволяет эмулировать масштабные сети (десятки тысяч) на одном физическом сервере. Рассматриваются различные методы управления временем, такие как дискретный, дискретно-событийный и непрерывный, среди которых для реализации выбирается дискретный метод, как наиболее подходящий к условиям функционирования СИМ. Отдельный раздел посвящён описанию графического интерфейса системы, преимуществом которого является, с одной стороны возможность эффективного описания и настройки ГКС для проведения экспериментов, а с другой – большое количество способов и средств визуализации результатов экспериментов, выдаваемых в том числе и в реальном времени. Описание системы заканчивается выводом о соответствии реализованной системы заявленным требованиям.

В **четвёртой главе** описываются набор экспериментальных исследований проведённых с целью проверки свойств и функционала реализованной СИМ. Структурно глава состоит из двух частей, в первой из которых проводятся эксперименты с отслеживанием динамики распространения вирусного ПО в сетях различного масштаба, а во второй исследуются особенности синхронизации времени между несколькими узлами кластера, на котором работает система. ***Первая часть*** начинается с описания механизма распространения вирусного ПО на примере сетевых червей. Далее описывается архитектура подсистемы, отвечающей за моделирование процесса распространения вирусного ПО. Первый эксперимент состоит в моделировании эпидемии сетевого червя CodeRedv2 на сети из нескольких сот тысяч хостов, которые группируются в 8 доменов. Моделирование

происходит на уровне сети доменов. Результаты сравниваются с реальными данными по количеству заражённых компьютеров в течение времени эпидемии. Второй эксперимент проводится на сети из 100000 отдельных хостов на примере эпидемии сетевого червя Sasser с учётом детального алгоритма его распространения. По результатам двух экспериментов Антоненко В.А. делается вывод о возможности применения системы для сетей масштаба ГКС. Во ***второй части*** главы для исследования синхронизации времени проводится эксперимент, в котором система запускается на кластере из двух узлов, на первом из которых запускается сервер DHCP и исследуется процесс получения новых IP-адресов хостами, которые обрабатываются на разных узлах. По результатам эксперимента делается вывод о корректной работе системы управления времени при правильной её настройке. Платой за это является примерно двукратное замедление модельного времени по сравнению с реальным.

В **заключении** приводятся результаты работы, выносимые на защиту, и формулируются направления для дальнейших работ по данной тематике.

Разработанная математическая модель и подход к иммитационному моделированию на основе техник легковесной виртуализации являются новыми научными результатами. Практическим результатом работы является реализованная распределённая система иммитационного моделирования, позволяющая строить эффективно масштабируемые сети и с высокой точностью воспроизводить процессы обработки и передачи сетевого трафика.

Модельные испытания показали, что реализованная система позволяет изучать динамику распространения вирусного ПО, а также успешно решает задачу синхронизации времени при маштабировании на несколько физических вычислителей.

Полученные автором результаты прошли апробацию на международных конференциях и семинарах.

В работе следует отметить следующие **недостатки**:

1. При описании состояния домена как пятёрки параметров (стр. 44), последние два параметра Res и FPS не рассмотрены. Первый из них рассмотрен в следующем разделе, а описание второго далее по тексту не встречается.
2. При описании модели заражения домена (стр. 45) недостаточно обосновано применение эпидемической модели – отсутствует описание ограничений на возможность её применения и её влияния на точность получаемых результатов.
3. В ходе описания экспериментов (стр. 90), в тексте фигурируют снимки экрана, подписи которых не соответствуют правилам нумерации и оформления, а сами рисунки не учтены при подсчётах – диссертация содержит 41 рисунок, а не 36, как указано во введении.

Отмеченные недостатки не влияют на общую положительную оценку работы.

Диссертационная работа В.А. Антоненко является законченным научным исследованием, выполненным автором самостоятельно. **Достоверность** исследования подтверждается проведенными модельными экспериментами. Результаты работы докладывались на конференциях и научных семинарах. Автореферат диссертации полно и правильно отражает ее содержание.

Диссертационная работа удовлетворяет всем требованиям, предъявляемым ВАК к кандидатским диссертациям, а ее автор Антоненко Виталий Александрович заслуживает присуждения ученой степени кандидата физико-математических наук по специальности 05.13.11 - «Математическое и программное обеспечение вычислительных машин, комплексов и компьютерных сетей».

Младший научный сотрудник
отдела компиляторных технологий
института системного программирования РАН,
к.ф.-м.н. А.И. Гетьман

Подпись А.И. Гетьмана удостоверяю
Учёный секретарь ИСП РАН, д.ф-м.н. А.И. Аветисян

04.09.2014 г.