

Метод и технология системно-динамического моделирования нагрузки на региональные информационно-коммуникационные сети

И. О. Датьев, В. А. Путилов, А. М. Федоров

Институт информатики и математического моделирования технологических процессов Кольского НЦ РАН, Апатиты

Моделирование нагрузки на различные элементы информационно-коммуникационной сети может использоваться при разработке рекомендаций по техническим мероприятиям, направленным на улучшение функционирования сети в целом [1, 2]. Знания о нагрузке, генерируемой пользователями, необходимы для эффективного развития информационных услуг. В этом контексте особенно важным является получение перспективных оценок, способствующих формированию наилучших стратегий развития компаний, предоставляющих информационные услуги. Кроме того, в современных условиях такие оценки необходимы и для эффективного решения задач прогнозирования и выработки стратегий информационного развития регионов в целом, анализа существующего состояния информационных услуг, разработки и внедрения новых услуг и информационных технологий.

Меры, направленные на развитие сети, могут носить как оперативный характер (быстрая модификация для решения существующих проблем), так и стратегический (некоторые превентивные меры для сведения к минимуму затрат на модификацию сети в будущем). При принятии решений по выбору и рациональному обоснованию технических мероприятий, носящих стратегический характер, необходимо уметь прогнозировать различные характеристики, влияющие на принятие решений и зависящие от этих решений.

В случае, когда накоплено достаточное количество статистических данных за некоторый промежуток времени, для прогнозирования значений основных показателей можно использовать методы математической статистики. Главной проблемой при этом является необходимость оперирования большими объемами статистических данных за длительный период времени (рекомендуемый в методиках экономического анализа период — не менее 10 лет). Зачастую, таких статистических данных не существует.

Однако даже наличие необходимого количества статистических данных не является решением проблемы прогнозирования нагрузки на элементы сетей. Это связано, с периодическим обновлением технологий передачи данных, существенно влияющих на производительность сетей и на статистические характеристики информационно-коммуникационного трафика. Таким образом, с внедрением новых технологий, накопленные статистические данные становятся неактуальными для прогнозирования сетевой нагрузки. Как правило, длительность периодов до очередной смены технологий передачи данных в информационно-коммуникационных сетях не превышает 5–7 лет.

Поэтому представляется целесообразным использовать моделирование с целью среднесрочного (5–7 лет) прогнозирования развития сети при различных исходных значениях воздействующих факторов.

Нагрузка на информационно-коммуникационные сети зависит от показателей пользовательской среды. Основным показателем пользовательской среды в системах моделирования на сегодняшний день является количество пользователей [3–5]. Каждый пользователь создает некоторую нагрузку на сеть в зависимости от различных факторов. Одним из этих факторов является социально-экономическое положение пользователя: различные цели использования сети и, соответственно, различные информационные ресурсы порождают различный сетевой трафик. Для решения задач прогнозирования нагрузки необходимо также прогнозировать число пользователей сети различных социально-экономических групп, т. е. прогнозировать значения параметров социально-экономической среды региона в целом. Одним из эффективных методов изучения сложных социально-экономических систем в настоящее время, успешно развивающимся во многих странах, является предложенный в 1960-х годах Джеймсом Форрестером специализированный метод имитационного моделирования — метод системной динамики [6]. Метод системной динамики позволяет исследовать поведение сложных систем, опираясь на возможности компьютерного моделирования. В отличие от «традиционных» методов компьютерного моделирования системная динамика не требует построения математической модели исследуемого объекта в традиционной форме,

а дает исследователю инструментарий для моделирования в виде реализованных на компьютере аналитических описаний системных элементов и связей между ними.

Таким образом, в условиях ограниченной доступности достоверных статистических данных о региональных информационно-коммуникационных сетях, а также высоких требований к оперативности получения результатов, наиболее приемлемым способом определения нагрузки на сеть, оказывается построение системно-динамической модели региональной информационно-коммуникационной сети, основанной на легкодоступных (в смысле скорости и стоимости) данных. Подобная модель, разумеется, не будет обладать высокой точностью, однако будет существенно превосходить другие методы получения оценок по скорости и стоимости получения результата.

Для решения проблемы прогнозирования трафика информационно-коммуникационных сетей региона предлагаемым методом, необходимы следующие модели:

- Системно-динамические модели региональной социально-экономической среды, позволяющие получать прогностические оценки численности групп пользователей региональных информационно-коммуникационных сетей.
- Статистических моделей информационных ресурсов, определяющих классы ресурсов и структуру контента для каждого класса.
- Статистических моделей пользователей сети различных групп в терминах классов потребляемых информационных ресурсов.

Модель региональной социально-экономической среды

Информационно-коммуникационные системы регионального масштаба являются довольно крупными системами, их реализация является долгосрочным процессом. Вследствие этого, такие системы должны проектироваться исходя из, по крайней мере, среднесрочного прогноза нагрузки на их элементы. Поэтому при построении моделей информационных систем так же необходимо учитывать динамику показателей, влияющих на пользовательскую среду. К таким показателям могут быть отнесены:

- демографические;
- социальноэкономические;
- показатели, описывающие экономическое и техническое состояние региональных информационных систем.

Для решения данной проблемы была разработана системно-динамическая модель региона, позволяющая получать прогнозные оценки вышеописанных показателей.

Модель социально-экономической системы региона представляет собой совокупность моделей территориальных образований (населенных пунктов), составляющих регион. Модель каждого населенного пункта состоит из следующих компонентов:

- модель демографического сектора,
- модели социоэкономических групп,
- модели поставщиков информационных услуг.

Модель демографического сектора позволяет получать прогнозные оценки количества населения различных возрастных категорий [9].

Модели социоэкономических групп пользователей позволяют получать прогнозные оценки количества населения, принадлежащего каждой из групп.

Социоэкономические группы можно разделить на два основных вида:

1. Не зависящие от выбора человека.
2. Зависящие от выбора человека.

Примером групп первого вида являются группы «школьники», «пенсионеры». Примерами групп второго вида являются группы, занятые в различных экономических отраслях, обществах, организациях.

Базовой компонентой модели «Группы пользователей» является модель одной группы. Ключевыми атрибутами модели «Группа пользователей» являются:

- численность группы,
- популярность (престиж).

Разбиение пользователей на группы может производиться на основе различных социоэкономических аспектов, таких как:

- занятость в экономической отрасли,
- принадлежность организации,
- принадлежность социальной группе.

Модель экономических отраслей позволяет получать прогнозные оценки относительно количества населения, занятого в каждой из отраслей.

Тенденции развития экономических отраслей могут быть заданы экспертом. Модель каждой экономической отрасли реализуется посредством параметризации шаблона системно-динамической модели. При необходимости этот шаблон может быть расширен путем добавления специфических для моделируемой отрасли параметров, влияющих на ее развитие.

В основе модели лежат следующие характеристики:

- количество населения, задействованного в данной отрасли,
- уровень зарплаты по отрасли,
- уровень привлекательности,
- уровень трудоемкости получения единицы условного продукта.

Выбор отраслей, входящих в системно-динамическую модель социально-экономической системы, производится на основе таких оценок, как численность отрасли и интенсивность использования информационно-коммуникационных сетей ее типичными представителями.

Практически любая отрасль, с точки зрения населения, тем привлекательнее, чем выше уровень заработной платы и ниже уровень трудоемкости в этой отрасли.

Модель социальных групп пользователей имеет структуру, схожую с моделью экономических отраслей. Ключевым параметром данной модели также является численность каждой из групп. Разработаны шаблоны базовых социальных групп: «школьники», «пенсионеры», «безработные».

Входными данными для моделей социоэкономических групп пользователей являются прогнозные оценки количества населения различных возрастных групп, полученные в модели демографического сектора.

Разработанные модели поставщиков информационных услуг позволяют получать прогнозные оценки количества пользователей различных категорий для каждого поставщика информационных услуг. Входными данными для моделей поставщиков информационных услуг являются прогнозные оценки количества населения, занятого в каждой из отраслей, или принадлежащего определенной группе, полученные в моделях экономических отраслей или социальных групп.

Модель каждого поставщика информационных услуг реализуется посредством параметризации шаблона системно-динамической модели. Параметризация может быть осуществлена как на основе реального состояния регионального рынка информационных услуг, так и на основе экспертных оценок. В основе модели лежат следующие параметры:

- количество пользователей различных групп,
- цена информационных услуг,
- технические характеристики предоставляемых услуг.

Разработанные модели в дальнейшем используются при создании и реализации технологии системно-динамического среднесрочного прогнозирования нагрузки на элементы информационно-коммуникационных сетей.

Модели пользователя и информационного ресурса

Нагрузка на региональную информационно-коммуникационную сеть может быть представлена в виде суммы нагрузок, генерируемых населенными пунктами, входящими в состав региона [7]. В свою очередь, для расчета нагрузки, генерируемой одним населенным пунктом необходимо учитывать нагрузку, создаваемую каждым из провайдеров населенного пункта. Трафик провайдера состоит из трафика пользователей провайдера. Можно выделить пользователей различных групп, генерирующих различающийся трафик. Зная количественное соотношение пользователей провайдера каждой группы можно вычислить суммарный трафик провайдера.

Пусть G — трафик региона, тогда $G = \{T_i\}$, где T_i — трафик i -того населенного пункта, $i = [1, n]$, где n — количество населенных пунктов региона, причем $\bigcap_{i=1}^n T_i = 0$.

Чтобы смоделировать трафик населенного пункта необходимо смоделировать трафик провайдеров данного населенного пункта: $T_i = \{P_{ij}\}$, где P_j — трафик j -го провайдера i -того населенного пункта, $j = [1, m]$, m — количество провайдеров i -того населенного пункта, $\bigcap_{j=1}^m P_{ij} = 0$.

Для моделирования трафика какого-либо провайдера необходимо смоделировать трафик, генерируемый некоторым количеством пользователей определенной группы: $P_j = \{U_l^k\}$, где U_l^k — трафик пользователей l -той группы, $l = [1, Z]$, Z — количество групп пользователей.

Таким образом, базовым компонентом для расчета нагрузки на информационно-коммуникационные сети региона является модель пользователя.

Разработанная модель пользователя представляет собой набор статистических характеристик, описывающих поведенческие особенности пользователя, а так же структурно-технические особенности используемых пользователем информационных ресурсов. К поведенческим характеристикам пользователя относятся: статистическое распределение начала сеансов обращения к информационным ресурсам, статистическое распределение продолжительности сеансов, статистическое распределение временных интервалов между запросами на использование объектов информационных ресурсов.

Характеристики модели пользователя, описывающие структурно-технические особенности информационных ресурсов следующие: процентное соотношение используемых типов ресурсов, количественные параметры, описывающие форматы и объемы данных, передающихся между узлами ИС при использовании информационных ресурсов различных типов.

Модель пользователя информационных систем позволяет разделять множество пользователей на группы с точки зрения различных социально-экономических аспектов (принадлежность пользователя к некоторой экономической отрасли или социальной группе, принадлежность пользователя к конкретной организации), предоставляя возможность наиболее точно учитывать цели использования ИС.

В результате, модель пользователя позволяет учитывать не только общее количество пользователей информационных систем, но и специфику информационных ресурсов, необходимых для решения задач типичного представителя различных групп пользователей.

С точки зрения моделирования сетевого трафика необходимо выделять такие типы ресурсов, которые порождают различный трафик. По отношению к контенту данных ресурсов, можно сказать, что они должны обладать некоторыми структурными (дизайнерскими) особенностями — соотношением текста и графики на веб-страницах, размеры загружаемых веб-страниц [8].

Атрибуты модели ресурса могут быть представлены следующим образом:

$R = \{P, O\}$, где P — количество объектов на странице; O — размеры объектов. Описание ресурсов в таком виде позволяет исследовать процессы, связанные с генерацией трафика различными категориями пользователей в отношении различных типов ресурсов.

Зачастую в существующих работах, связанных с моделированием трафика отсутствует связь между моделью пользователя и моделью ресурса: предполагается, что статистические распределения, описывающие параметры данных моделей являются независимыми, что на самом деле практически никогда не выполняется. В данной работе такая связь осуществляется посредством добавления в модель дополнительных параметров.

В модель ресурса добавляется параметр, определяющий множество групп информационных ресурсов определенного характера $S = \{s_i\}$. В этом случае группа представляет собой агрегированную характеристику, являющуюся сочетанием количества и объема объектов ресурса и поставленным ей в соответствие идентификатором.

Вместе с тем в модель пользователя добавляется параметр, определяющий соотношение типов ресурсов предпочитаемых пользователем: $E = \{s_i, d_i\}$, где $s_i \in S$, d_i — доля в трафике пользователя. В результате, множество параметров модели ресурса R и модели пользователя U представляются следующим образом: $R = \{P, O, S_i d\}$ $U = \{N, T, E, q, u, U_i d\}$, где:

N — статистическое распределение количества запрошенных в течение сеанса работы веб-страниц;

T — статистическое распределение времени между запросами страниц;

P — статистическое распределение количества объектов на странице;

O — статистическое распределение размеров объектов;

$S_i d$ — идентификатор типа ресурса определенного характера;

$E = \{s_i, d_i\}$ — соотношения типов предпочитаемых ресурсов, где $s_i \in S$ (S — множество типов информационных ресурсов); d_i — доля в трафике пользователя;

$U_i d$ — идентификатор группы пользователя;

q — задает предпочтение пользователя: q близко к 1, важна цена, q близко к 0 — важна скорость;

u — значимость времени бесперебойной работы поставщика информационных услуг.

Таким образом, модель ресурса и модель пользователя являются взаимосвязанными моделями, на основе которых можно строить достаточно эффективные модели трафика и проводить исследования процессов, связанных с сетевым трафиком.

Технология моделирования нагрузки на информационно-коммуникационную сеть

Последовательность действий, которую необходимо выполнить для моделирования нагрузки схематично изображена на рис. 1.

Алгоритм построения модели нагрузки на информационно-коммуникационную сеть региона, представленный в виде словесного описания выглядит следующим образом:

1. На основе статистических данных о населении региона задать исходные значения параметров модели демографического сектора населенного пункта (территориального образования).
2. В общую схему модели населенного пункта добавить из библиотеки шаблоны системно-динамических моделей экономических отраслей и социальных групп, присутствующих в регионе. При необходимости, обусловленной специфичностью региона, модифицировать шаблоны, добавленные из библиотеки.
3. Задать исходные значения параметров моделей экономических отраслей и социальных групп.
4. Для модифицированных групп, задать соответствующие характеристики их пользователей в блоке расчета нагрузки.
5. На общей схеме модели населенного пункта разместить необходимое количество шаблонов системно-динамических моделей поставщиков информационных услуг.

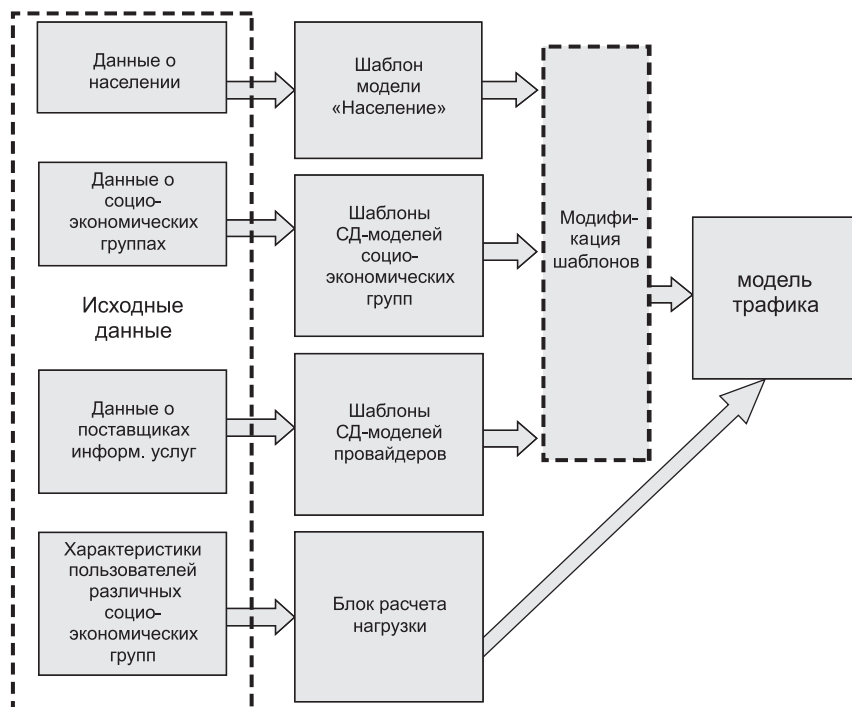


Рис. 1. Этапы моделирования трафика

6. Задать исходные значения параметров моделей поставщиков информационных услуг.
7. Для остальных населенных пунктов региона выполнить действия 1–7.
8. Задать временной диапазон и произвести прогон модели.

Результаты моделирования

Предлагаемый метод среднесрочного моделирования сетевой нагрузки был реализован в виде технологии на примере Апатитско-Кировского региона Мурманской области.

В качестве исходных данных для наполнения шаблонов моделей региональной социально-экономической среды использовались статистические данные и экспертные оценки следующего рода:

- количество населения различных возрастных групп, данные о смертности, миграции;

Таблица 1

Результаты моделирования сетевой нагрузки Апатитско-Кировского региона

Сетевая нагрузка, Гбайт/мес	2006	2007	2008	2009	2010	2011	2012
Апатиты	2328,512	3417,422	4321,485	4958,993	5401,164	5842,331	6196,247
Кировск	1209,095	1753,744	2245,813	2688,264	3053,656	3279,698	3454,872
Всего	3537,607	5171,166	6567,298	7647,257	8454,820	9122,029	9651,119

- оценка состава socioeconomicальных групп, тенденции развития, характеристики пользователей socioeconomicальных групп;
- информация о поставщиках информационных услуг (количество провайдеров, цены, технические характеристики).

Суммарные оценки сетевой нагрузки по городам, полученные в результате моделирования представлены в табл. 1.

Область применения

Реализованная технология позволяет получать прогностические оценки нагрузки на сеть среднесрочного периода и может быть полезной на следующих административных уровнях:

1. Продвинутым пользователям для прогнозирования привлекательности информационных услуг провайдеров.
2. Техническим работникам фирм-провайдеров (системным администраторам, инженерам) для получения перспективных оценок нагрузки на сеть провайдера.
3. Владельцам фирм, предоставляющих информационные услуги для прогнозирования и планирования развития регионального рынка информационных услуг (характеристики конкурентов, количество пользователей).
4. Лицам, принимающим решения в области планирования развития информационно-коммуникационных сетей регионального уровня (загруженность сетей, количество пользователей, характеристики рынка информационных услуг).

Разработанная технология обладает гибкостью в смысле расширения ее применения на классы задач, где могут быть выделены группы пользователей и определены предпочтения выделенными группами пользователей каких-либо ресурсов, причем использование каждой из групп

ресурсов обладает специфическими особенностями с точки зрения оказываемой нагрузки на сеть.

В целом, технология представляет собой механизм расчета следующих параметров:

- нагрузка на территориальные субъекты (сеть вторичного провайдера, сеть населенного пункта, сеть региона);
- объем трафика, ассоциированный с каждой из групп пользователей;
- объем предоставленных информационных услуг в денежном эквиваленте каждым провайдером (суммарная оценка, оценка по группам).

Заключение

В работе предложен метод среднесрочного прогнозного моделирования нагрузки на информационно-коммуникационные сети региона. При создании технологии на основе данного метода использован комплексный подход, позволяющий учитывать различные аспекты, влияющие на формирование нагрузки на информационно-коммуникационные сети:

- контент информационных ресурсов, влияющий на сетевой трафик, генерируемый при использовании этих ресурсов;
- различная интенсивность влияния разных групп пользователей на формирование трафика информационно-коммуникационных сетей;
- динамика развития социально-экономической среды для прогнозирования численности групп пользователей.

В процессе разработки и реализации предлагаемого метода моделирования получены следующие результаты:

- разработаны системно-динамические модели региональной социально-экономической среды, позволяющие получать прогностические оценки, связанные с численностью групп пользователей региональных информационно-коммуникационных сетей;
- разработаны статистические модели информационных ресурсов, определяющие классы ресурсов и структуру контента для каждого класса;
- разработаны статистические модели пользователей сети различных групп в терминах классов потребляемых информационных ресурсов, позволяющие моделировать трафик, генерируемый отдельным пользователем каждой группы;
- на основе системно-динамических моделей региональной социально-экономической среды, моделей информационных ресурсов и пользователей сети различных групп создана технология среднесрочного прогнозного моделирования трафика региональных информационно-коммуникационных сетей;

- на основе статистических данных о составе населения, характеристиках информационных услуг Апатиты-Кировского региона, реализована технология построения модели региональной информационно-коммуникационной сети на примере этого региона.

Литература

1. Основы построения больших информационно-вычислительных сетей // Под общей ред. Д. Г. Жимерина и В. И. Максименко. М.: Статистика, 1976. 296 с., ил.
2. Вишневецкий В. М. Теоретические основы проектирования компьютерных сетей. М.: Техносфера, 2003. 512 с.
3. Jin Cao, Cleveland William S., Yuan Gao, Jeffay K., Smith Donelson F., Weigle M. Stochastic Models for Generating Synthetic HTTP Source Traffic // Proceedings of IEEE INFOCOM. Hong Kong, March 2004.
4. Fernandos-Campos H., Jeffay K., Smith Donelson F. Tracking The Evolution Of Web Traffic // In Proceedings IEEE/ACM MASCOTS 2003. P. 16–25.
5. D-ITG, Distributed Internet Traffic Generator. Other Internet Traffic Generators. [Электронный ресурс] <http://www.grid.unina.it/software/ITG/link.php>
6. Форрестер Дж. Динамика развития города М.: Прогресс, 1974. 285 с.
7. Шишаев М. Г., Датьев И. О., Шемякин А. С. Методика оценки уровня информационных потребностей региональных пользователей // Информационные ресурсы России. 2006. № 6(94). С. 29–32. -ISSN 0204–3653.
8. Датьев И. О. Анализ TCP-трафика в задаче мониторинга информационной активности пользователей корпоративной сети // Труды Института системного анализа Российской академии наук (ИСА РАН). Прикладные проблемы управления макросистемами. Т. 28. М.: КомКнига, 2006. 368 с.
9. Горохов А. В., Малыгина С. Н. Концептуально динамическая модель устойчивого развития типового города Севера России // Имитационное моделирование в исследованиях проблем регионального развития. Апатиты: Изд-во Кольского научного центра РАН, 2000. С. 25–34.