На правах рукописи

Антоненко Виталий Александрович

Разработка и исследование модели функционирования глобальной сети для анализа динамики распространения вредоносного программного обеспечения

Специальность 05.13.11 — «Математическое обеспечение вычислительных машин, комплексов и компьютерных сетей.»

Автореферат

диссертации на соискание учёной степени

кандидата физико-математических наук

Москва — 2014

Работа выполнена в Московском государственном университете имени М.В. Ломоносова на факультете вычислительной математики и кибернетики.

Научный руководитель:	член-корреспондент РАН, профессор Смелянский Руслан Леонидович
Официальные оппоненты:	доктор физико-математических наук, заведующий кафедрой теоретической информатики Ярославского государственного университета им. П.Г. Демидова, профессор Соколов Валерий Анатольевич
	кандидат физико-математических наук , младший научный сотрудник Института системного программирования Российской академии наук Гетьман Александр Игоревич
Ведущая организация:	Федеральное государственное бюджетное учреждение науки Институт прикладной математики им. М.В. Келдыша Российской академии наук.

Защита состоится 19 сентября 2014 г. в 11^{00} часов на заседании диссертационного совета Д. 501.001.44 при Московском государственном университете имени М.В. Ломоносова по адресу: 119991, ГСП-1, Москва, Ленинские горы, МГУ, 2-й учебный корпус, факультет ВМК, аудитория 685.

С диссертацией можно ознакомиться в Фундаментальной библиотеке МГУ.

С текстом автореферата и диссертации можно ознакомиться на официальном сайте факультета ВМК МГУ http://cs.msu.ru/в разделе «Диссертации».

Автореферат разослан: _____.

Ученый секретарь
диссертационного совета
к. т. н., ведущий научный сотрудник Костенко В.А.

Общая характеристика работы

Актуальность темы. Под термином Глобальная Компьютерная Сеть (ГКС) будем понимать компьютерную сеть, состоящую не менее чем из 10^5 узлов. Узел сети характеризуется определенным набором параметров. Значения параметров узла сети определяют его состояние; совокупное состояние всех узлов определяет состояние сети. Узлы соединяются каналами и тем самым образуют структуру сети, называемую топологией.

В современных ГКС актуально уметь оперативно прогнозировать динамику изменения состояния сети, например:

- прогнозировать динамику распространения вредоносного программного обеспечения (ВПО) и оценивать наносимый ущерб ГКС. Ущербом будем считать изменение параметров качества сервиса (задержек и процента потерь легитимного трафика);
- прогнозировать и оценивать задержку при доставке контента от сервера хранения до получателя;
- прогнозировать скорость сходимости протоколов маршрутизации и оценивать накладные расходы (например, количество служебного трафика в ходе функционирования исследуемого протокола).

Исходя из большой размерности и сложности структуры ГКС, эксперименты без использования моделирования затруднены по финансовым причинам, а также из-за невозможности физического воссоздания сети столь большого размера. Имитационная модель ГКС — это с одной стороны комбинация математической модели и ее реализация на ВС, с другой — результат компромисса между:

- уровнем детальности описания;
- сложностью описания функционирования;
- точностью предсказания поведения;
- сложностью идентификации и калибровки построенной модели.

Необходимая детализация имитационной модели ГКС зависит от целей моделирования и определяется исследователем при подготовке эксперимента моделирования. Подробность и точность имитационной модели зависит от выбора уровня абстракции объекта моделирования, а также от выбора математического аппарата, в терминах которого строится модель. В настоящее

время исследователи для моделирования ГКС чаще всего используют вероятностный математический аппарат.

Использование вероятностного математического аппарата предполагает статистическое усреднение многих параметров функционирования ГКС, либо их представление в форме функции распределения соответствующего вида. Статистическое усреднение используется для:

- упрощения модели;
- сокращения размерности модели;
- уменьшения вычислительной сложности реализации модели.

Целью данной работы является разработка и реализация системы моделирования процесса функционирования ГКС с возможностью анализа динамики распространения ВПО.

Для достижения поставленной цели необходимо было решить следующие задачи:

1. составить обзор математических методов описания/построения моделей функционирования сети и средств реализации моделей сети. Оценить их с точки зрения следующих критериев:
   (a) точность моделирования функционирования сети (точность);
   (b) требовательность к вычислительным ресурсам для вычисления результатов моделирования (ресурсоемкость);
   (c) зависимость количества узлов в моделируемой сети от количества вычислительных ресурсов, используемых в процессе моделирования (масштабируемость);
2. построить формальную модель ГКС, которая позволяет моделировать функционирование ГКС с возможностью анализа динамики распространения ВПО;
3. исследовать применимость эпидемических моделей, известных из медицины и биологии, для описания процесса распространения ВПО;
4. разработать систему имитационного моделирования ГКС с возможностью точного моделирования процесса распространения ВПО, то есть моделирования всех стадий (выбор жертвы, сканирование, заражение и т.д.) жизненного цикла ВПО;
5. исследовать динамику распространения тестового набора ВПО в разработанной системе имитационного моделирования ГКС.

Основные положения, выносимые на защиту:

1. Построена математическая модель, позволяющая моделировать функционирование ГКС, при этом результат моделирования близок к функционированию реальной сети. В терминах этой модели описана задача прогнозирования динамики распространения ВПО и показано, что она разрешима.

2. На основе техники легковесной виртуализации предложен новый подход к моделированию функционирования ГКС, позволяющий строить модели сетей необходимого размера. Отличительной чертой подхода является высокая точность моделирования процесса функционирования ГКС по сравнению с существующими подходами.

3. Разработана и реализована уникальная распределенная система имитационного моделирования (СИМ) сети с использованием «виртуальных контейнеров», названная Network Prototype Simulator (NPS). Результаты моделирования в СИМ NPS продемонстрированы применительно к задачи исследования динамики распространения ВПО.

Научная новизна заключается в разработке подхода к построению имитационных моделей на основе техники легковесной виртуализации, которая позволяет:

- масштабировать модель вплоть до размеров ГКС;
- сократить затраты на калибровку и идентификацию модели;
- избежать необходимости доказательства корректности конкретной модели сети, то есть доказательства факта воспроизведения процессов обработки и передачи сетевого трафика в заданной пользователем топологии сети.

Практическая значимость в создании системы имитационного моделирования на основе техники легковесной виртуализации заключается в упорядочении и упрощении процесса построения имитационной модели ГКС. Подобная система предназначена для:

- исследователей в области компьютерных сетей при анализе различных сетевых обменов и их влияния на различные сетевые характеристики;
- разработчиков сетевых приложений и протоколов для определения корректности работы приложения или, например, для исследования сходимости нового протокола маршрутизации;

- сетевых архитекторов на различных стадиях проектирования и реализации сети.

Апробация работы. Основные результаты работы докладывались на:

- 17h GENI ENGINEERING CONFERENCE (GEC17);
- SIGCOMM 2013;
- YET ANOTHER CONFERENCE 2013;
- SOFTWARE ENGINEERING CONFERENCE IN RUSSIA 2013.

Диссертационная работа была выполнена при поддержке грантов:

1. Российского фонда фундаментальных исследований 10-0144/01 от 25.03.2010;
2. Инновационного центра Сколково 79 от 02.07.2012.

Публикации. Основные результаты по теме диссертации изложены в печатных изданиях [1, 2], два из которых изданы в журналах, рекомендованных ВАК [1, 2].

Объем и структура работы. Диссертация состоит из введения, четырех глав и заключения. Полный объем диссертации составляет 108 страниц с 36 рисунками и 5 таблицами. Список литературы содержит 115 наименований.

Содержание работы

Во введении обосновывается актуальность исследований, проводимых в рамках данной диссертационной работы, приводится обзор научной литературы по изучаемой проблеме, формулируется цель, ставятся задачи работы, сформулированы научная новизна и практическая значимость представляемой работы.

Первая глава посвящена обзору математических методов описания, построения моделей функционирования сети и средств реализации моделей сети. Целью данного обзора является рассмотрение средств реализации и основных математических аппаратов, используемых для построения моделей сетей с точки зрения точности моделирования функционирования сети, требовательности к вычислительным ресурсам, зависимости количества узлов в моделируемой сети от количества вычислительных ресурсов, используемых в процессе моделирования.

В главе описана предметная область и сформулированы основные понятия. Так ГКС определена как сеть, состоящая не менее чем из 10^5 узлов.

Узлом называется точка в сети, имеющая уникальный сетевой адрес. Примером узла может быть либо хост, либо сетевое устройство, которое объединяет хосты между собой, например, коммутатор или машрутизатор. Хост (компьютер с сетевым интерфейсом) — это потребитель сетевого сервиса ГКС, на котором задан уникальный сетевой адрес. Предполагается, что ГКС обладает следующими свойствами:

- размер сети не менее 10^5 хостов. Хостом является только вершина степени один в графе сети;
- сеть разделена на домены. Домен — это множество связанных между собой хостов. Считается, что множества хостов различных доменов не пересекаются. Мобильные хосты не рассматриваются, то есть при переходе хоста из одного домена в другой первоначальные свойства домена не изменяются;
- различные домены связаны между собой каналами типа «точка-точка». Под каналом понимается пара сетевых интерфейсов, которая связывает хосты с сетевыми устройствами или сетевые устройства между собой.

Сетевое приложение определяется как распределенная система, разные части которой обмениваются между собой данными. Между сетевыми приложениями возникают потоки данных, совокупность которых образует трафик сети.

Сетевая активность ГКС определяется потоками трафика, передаваемыми между хостами. Эти потоки разделяются на два класса :

- легитимный трафик — трафик, генерируемый сетевыми приложениями и служебными сетевыми протоколами (ICMP, DNS, DHCP и др.);
- вредоносный трафик — трафик, генерируемый при передаче экземпляра вредоносного программного обеспечения (ВПО) и сетевой активности каждого экземпляра ВПО. Вредоносное ПО (ВПО) – любое сетевое приложение, целью которого является нанесение ущерба отдельному узлу или всей сети.

Процесс построения имитационной модели делится на два этапа:

1. выбор математического аппарата и построение математической модели сети;
2. решение математической модели через ее программную реализацию на вычислительной системе.

Точность и детальность моделирования зависят от выбора математического аппарата, на основе которого строится имитационная модель. Моделирование функционирования ГКС выдвигает особые требования, которые необходимо учитывать при выборе математического аппарата:

- Масштабируемость. ГКС по определению является сетью с большим числом узлов, модель должна позволять варьировать в широком диапазоне число узлов в сети без особых усилий для перестроения модели.
- Ресурсоемкость. Требовательность к вычислительным ресурсам, необходимым для моделирования сети, заданного пользователем размера.
- Точность. Обычно подходы, используемые для построения моделей ГКС, обладают низкой точностью. Это связанно с высокой ресурсоемкостью «точных» моделей.

Исходя из анализа открытых источников, были проанализированы существующие математические аппараты (теория вероятностей, сети Петри, теория графов, теория автоматов) по критериям: точность, ресурсоемкость и масштабируемость.

Сравнительный анализ математических аппаратов с целью исследования возможности их использования для построения моделей ГКС, представлен в таблице 1.

Таблица 1: Сравнительный анализ математических аппаратов.

Математический аппарат	Точность	Ресурсоемкость	Масштабируемость
ТВ	-	+	-
ТМО	-	-	+
СП	-	-/+	+/-
ТГ	+/-	-/+ (*)	+
ТА	+	-	-

(*) — зависит от необходимости моделирования сетевого трафика. В случае моделирования только структуры сети — это «чистый плюс».

В итоге сделан вывод о том, что ни один из представленных подходов к построению математической модели функционирования сети не удовлетворяет требованиям к моделированию ГКС, а так же о необходимости предложения гибридного варианта, который будет удовлетворять требованиям: точности, масштабируемости и ресурсоемкости. Данный подход был предложен, и на его основе построена формальная модель ГКС, описанная во второй главе.

Процесс реализации имитационной модели сети заключается в разработке программы, которая шаг за шагом воспроизводит события, происходящие в моделируемой сети. Обычно говорят о двух способах реализации имитационной модели:

1. использование специализированных языков моделирования;
2. использование специализированной системы имитационного моделирования. В данном случае разработчик строит модели, используя систему имитационного моделирования (СИМ), базируясь на математический аппарат, заложенный в основу конкретной СИМ.

В настоящее время к средствам реализации имитационной модели предъявляется существенный ряд требований, как то:

- описание исследуемой системы на различных уровнях детализации без ограничений на их количество;
- проектирование модели в графическом режиме на основе готовых блоков при минимальном объеме программирования;
- удобное и полное представление работы программного обеспечения сети (работы сетевых приложений);
- наглядное отображение процесса моделирования (использование методов анимации);
- снижение затрат времени моделирования на сбор статистики (накопление данных по требованию);
- распределенное моделирование сети на нескольких вычислителях;
- точное описание процесса обработки PDU;
- избегание процессов идентификации и калибровки модели сети.

В работе рассмотрены следующие средства реализации имитационной модели:

- специализированные языки моделирования: SIMULA и GPSS;
- системы имитационного моделирования: OPNET, OMNET++, AnyLogic, NS-2, NS-3.

Сравнительный анализ средств реализации имитационных моделей сетей с целью исследования возможности их использования для построения моделей ГКС представлен в таблице 2.

Таблица 2: Сравнительный анализ средств реализации имитационных моделей сетей.

Требование	SIMULA	GSPP	NS-2	NS-3	OPNET	OMnet++	Anylogic
различные уровни детализации	+	+	-	-	-	-	+
графический режим	-	-	+	+	+	+	+
представление работы ПО сети	-/+	-/+	+/-	+/-	-/+	-/+	-
отображение процесса моделирования	-	-	+	+	+	-	+
накопление данных по требованию	+	+	-	-	+	-	-
распределенное моделирование	-/+	-/+	-	-	-	-	-
описание процесса обработки PDU	+/-	+/-	+	+	+	+	-
избегание процессов идентификации и калибровки (*)	-	-	-/+	-/+	-/+	-/+	-/+
моделирование сетей размера ГКС	-/+	-/+	-	-	-	-	-

(*) — «-/+» означает отсутствие калибровки только на заранее подготовленных моделей сетевых приложений и протоколов.

Отдельно были выделены СИМ (названные Hi-Fi СИМ), моделирующие с высокой точностью процессы обработки и передачи сетевого трафика в заданной топологии сети.

В результате обзора математических аппаратов для моделировании ГКС было показано, что невозможно выбрать единственный математический аппарат для моделирования функционирования ГКС. Предлагалось использовать комбинацию математических аппаратов: теории графов, теории массового обслуживания, теории автоматов. Построенная модель ГКС описана во второй главе. Доказательство соответствия модели ГКС требованиям точности, ресурсоемкости и масштабируемости приводится в экспериментальном исследовании, представленном в четвертой главе.

В результате обзора существующих средств реализации имитационной модели обнаружилось, что ни одно не удовлетворяет описанным требованиям. Наиболее перспективными для реализации модели ГКС были выбраны Hi-Fi СИМ. Однако по причине отсутствия готовых СИМ, удовлетворяющих требованиям модели ГКС, необходима разработка и реализация собственной СИМ. Описание реализованной СИМ представлено в третьей главе.

Во второй главе приведено описание формальной модели ГКС при помощи математических аппаратов: теории графов, элементов теории массового обслуживания, а также теории автоматов. Основная цель данной формальной модели — спрогнозировать динамику распространения ВПО. Приведены основные понятия и термины. Доказано, что задача исследования динамики распространения ВПО в построенной формальной модели имеет решение.

Будем называть доменом — множество хостов и сетевых устройств. Хосты и сетевые устройства обмениваются между собой данными. Процесс обмена представляется в виде потока данных между доменами, который будем называть поток данных или поток. Каждый домен имеет определенный набор истоков и стоков:

- исток — это точка подключения канала к домену, через которую поток поступает в домен;
- сток — это точка подключения канала к домену, через которую поток уходит из домена.

Домены связаны между собой каналами связи или каналами. По каждому каналу поток может проходить только в одном направлении. Каждый канал имеет фиксированную пропускную способность. Канал может быть подключен только к одному стоку и одному истоку соответствующего домена. Другими словами, канал не может соединять один и тот же сток с

несколькими истоками как одного, так и разных доменов. Верно и обратное: несколько стоков не может быть соединено с одним и тем же истоком.

В модели каждый соединяющий домены канал однонаправлен и ориентирован от стока к истоку. Домены, соединенные каналом, будем называть смежными. Дуплексный канал в модели ГКС будет представлен двумя разнонаправленными каналами. В этих предположениях сеть доменов может быть описана в терминах конечно порожденной частичной алгебры.

Трафик в сети рассматривается в виде совокупности потоков, где каждый поток характеризует обмен данными между сетевыми приложениями. Сетевое приложение — это источник и/или потребитель сетевого трафика, ассоциированный с доменом. Совокупность потоков, проходящих через один и тот же канал, будем называть трафиком канала. Под интенсивностью потока будем понимать количество данных, прошедших по каналу в единицу времени.

Правомерность использования понятия потока основана на том, что в современных сетях преобладает использование виртуальных соединений между взаимодействующими сетевыми объектами. Говоря о потоке, мы фокусируем внимание на взаимодействии двух сетевых объектов, а не пытаемся охватить все взаимодействия, проходящие через определенный канал.

Рассматривая сеть на уровне доменов, мы пренебрегаем локальным трафиком внутри домена, который для оценки динамики распространения ВПО и влияния трафика ВПО на уровне потоков не важен. В рамках поставленной задачи необходимо оценить динамику заражения ВПО хостов внутри домена, что достаточно точно описывается эпидемическими моделями.

Для описания процесса заражения хостов внутри домена используется эпидемическая модель. Зная количество экземпляров ВПО, попавших в домен в заданный промежуток времени и защищенность домена от данного типа ВПО, возможно вычислить количество ВПО, прошедшего через защиту домена.

Стоит отметить, что использование эпидемических моделей не является обязательным, и выбор модели распространения ВПО зависит от известной информации о распространении конкретного экземпляра ВПО. Наиболее точным будет моделирование, которое основывается:

- на алгоритме распространения исследуемого ВПО;
- данных о его размерах;
- характеристиках сетевого приложения, при помощи которого осуществляется распространение.

Заметим, что домен в предлагаемом подходе играет двоякую роль. С одной стороны, в нем сконцентрированы ресурсы (хосты, сетевые приложения и т.п.), с другой — он выполняет функции коммутатора.

Предположим, что у нас есть сеть взаимосвязанных доменов, по которым проходят потоки различных типов: легитимные и вредоносные. Поток ориентирован от стока одного домена к истоку смежного домена. Внутри домена может быть порожден новый поток. Некоторые потоки, вошедшие в домен, могут быть «потреблены» доменом, который является конечной точкой маршрута, или же часть потока может быть потеряна по причине исчерпания ресурсов домена или перегрузкой канала. Динамику обработки потоков внутри домена будем описывать в виде автомата.

В рамках построенной модели рассматривается следующая задача: при заданных начальных условиях оценить динамику распространения ВПО среди хостов в ГКС и долю вредоносного потока в общем трафике ГКС.

При построении имитационных моделей критичным является вопрос о представлении модельного времени. Будем предполагать, что в сети один наблюдатель с едиными часами, значение которых изменяется дискретно. С целью упростить технику описания нашей модели будем считать, что время едино и изменяется дискретами. Предположение о едином наблюдателе для одного домена оправдано в случае использования распределенной системы вычислителей для реализации модели. В этом случае у каждого домена будут свои часы, свой наблюдатель. Случай моделирования с несколькими часами и описанием процесса их синхронизации рассмотрен в четвертой главе.

Корректность описанной формальной модели ГКС была сформулирована в виде теоремы.

Теорема 1. Пусть дана сеть с источниками и получателями потоков, тогда если в сети выполнены следующие условия:

- каждому потоку приписана функция начальной интенсивности;
- каждому каналу приписана пропускная способность;
- каждой вершине приписана величина, обозначающая ресурсы и конечный автомат для изменения потоков;

то существует алгоритм воспроизведения динамики заражения хостов и доли вредоносного трафика в ГКС, представляющее оценку числа зараженных хостов в каждой вершине сети и доли вредоносных потоков в общем трафике сети.

Доказательство теоремы приводится в тексте второй главы.

Третья глава описывает архитектуру системы имитационного моделирования (СИМ) Network Prototype Simulator (NPS). Целью разработки СИМ NPS является реализация формальной модели ГКС, описанной во второй главе.

СИМ NPS состоит из четырех основных частей:

- NPS главная управляющая консоль — это консоль управления, способная конфигурировать и выводить результаты выполнения команд с нескольких NPS узлов кластера, используя SSH соединение;
- NPS узел кластера — это виртуальная или физическая машина с установленной ОС Linux и с набором установленных пакетов: Python2.7, Pythonscapy, Mininet;
- NPS ПКС контролер используется для программирования сетевой активности внутри сегментов сетей в NPS узлах кластера и между данными сегментами;
- NPS библиотека сетевых приложений, которая содержит интерфейсы для запуска сетевых приложений на большом числе хостов. Архитектура библиотеки сетевых приложений дает возможность пользователю добавлять собственные приложения.

Под NPS кластером понимается несколько вычислительных машин, объединенных между собой в сеть. В этом случае не имеет значения использование виртуальных или физических вычислительных машин. Вычислительная машина, входящая в состав NPS кластера, является NPS узлом кластера.

Требование масштабируемости удовлетворяется СИМ NPS за счет увеличения максимального количества узлов в модели сети при добавление новых NPS узлов кластера. СИМ NPS сама решает, как будет разбита топология модели сети на NPS узлах кластера. Это свойство существенно упрощает перестройку модели при изменении количества моделируемых узлов в сети, так как от пользователя требуется только предоставление вычислительных ресурсов (в виде NPS узлов кластера).

Возможный размер моделируемой NPS сети зависит от количества NPS узлов в кластере. Беря во внимание, что один NPS узел кластера может моделировать топологию размером до 2000 узлов (хостов или сетевых устройств), а современный сервер может поддерживать работу порядка 25 NPS узлов кластера, итого получаем порядка 50000 хостов, моделируемых одним сервером. Учитывая, что разработанная СИМ NPS обладает свойством масштабируемости, мы получаем возможность моделировать сети масштаба ГКС.

Алгоритм построение NPS кластера состоит из нескольких шагов:

1. установка SSH соединений с каждым NPS узлом кластера;
2. конфигурирование выделенного интерфейса при построении виртуального сегмента сети на каждом NPS узле кластера;
3. разделение стандартного скрипта инициализации системы с описанием структуры модели сети на части. Количество этих частей равно количеству NPS узлов кластера;
4. подготовка скрипта инициализации для соответствующих NPS узлов кластера;
5. отправка частей скрипта инициализации на соответствующие NPS узлы кластера для построения топологии каждой части сети на NPS узле кластера;
6. отправка вспомогательных скриптов для генерации и установлении факта получения специализированных сетевых пакетов. Данные скрипты будут в дальнейшем использоваться для моделирования процесса распространения ВПО;
7. после построения сетевой структуры NPS главная управляющая консоль начинает выполнение сценария, предусмотренного экспериментом;
8. после выполнения сценария главная управляющая консоль сворачивает все экземпляры системы на NSP узлах кластера и закрывает SSH соединения;
9. (Дополнительно) визуализация результатов.

Существенным моментом для пользователя любой СИМ является процесс описания входных данных для эксперимента. Немаловажно сколько времени у пользователя тратится на описание всех входных данных для эксперимента (описание топологии, ресурсов сети, списка функционирующих приложений, процесса маршрутизации трафика, начальной популяции ВПО и т.д.). Для СИМ NPS был разработан специализированный графический интерфейс пользователя, который позволяет существенно ускорить и упростить процесс описания модельного эксперимента.

Основные возможности графического интерфейса пользователя (ГИП) СИМ NPS:

- описание структуры сети в виде графа (рисунок 1):

  - описание графа модели сети «вручную»;
  - генерация случайного графа модели сети;
  - загрузка графа модели сети из текстового описания;

- описание ресурсов сети путем изменения параметров хостов, каналов и транзитных узлов графа модели сети;
- запуск ПКС контролера (при запуске эксперимента) с возможностью конфигурирования его параметров. Помимо стандартных параметров (IP адреса и порта контролера) имеется возможность задания набора приложений контролера, отвечающих за маршрутизацию в моделируемой сети. Такой подход позволяет быстро и гибко задавать роль каждого транзитного узла в модели сети, сделав его коммутатором, маршрутизатором, или другим сетевым устройством;
- задание списка сетевых приложений, ассоциированных с каждым хостом в модели сети. При этом имеется возможность управлять параметрами запуска сетевых приложений, а так же задавать профили фоновой активности (например, HTTP-приложениям отправлять случайные запросы на HTTP-сервера в сети, или SMTP-приложениям отправлять случайные письма между активными SMTP-клиентами запущенными в модели сети). Последнее позволяет задавать фоновую активность, что существенно увеличивает ценность результатов проводимого эксперимента;
- сохранение (и загрузка) эксперимента в специальный текстовый формат файла (формат *.nps).

Основой задачей ГИП СИМ NSP является визуализации результатов проведения эксперимента. При этом важна визуализация как модельных данных («выхода» модели), так и процесса работы СИМ NPS (процесса разбиения модели сети на узлы кластера, загрузка ресурсов узла кластера, статистическая информация и т.д.).

Описанная архитектура СИМ NPS удовлетворяет требованиям, описанным в первой главе. Так масштабируемость достигается за счет добавление новых NPS узлов кластера при подготовки эксперимента? точность — за счет использования аппарата легковесной виртуализации ОС Linux, а также NPS библиотеки сетевых приложений. Возможность системы автоматического разбиения топологии модели сети на NPS узлы кластера позволяет без особых временных затрат изменять количество узлов в модели сети.

В четвертой главе представлено экспериментальное исследование разработанной и реализованной СИМ NPS на примере моделирования динамики

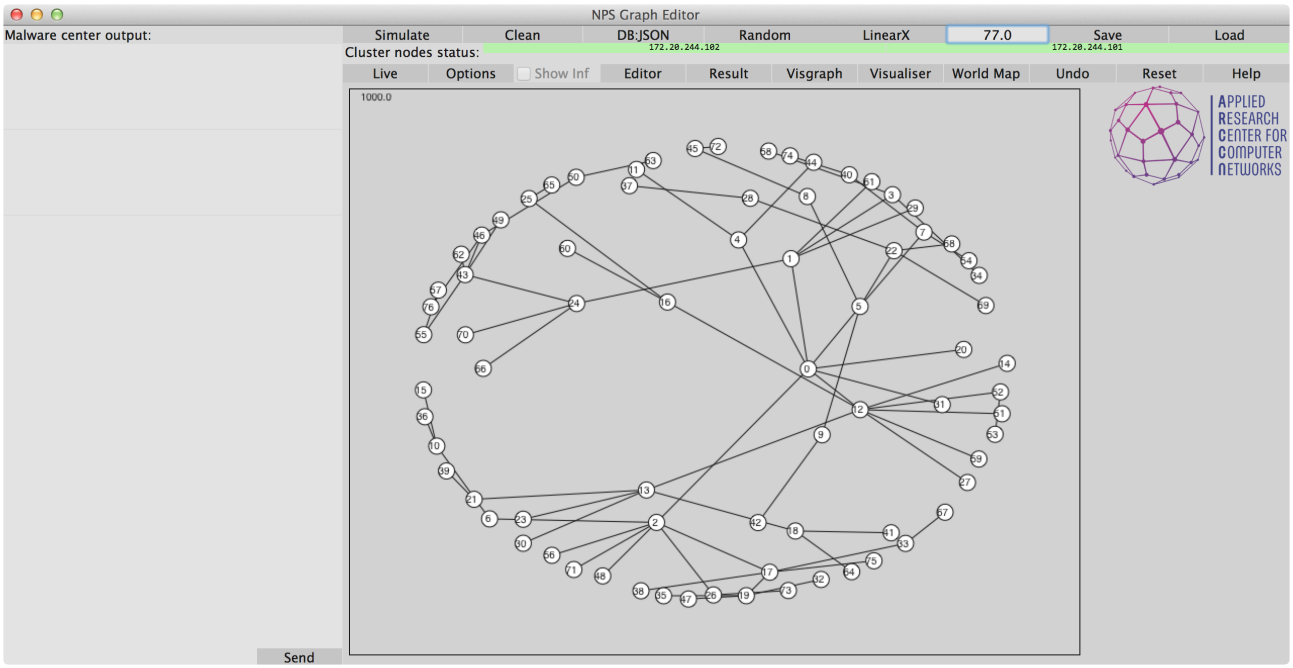

Рис. 1: Процесс редактирования графа модели сети.

распространения активных сетевых червей (один из типов ВПО). Продемонстрирована работа системы управления временем, входящей в состав СИМ NPS.

Экспериментальное исследование разбито на две части:

- экспериментальное исследование свойства масштабируемости разработанной СИМ NPS на примере моделирования динамики распространения ВПО;
- экспериментальное исследование синхронизации времени между компонентами СИМ NPS.

Возможность моделировать ГКС в СИМ NPS проверялась на примере распространения активных сетевых червей CodeRedv2 и Sasser. Данные сетевые черви являются одними из наиболее известных представителей своего класса. Использование в экспериментальном исследовании именно этих сетевых червей обусловлено наличием подробной аналитической информации либо о процессе распространения сетевого червя, либо о реализации экземпляра сетевого червя (например, алгоритм поиска новой жертвы, используемые порты при распространении или количество сетевых пакетов, содержащих экземпляр сетевого червя).

Сетевой червь — это один за классов ВПО. Его отличает самостоятельность предпринимаемых попыток найти новую жертву. Характеристикой этого класса ВПО является быстрое распространение в сети.

Потенциальными жертвами сетевого червя являются хосты, участвующие в сетевом взаимодействии. Хосты называются уязвимыми, если они обладают уязвимостью, используемой червем в процессе распространения. Соответственно, по аналогии, определяются неуязвимые хосты.

Цель первого эксперимента — продемонстрировать работу СИМ NPS с сетями масштаба ГКС, когда домен представляется более чем одним хостом. Этот случай интересен тем, что внутри домена для моделирования распространения ВПО будут использоваться эпидемические модели. Использование данного механизма позволяет моделировать сети большого размера, несмотря на отсутствие подробной информации о структуре сети или нехватки вычислительных ресурсов.

В экспериментальной топологии, представленной на рисунке 2 каждая вершина в графе является автономной системой (АС). Приведенная топология построена по данным о связанности АС за июль 2001 года (время эпидемии червя CodeRedv2). В текущем эксперименте с позиции СИМ NPS каждая АС является доменом. Распределение уязвимых хостов по доменам представлено в таблице 3.

Таблица 3: Распределение хостов по доменам.

Домен	Число уязвимых хостов	Начальная популяция
AS10	42.845	0
AS20	47.354	0
AS30	40.862	0
AS50	38.257	217
AS60	50.768	278
AS70	30.591	0
AS100	52.945	0
AS200	45.846	0
Итого	359.477	495

В результате моделирования процесса распространения червя CodeRedv2 в СИМ NPS получаем число зараженных хостов. Помимо общего числа зараженных хостов СИМ NPS фиксирует динамику изменения данного параметра, что позволяет сравнить моделируемый процесс распространения червя CodeRedv2 с данными о реальной эпидемии червя CodeRedv2. Результаты данного сравнения представлены на графиках (см. рисунки 3, 4).

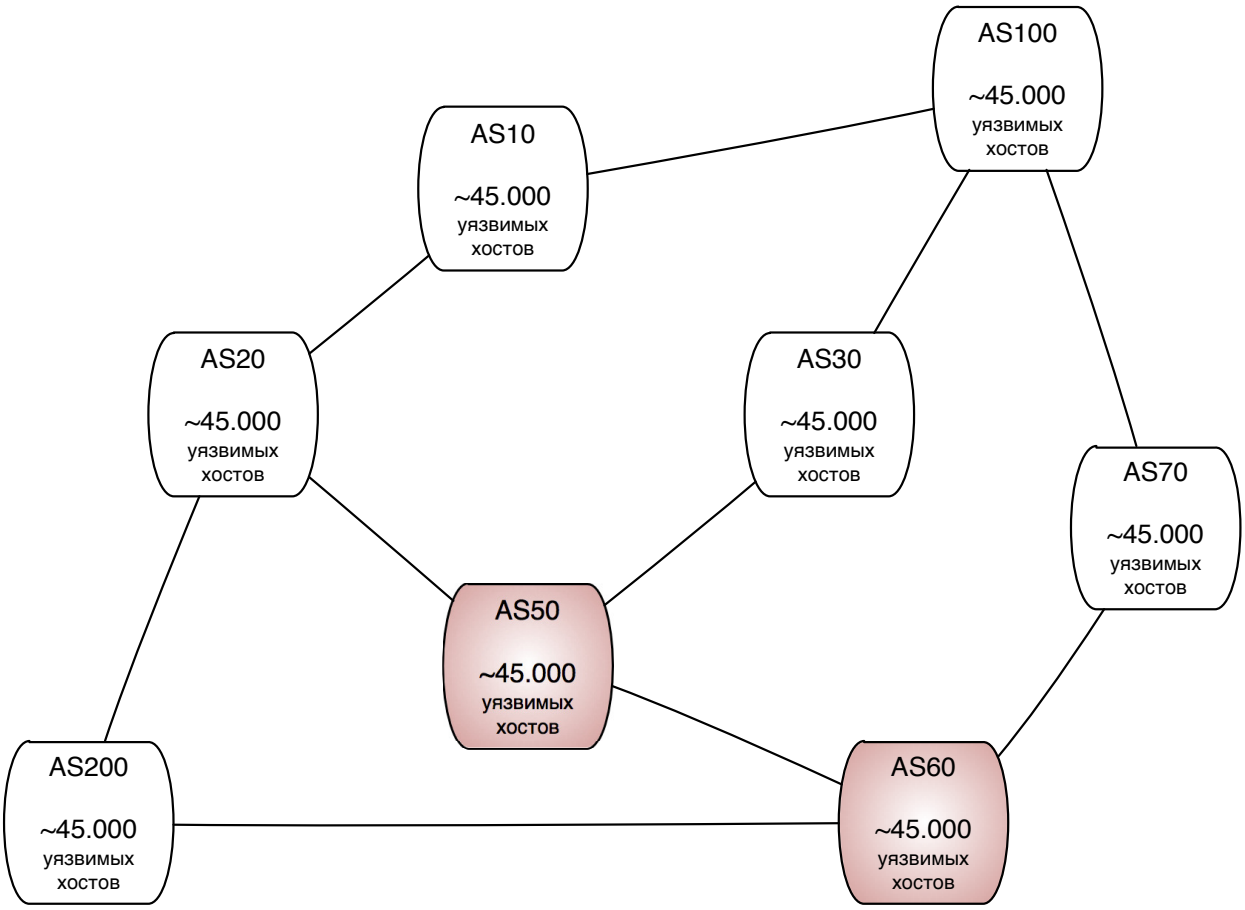

Рис. 2: Экспериментальная топология сети.

В результате моделирования распространения сетевого червя CodeRedv2 были заражены все уязвимые хосты. При этом наблюдалась схожая динамика заражения с реальной эпидемией CodeRedv2.

Проведенный эксперимент наглядно показывает возможности масштабирования математической модели ГКС, создаваемой при помощи СИМ NPS, за счет повышении уровня абстракции узла: с хоста до домена (набора хостов). Стоит отметить, что от уровня абстракции зависит точность имитационного моделирования ГКС, однако в данном эксперименте потеря в точности моделирования сетевого обмена между хостами внутри домена не влияет на успешность результатов, так как сравнение идет с обобщенными данными о реальном распространении сетевого червя CodeRedv2, в которых отсутствует информация о топологии сети.

Данные результаты были опубликованы в работе [1].

Цель второго эксперимента — доказать адекватность модели функционирования ГКС, построенной в СИМ NPS, когда домен представляется единичным хостом. Этот случай интересен тем, что на каждом хосте полностью моделируются все стадии распространения сетевого червя на примере Sasser.

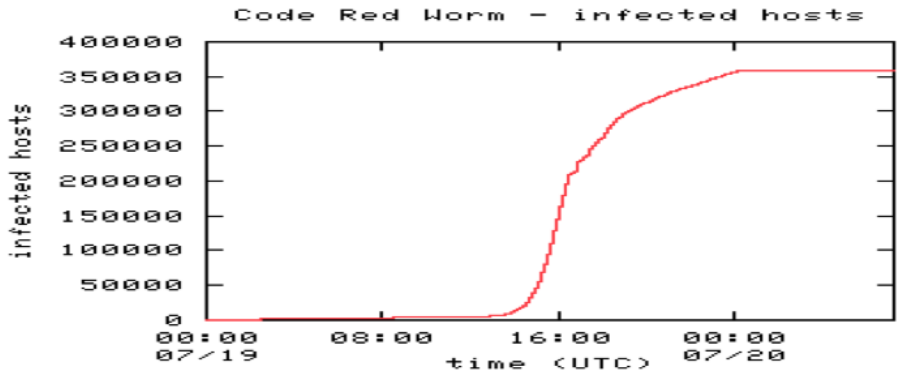

Рис. 3: Реальные данные о распространению червя CodeRedv2.

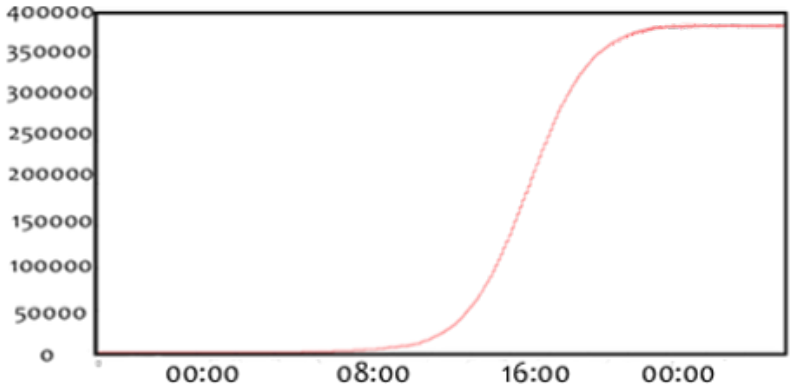

Рис. 4: Результаты моделирования распространения червя CodeRedv2.

Следовательно, моделирование динамики распространения сетевого червя будет максимально приближенным к реальному распространению Sasser.

В этом эксперименте считается, что домен состоит из одного хоста, далее по тексту будем использовать термин хост вместо термина домен.

Процесс распространения червя Sasser разделен на несколько шагов. Каждый шаг соответствует этапу жизненного цикла ВПО (выбор жертвы, сканирование, заражение, самораспространение и вредоносная активность)

Сетевой червь Sasser заражал компьютеры с уязвимой версией ОС Windows XP и Windows 2000. Sasser был впервые обнаружен 30 апреля 2004 года. Червь был назван Sasser из-за названия эксплуатируемого им сетевого приложения. В ходе своего распространения червь использовал ошибку переполнения буфера в компоненте ОС LSASS (Local Security Authority Subsystem Service). Червь сканировал различные наборы IP адресов и пытался соединиться с хостом жертвы через 445 TCP порт. Анализ исходного кода червя показал также, что в некоторых случаях использовался 139 порт.

В ходе эксперимента была случайным образом создана топология состоящая из 100.000 узлов. Данные узлы были распределены между 50 узлами NPS кластера примерно по 2000 вершин на каждый узел. Узел NPS кластера представлял из себя виртуальную машину на базе ОС Ubuntu.

В эксперименте червь Sasser использует случайный алгоритм поиска новой жертвы, следовательно топология незначительно влияет на результаты моделирования. Для доказательства этого утверждения была проведена серия экспериментов со случайными топологиями разных размеров. Количественные характеристики входных данных каждого эксперимента из серии приведены в таблице 4. Средняя скорость заражения измеряется в количестве хостов, зараженных за один цикл распространения сетевого червя.

Таблица 4: Количественные данные экспериментов с разными по размерам случайными топологиями.

Количество NPS узлов кластера	Число вершин	Средняя скорость заражения
10	7.500	210
30	30.000	850
100	100.000	2690

Из таблицы видно, что средняя скорость заражение растет пропорционально росту размера сети. Это обуславливается, как было сказано ранее, особенностью алгоритма распространения. Далее будем приводить результаты самого крупного эксперимента (100.000 хостов).

Начальная популяция сетевого червя Sasser составляла 5.257 хостов (хосты выбирались случайным образом). Скорость распространения равнялась одной попытке захвата жертвы за один круг жизненного цикла ВПО, описанного в разделе ??. Результаты моделирования представлены на графиках (см. рисунки 5, 6).

«Серый» (верхний) график на рисунке 5 представляет общее число уязвимых хостов. Поскольку предполагалось, что все хосты уязвимы к заражению, данное значение является константой. «Синий» (средний) график представляет число зараженных хостов в сети. Так как в ходе эксперимента никак не моделировались контрмеры против распространения червя, «синий» график монотонно возрастает. «Красный» (нижний) график представляет число успешных попыток заражения новых жертв в ходе одного шага распространения. Как можно было заметить, число таких попыток медленно, но растет (см. рисунок 6).

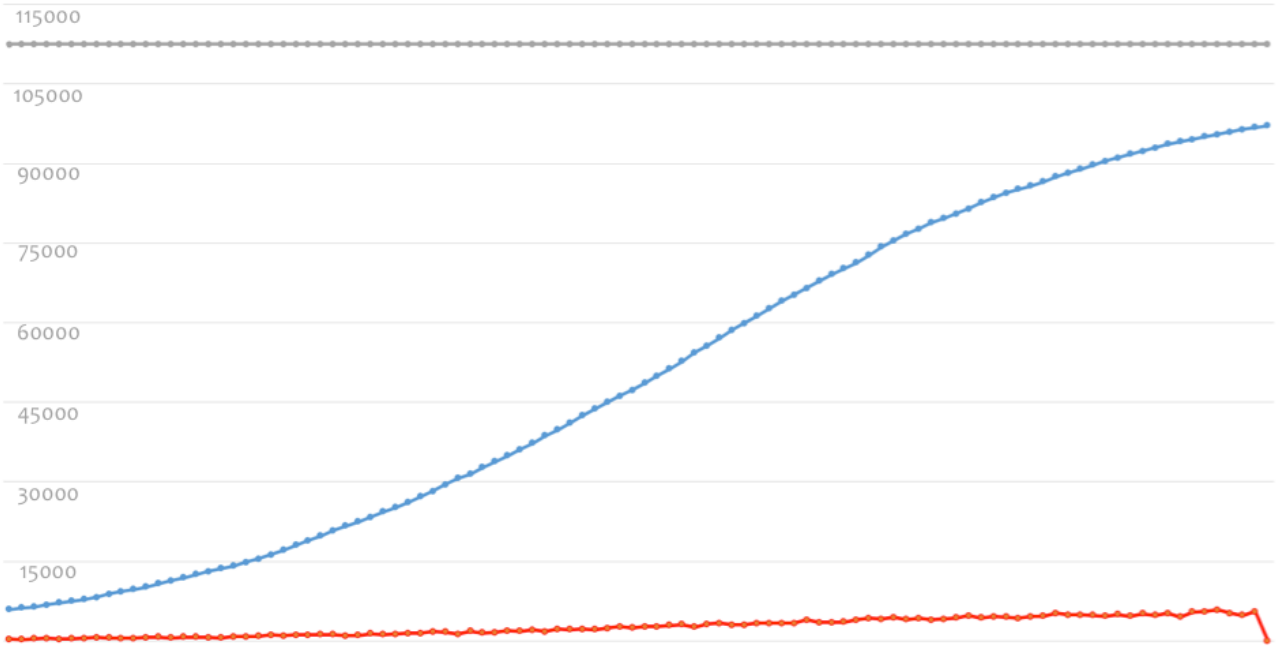

Рис. 5: Результаты моделирования распространения червя Sasser.

Полученные результаты были сопоставлены с результатами наблюдения за активностью, связанной с портом 445 в глобальной сети во время эпидемии червя Sasser 7. Черная вертикальная линия показывает период, когда компания Microsoft выпустила исправления, которые устраняют уязвимость в сетевом приложении LSASS. «Черный» график — это интерпретация результатов моделирования, описанных выше. Данные о реальной эпидемии сетевого червя Sasser носят косвенных характер, так как наблюдается только повышение обращений к порту 445. Поэтому корреляцию данных можно только визуально сопоставить на графике.

В результате данного эксперимента показана возможность СИМ NPS моделировать все этапы жизненного цикла ВПО. Продемонстрирована возможность работы СИМ NPS с ГКС без применения механизма абстрагирования от уровня хоста до уровня домена.

Проведенные эксперименты с моделями распространения ВПО наглядно продемонстрировали возможности СИМ NPS моделировать функционирование сетей машстаба ГКС.

В этих экспериментах был воспроизведен процесс полного жизненного цикла ВПО, что позволяет говорить о СИМ NPS, как об уникальной Hi-Fi системе распространения ВПО.

Стоит отметить, что моделирование распространения ВПО является одним из многих вариантов использования СИМ NPS. Разработанная система может использоваться, например, для исследования корректности функ-

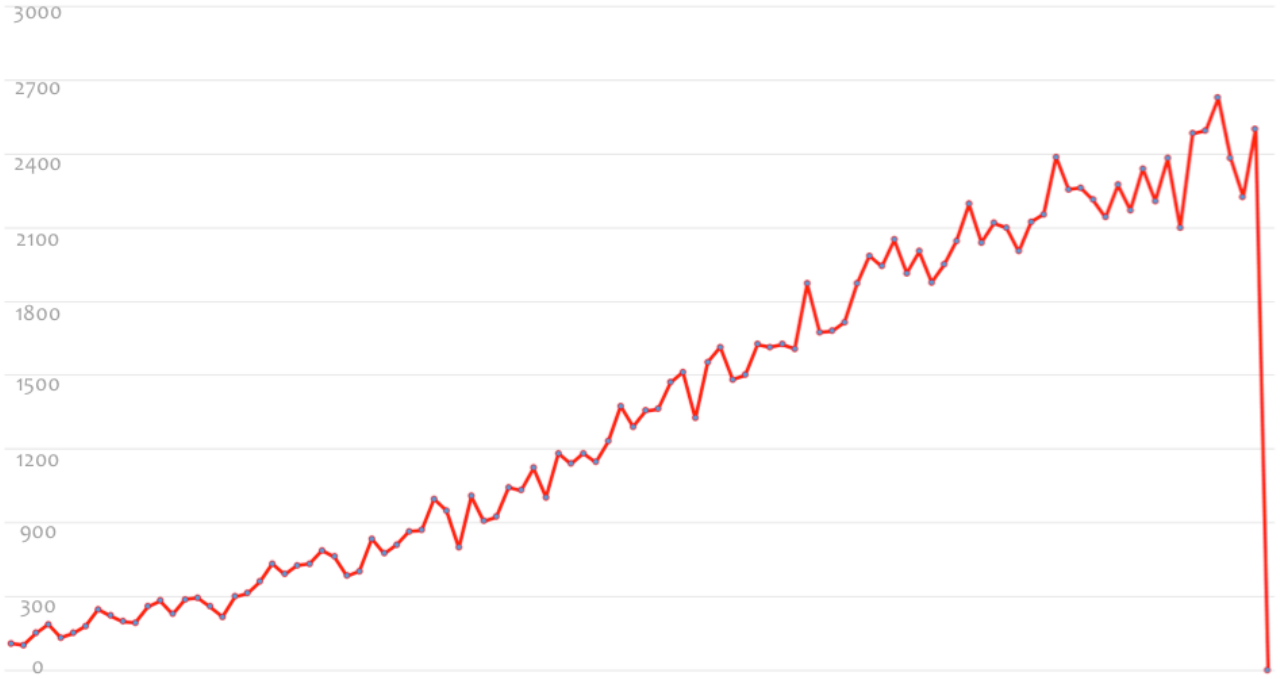

Рис. 6: График скорости заражения новых жертв червем Sasser. По оси абсцисс — время распространения; по оси ординат — средняя скорость заражения.

ционирования новых протоколов маршрутизации или построения тестовой площадки для новых сетевых приложений.

Данные результаты были опубликованы в работе [2].

Целью третьего эксперимента является демонстрация возможности системы управления модельным временем (СУМВ) синхронизации между частями модели, расположенными на разных узлах NPS кластера. Данный эксперимент доказывает корректность работы со временем в разработанной распределенной системе NPS. Под корректностью работы СУМВ понимается предотвращение потерь сетевого трафика из-за несоответствия временных характеристик модельных каналов с физическими каналами, поверх которых они прокладываются.

Эксперимент проводился на NPS кластере из двух узлов. В соответствии с процедурой развертывания модели поверх NPS кластера граф сети был разделен на две части, где:

- каждая из этих частей запустится на своем вычислителе в NPS кластере;
- как бы не было выполнено разбиение топологии моделируемой сети, один из каналов будет отображен на физический канал, соединяющий два вычислителя в NPS кластере.

При установлении DHCP сервера на заданный хост внутри одной из двух частей топологии сети, а также задания таймаута на подключение к

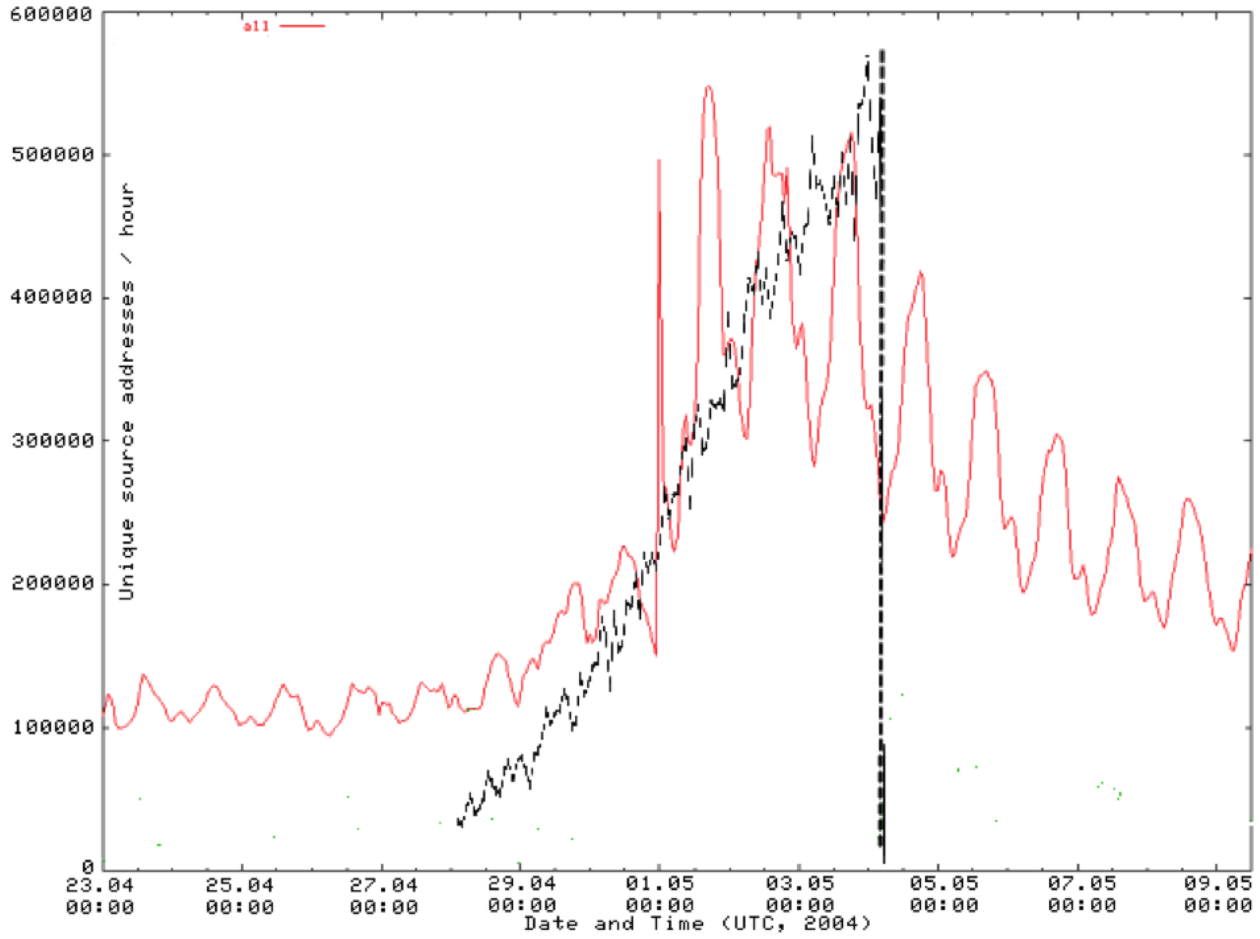

Рис. 7: Сравнение результатов моделирования с реальными данными об эпидемии червя Sasser.

DHCP серверу равным наибольшему значению RTT между хостами, получаем что:

- все хосты, которые находятся на одном с DHCP сервером узле NPS кластера, получили новые IP адреса;
- все хосты, которые находятся на разных с DHCP сервером узлах NPS кластера, остались со старыми IP адресами.

При использовании СУМВ происходит нормализация всех модельных каналов по найденному отклонению. Также нормализуются все временные характеристики сетевых приложений, функционирующих в модели сети. В итоге отмечаем, что поведение модели де-факто соответствует ожидаемому поведению, так как всеми хостами получен новый IP адрес. Платой за корректность в работе с модельным временем является увеличение астрономического времени проведения эксперимента. Так время проведения эксперимента увеличилось с 88.35 сек до 187.6 сек.

Представленный эксперимент демонстрирует успешное решение проблемы несоответствия временных характеристик физических и моделируемых каналов.

Экспериментальное исследование показало, что разработанная и реализованная распределенная система имитационного моделирования ГКС полностью удовлетворяет требованиям (точности моделирования процесса сетевого обмена данными и масштабируемости модели сети, создаваемой в СИМ NPS).

В ходе экспериментов с сетевыми червями было показано, что процесс распространения ВПО в модели практически полностью совпадает с данными о реальном распространении ВПО. Различия между моделью и реальностью минимальны и связанны с неполнотой данных о реальном распространении ВПО (информацией и топологии, либо о количестве уязвимых хостов).

Отдельно был проведен эксперимент с использованием СУМВ. В ходе этого эксперимента было продемонстрировано успешное решение проблемы рассинхронизации между различными частями модели сети, расположенными на разных вычислителях.

В заключении приведены основные результаты работы, которые заключаются в следующем:

1. Построена математическая модель, позволяющая корректно и адекватно моделировать функционирование ГКС. В рамках модели поставлена задача прогнозирования динамики распространения ВПО и доказано, что она имеет решение.

2. Предложен подход к имитационному моделированию сети на основе техник легковесной виртуализации, позволяющий строить имитационные модели сетей большого размера так, что процесс обработки и передачи сетевого трафика воспроизводится аналогично процессам обмена трафиком в реальной сети.

3. Разработана и реализована уникальная распределенная система имитационного моделирования (СИМ) сети, названная Network Prototype Simulator (NPS). В основе СИМ лежит аппарат легковесной виртуализации, который позволяет строить эффективно масштабируемые модели сети с высокой точностью воспроизведения процессов обработки и передачи сетевого трафика.

СИМ NPS обладает широкими перспективами развития. Возможные дальнейшие направления развития NPS заключаются в:

- расширении списка сетевых приложений в NPS библиотеке сетевых приложений;
- создании версии для других сетевых стеков, используемых при соединении NPS узлов кластера, например, это может быть полезно для запуска процесса моделирования на суперкомпьютерных вычислителях с использованием Infiniband-соединений;
- разработке перспективных средств визуализации больших топологий компьютерных сетей на базе графического интерфейса пользователя СИМ NPS.

Предложенный в работе подход к имитационному моделированию функционирования ГКС открывает новые возможности для исследователей и разработчиков новых сетевых решений по тестированию и построению сетевого окружения. СИМ NPS является открытой, уникальной площадкой для тестирования сетевого оборудования и сетевых протоколов, а так же универсальной исследовательской площадкой для изучения различных сетевых активностей в ГКС.

Публикации автора по теме диссертации

1. Antonenko Vitaly, Smelyanskiy Ruslan. Global Network Modelling Based on Mininet Approach. // Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking. HotSDN '13. New York, NY, USA: ACM, 2013. С. 145–146. URL: http://doi.acm.org/10.1145/2491185.2491211.
2. Антоненко В.А., Смелянский Р.Л. Моделирование вредоносной активности в глобальной компьютерной сети // Программирование. 2013. Т. 1. С. 60–72.