

Вирішується задача апроксимації контрольованих процесів в комп'ютерній мережі з метою зниження рівня службового трафіку, що генерується при моніторингу контрольованих процесів. Запропоновано метод оцінювання інформаційних моделей контрольованих процесів в комп'ютерній мережі. Метод дозволяє розрахувати значення параметрів інформаційних моделей, на підставі яких може бути обрана адекватна апроксимуюча модель контрольованого процесу

Ключові слова: апроксимація, моніторинг комп'ютерної мережі, спостережуваний процес, Windows Management Instrumentation (WMI)

Решается задача аппроксимации контролируемых процессов в компьютерной сети с целью снижения уровня служебного трафика, генерируемого при мониторинге контролируемых процессов. Предложен метод оценивания информационных моделей контролируемых процессов в компьютерной сети. Метод позволяет рассчитать значения параметров информационных моделей, на основании которых может быть выбрана адекватная аппроксимирующая модель контролируемого процесса

Ключевые слова: аппроксимация, мониторинг компьютерной сети, наблюдаемый процесс, Windows Management Instrumentation (WMI)

УДК 004.738.52:004.031

DOI: 10.15587/1729-4061.2015.36277

МЕТОД ОЦЕНИВАНИЯ ИНФОРМАЦИОННЫХ МОДЕЛЕЙ НАБЛЮДАЕМЫХ ПРОЦЕССОВ В КОМПЬЮТЕРНОЙ СЕТИ

А. И. Гриценко

Аспирант

Кафедра информационных управляющих систем
Харьковский национальный
университет радиоэлектроники
пр. Ленина 14, г. Харьков, Украина, 61000
E-mail: alegri@inbox.ru

1. Введение

Мониторинг компьютерной сети – это множество мероприятий организационно и технического характера, осуществляемых для решения задач сбора, хранения и анализа информации о состоянии элементов компьютерной сети. Решение этих задач необходимо для обеспечения требуемой функциональности компьютерной сети: своевременного обнаружения и исправления неисправностей, сбора статистической информации, обеспечения требуемого уровня качества сервиса и т. д.

Существуют различные подходы и способы реализации мониторинга сети [1–3], зависящие от целей мониторинга, технических условий, организационных возможностей. Большой выбор программных средств позволяет решать практически любые по степени сложности задачи удаленного сбора и анализа данных.

2. Анализ литературных данных и постановка проблемы

Как правило, сбор данных во время мониторинга компьютерной сети производится централизованно с помощью клиент-серверных технологий. С одной стороны это повышает оперативность и надежность данных, с другой стороны приводит к появлению дополнительного служебного трафика, снижающего по-

лезную пропускную способность сети. Важно выбрать правильный способ сбора данных, позволяющий не только обеспечивать требуемый уровень оперативности и точности данных, но и приемлемый уровень служебного трафика. Для решения этой проблемы разрабатываются различные методы, перераспределяющие нагрузку на сеть [4, 5], использующие компрессию [6] или аппроксимацию [7, 8]. Для более глубокого понимания проблемы и поиска новых способов ее решения проводится анализ существующих средств мониторинга [9, 10].

Мониторинг компьютерной сети неизбежно связан с накоплением данных, соответствующих результатам наблюдений за определенными переменными $v_i \in V$, где V – все множество наблюдаемых переменных. В итоге, для каждой переменной можно получить представление в виде процесса $\varphi_i(t)$. В ряде задач мониторинга компьютерных сетей требуется аппроксимация $\varphi_i(t)$ [7, 8]. Для этого строится соответствующая аппроксимирующая модель $f_i(t)$. Построение $f_i(t)$ состоит из этапов структурной и параметрической идентификации, для которых важна информация о поведении $\varphi_i(t)$. От точности, полноты и непротиворечивости $f_i(t)$ зависит эффективность решения задачи аппроксимации, а следовательно, и задачи мониторинга сети в целом.

Эффективность аппроксимации наблюдаемых процессов с помощью существующих методов (например, [11–13]) значительно зависит от качества

экспертного оценивания. Только при наличии необходимой и достаточной информации о поведении случайного процесса $\varphi_i(t)$ можно построить адекватную аппроксимирующую модель. Основываясь на том, что все процессы в компьютерной сети заранее частично определены, в работе [14] был сформирован набор значимых для аппроксимации характеристик наблюдаемых процессов, на основании которых были разработаны две информационные модели наблюдаемых процессов.

Актуальной проблемой рассматриваемых информационных моделей является отсутствие формализованного подхода к их оцениванию. Решение проблемы предполагает разработку метода, позволяющего последовательно вычислить параметры моделей с помощью специальных правил, описанных в работе [14].

3. Цель и задачи исследования

Целью работы является разработка формализованного подхода к оцениванию информационных моделей наблюдаемых процессов. В работе [14] дано определение рассматриваемым информационным моделям.

Для достижения поставленной цели рассматривается следующая постановка задачи: предположим, что планируется мониторинг некоторых процессов в компьютерной сети. Пусть известно, что во время мониторинга будут аппроксимироваться наблюдаемые процессы с целью минимизации уровня служебного трафика. При этом для повышения эффективности аппроксимации формируются специальные информационные модели [14].

Основной задачей является разработка метода оценивания информационных моделей наблюдаемых процессов в компьютерной сети.

Для достижения поставленной цели необходимо выполнить следующие задачи:

1. Согласовать этапы оценивания моделей с этапами организации мониторинга компьютерной сети.
2. Проанализировать и улучшить правила оценивания моделей.
3. Разработать алгоритм вычисления значений параметров моделей.
4. Разработать и описать метод оценивания моделей.

4. Материалы и методы исследования способов оценивания информационных моделей наблюдаемых процессов в компьютерной сети

4.1. Положения об информационных моделях наблюдаемых процессов в компьютерной сети

Рассматриваются информационные модели наблюдаемых процессов в компьютерной сети, предложенные в статье [14]. В статье указывается, что оценивание моделей осуществляется последовательно в два этапа (рис. 1): на этапе предварительного оценивания формируется априорная модель S_{-1} , на этапе экспериментального оценивания формируется апостериорная модель S_1 .

Однако в работе не рассматривается вопрос имплементации этапов оценивания моделей в контексте

общей схемы организации мониторинга. Основные этапы мониторинга описаны в [1]: инициализация, планирование, выбор инструментария, проведения измерений и анализ результатов. Исходя из того, что модели должны быть сформированы до проведения измерений, но после получения всей необходимой информации об объекте мониторинга, будем полагать, что этапы оценивания рассматриваемых информационных моделей должны следовать непосредственно после этапа выбора инструментария (рис. 2).

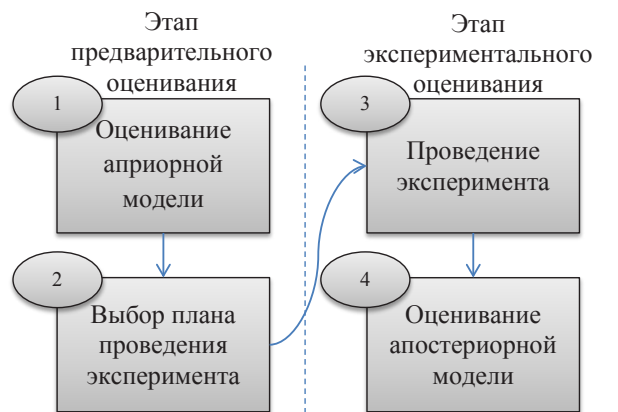


Рис. 1. Этапы оценивания информационных моделей

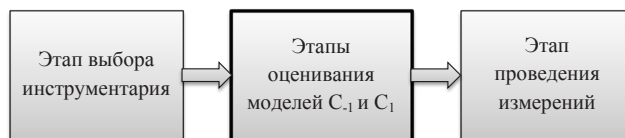


Рис. 2. Этапы оценивания информационных моделей в контексте этапов организации мониторинга компьютерной сети

Таким образом, для оценивания моделей S_{-1} и S_1 может быть использована информация, полученная на первых трех этапах организации мониторинга. Обозначим ее как спецификацию мониторинга $M = \{G, F, S, D, O, V, I\}$, где G – цели мониторинга, F – задачи мониторинга, S – среда мониторинга, D – предметная область, O – объекты мониторинга, V – наблюдаемые переменные, I – инструментальные средства мониторинга.

4.2. Описание этапа предварительного оценивания

Оценивание априорной информационной модели. Описанная в [14] модель S_{-1} оценивается экспертным путем на основании априорных данных, то есть исходной спецификации мониторинга M . Задача оценивания S_{-1} заключается в расчете определенных статистических характеристик наблюдаемых процессов. Формально эти характеристики описаны как параметры модели $S_{-1} = \{d_\varphi, l_\varphi, s_\varphi, t_\varphi\}$, где d_φ – динамичность, l_φ – линейность, s_φ – стационарность, t_φ – время жизни. Расчет параметров модели S_{-1} осуществляется с помощью использования правил R_{-1} в ходе графо-визуального анализа измеренных значений наблюдаемых процессов.

Формирование плана проведения эксперимента. Основным результатом этапа предварительного оце-

нивания является план проведения эксперимента E , состоящий из множества испытаний $e_i \in E$. Каждое испытание e_i представляет собой систематическое измерение наблюдаемых процессов, проводимое в соответствии со спецификацией $e_i = \{T, T_0, \Delta T_c, \Delta t_c, P_c\}$, где T – длительность испытания, T_0 – начальное время испытания, ΔT_c – длина интервала актуализации данных, Δt_c – длина интервала измерений, P_c – произвольный набор правил преобразований значений.

4. 3. Описание этапа экспериментального оценивания

Проведение эксперимента. Согласно выбранному плану проведения эксперимента E проводится ряд испытаний, в ходе которых, с помощью инструментальных средств I , измеряются наблюдаемые процессы V и формируется множество данных X_{ii} .

Оценивание апостериорной информационной модели. В соответствии с [14] модель C_1 формируется на основании спецификации мониторинга M , ранее оцененной модели C_{-1} и полученных в ходе эксперимента данных X_{ii} . Модель C_1 состоит из всех ранее перечисленных параметров модели C_{-1} , а также дополнительных параметров, отражающих математические характеристики функций наблюдаемых процессов. Модель описывается как $C_1 = \{d_\phi, l_\phi, s_\phi, t_\phi, l_f, c_f, m_f, u_f\}$, где d_ϕ – динамичность, l_ϕ – линейность, s_ϕ – стационарность, t_ϕ – время жизни, l_f – степень нелинейности, c_f – непрерывность, m_f – монотонность, u_f – выпуклость/вогнутость. Расчет параметров модели осуществляется с помощью специальных правил R_1 . Основным результатом этапа экспериментального оценивания является оцененная модель C_1 .

4. 4. Правила оценивания информационных моделей

В [14] не указывается, как использовать правила R_{-1} и R_1 при наличии выборок значений наблюдаемых процессов X_{-ii} и X_{ii} . Однако очевидно, что для получения адекватных априорной и апостериорной моделей, измеренные значения необходимо предварительно обработать. Будем использовать следующий алгоритм расчета параметров моделей на основании X_{-ii} и X_{ii} . В соответствии с ΔT_c выборки X_{-ii} и X_{ii} разбиваются на интервалы $\Delta T_c^{(i)}$. Для каждого интервала $\Delta T_c^{(i)}$ рассчитываются описанные в пп. 4.1, 4.2 характеристики наблюдаемых процессов с помощью правил R_{-1} и R_1 . Значения параметров моделей C_{-1} и C_1 выбираются на основании относительных или средних значений рассчитанных характеристик. Например, если все множество измеренных значений X_{ii} разбивается на 100 интервалов $\Delta T_c^{(i)}$ и для 80 % интервалов характеристика динамичности d_ϕ равна 1, то считаем процесс высокодинамичным. Все пороговые значения относительных показателей выбираются экспертным путем для каждого параметра модели. На этапе оценивания модели C_{-1} расчет априорных характеристик для интервалов $\Delta T_c^{(i)}$ проводится визуально, а правила R_{-1} носят рекомендательный характер.

Дополнительно к уточненным общим правилам оценивания моделей опишем улучшенные правила

множества R_{-1} , а именно правила расчета параметров динамичности, линейности и времени жизни.

Динамичность процесса (d_ϕ) будем оценивать на основании коэффициента вариации $CV_\phi = \sigma[X]/\mu[X]$, где σ – стандартное отклонение, μ – математическое ожидание, X – множество измеренных значений. Выберем пороговое значение коэффициента вариации CV_ϕ^* . Если $CV_\phi > CV_\phi^*$, то считаем процесс высокодинамичным ($d_\phi = 1$), иначе низкодинамичным ($d_\phi = 0$). Рекомендованное значение для $CV_\phi^* = 0.33$ (соответствует однородной совокупности).

Линейность процесса (l_ϕ) будем оценивать путем построения линейной регрессии и расчета ее коэффициент детерминации R^2 . Если $R^2 > R^{2*}$, то процесс считаем линейным ($l_\phi = 1$), иначе нелинейным ($l_\phi = 0$). Рекомендованное значение для $R^{2*} = 0.8$.

Время жизни процесса (t_ϕ) можно оценить на основании его абсолютного отклонения от среднего D . Если $D > 0$, то считаем процесс долговременным ($t_\phi = 1$), иначе кратковременным ($t_\phi = 0$).

5. Результаты исследований способов оценивания информационных моделей наблюдаемых процессов в компьютерной сети

5. 1. Описание метода оценивания информационных моделей наблюдаемых процессов в компьютерной сети

Предлагается метод оценивания информационных моделей наблюдаемых процессов, разработанных в [14]. Метод заключается в последовательном вычислении характеристик наблюдаемых процессов и формировании априорной и апостериорной информационных моделей. Входная и выходная информация, а также уточненные правила описаны в разделе 2 текущей статьи. Шаги метода, соответствующие этапам предварительного и экспериментального оценивания информационных моделей (рис. 1):

1. Формирование краткой экспертной характеристики наблюдаемых процессов.
2. Проведение кратковременной сессии измерений наблюдаемых процессов и представление полученных значений в графическом виде.
3. Оценивание априорной информационной модели $C_{-1} = \{d_\phi, l_\phi, s_\phi, t_\phi\}$ с помощью правил R_{-1} .
4. Выбор плана проведения эксперимента $E: e_i = \{T, T_0, \Delta T_c, \Delta t_c, P_c\}$, $e_i \in E$ на основании априорной модели C_{-1} и спецификации мониторинга M .
5. Проведение эксперимента в соответствии с выбранным планом E и формирование множества данных X_{ii} .
6. Оценивание апостериорной информационной модели $C_1 = \{d_\phi, l_\phi, s_\phi, t_\phi, l_f, c_f, m_f, u_f\}$ с помощью правил R_1 .

5. 2. Характеристика метода оценивания информационных моделей наблюдаемых процессов

Анализируя цели и задачи разработанного метода, особенности алгоритма его реализации и типы объектов, к которым он применяется, можно сформулировать его обобщенную характеристику.

Выделим следующие моменты, характеризующие предложенный метод:

1. Основывается на идее частичной априорной определенности наблюдаемых процессов.

2. Не требует высоких вычислительных затрат. Объем затрат прямо пропорционален количеству наблюдаемых процессов.

3. Может адаптироваться к изменениям среды мониторинга. Простота реализации метода и не высокие вычислительные затраты позволяют реализовать динамическую адаптацию информационных моделей к условиям изменения среды мониторинга.

4. Позволяет оценить уровень ошибочного оценивания моделей путем сравнения аппроксимируемых значений наблюдаемых процессов с реальными значениями измерений.

Таким образом, метод обладает гибкостью реализации, позволяющей учитывать особенности среды мониторинга, адаптировать к ее изменениям и оценивать риски ошибочных расчетов.

5. 3. Описание алгоритма метода оценивания информационных моделей наблюдаемых процессов

Реализация разработанного метода состоит из следующих шагов:

1. Формируем краткую экспертную характеристику наблюдаемых процессов V в произвольной описательной форме.

2. Проводим кратковременную сессию измерений наблюдаемых процессов V . Длительность измерений и длину интервала измерений выбираем такие, чтобы получаемые выборки были репрезентативными с экспертной точки зрения.

3. Для каждого наблюдаемого процесса $v_i \in V$ выбираем длину интервала актуализации данных ΔT_c . Значение ΔT_c зависит от характера поведения v_i . Если процесс высокодинамичный, нестационарный или не линейный, то рекомендуется выбирать небольшой интервал ΔT_c в диапазоне от 5 до 9 значений. Для низкодинамичного или линейного процесса ΔT_c может находиться в диапазоне от 10 до 20 значений.

4. Оформляем измеренные на 2 шаге значения в виде линейных графиков, на которых отмечаем интервалы $\Delta T_c^{(i)}$ в соответствии с ΔT_c .

5. С помощью визуального анализа поведения наблюдаемых процессов V и на основании правил R_{-1} оцениваем характеристики *динамичности, линейности, стационарности и времени жизни*.

6. Записываем рассчитанные характеристики в виде параметров априорной информационной модели $C_{-1} = \{d_\phi, l_\phi, s_\phi, t_\phi\}$.

7. На основании C_{-1} и M формируем план проверки эксперимента E . Экспертным путем выбираем количество и спецификацию испытаний $e_i \in E$, где $e_i = \{T, T_0, \Delta T_c, \Delta \tau_c, P_c\}$.

8. В соответствии с E измеряем V и формируем общую выборку X_{ii} .

9. Проводим постобработку X_{ii} в соответствии с правилами P_c (округление, преобразование и т. д.).

10. Для всех $\Delta T_c^{(i)}$ аналитическим и экспертным путем рассчитываем априорные и апостериорные характеристики с помощью правил R_{-1} и R_1 .

11. Записываем полученные характеристики в виде параметров апостериорной информационной модели $C_1 = \{d_\phi, l_\phi, s_\phi, t_\phi, l_f, c_f, m_f, u_f\}$.

6. Пример реализации метода оценивания информационных моделей наблюдаемых процессов в компьютерной сети

6. 1. Описание исходной информации о планируемых наблюдениях

Для примера использования разработанного метода проведем наблюдение нескольких процессов в компьютерной сети. Предположим, что задана спецификация мониторинга $M = \{G, F, S, D, O, I, V\}$, где G – оценка интенсивности использования рабочих станций сети, F – сбор статистических данных об интенсивности использования дисковой и сетевой подсистем, S – небольшая компьютерная сеть предприятия (до 400 узлов) с доступом в Интернет, рабочее время предприятия с 9 до 18 часов в будние дни, D – разработка программного обеспечения, O – умеренно нагруженные рабочие станции сети, I – программные средства мониторинга, основанные на WMI [15], V – параметры классов модели DMTF CIM [16] (табл. 1).

Таблица 1

Наблюдаемые переменные

CIM класс	Параметр класса
Win32_PerfFormattedData_PhysicalDisk_PerfDisk_	DiskTransfersPersec
Win32_PerfFormattedData_Tcpip_TCP	ConnectionsEstablished

6. 2. Этап предварительного оценивания (априорная модель)

Дадим краткую экспертную характеристику для процессов $v_i \in V$ (табл. 2).

Таблица 2

Экспертная характеристика наблюдаемых процессов

Процесс	Экспертная характеристика
DiskTransfersPersec	Переменная характеризует уровень использования (количество операций чтения и записи) физического носителя данных. Описывает высокодинамичный нелинейный нестационарный процесс
ConnectionsEstablished	Переменная характеризует количество TCP соединений, находящихся в статусе ESTABLISHED или CLOSE-WAIT. Описывает низкодинамичный кусочно линейный процесс

Чтобы собрать данные для предварительного анализа, произведем измерения значений наблюдаемых процессов с помощью системы мониторинга AggreGate, поддерживающей WMI. Выберем длительность наблюдения 5 минут с длиной интервала измерений 1 секунда.

Для каждого наблюдаемого процесса выберем длину интервала актуализации ΔT_c . Для высокодинамичного процесса DiskTransfersPersec выберем ΔT_c равным 5 значениям. Для низкодинамичного и линейного ConnectionsEstablished ΔT_c будет равняться 15 значениям.

Представим полученные значения наблюдаемых процессов в виде графиков, на которых отметим интервалы $\Delta T_c^{(i)}$ (рис. 3, а, б).



а



б

Рис. 3. Графики значений наблюдаемых процессов: а – DiskTransfersPersec; б – ConnectionsEstablished

Путем графо-визуального анализа поведения процессов на интервалах $\Delta T_c^{(i)}$, определим характеристики *динамичности, линейности, стационарности* и *времени жизни* наблюдаемых процессов с помощью правил R_{-1} , описанных в [14] и уточненных в п. 4.4 текущей работы.

Динамичность процесса. Для расчета этой характеристики используется коэффициент вариации, показывающий степень пульсации процесса. В ходе графо-визуального анализа определим, как часто и как сильно процесс отклоняется от своего среднего значения. В результате, DiskTransfersPersec считаем высокодинамичным процессом ($d_p = 1$), а ConnectionsEstablished низкодинамичным ($d_p = 0$).

Линейность процесса. Для расчета этой характеристики оценивается ошибка линейной регрессионной модели. В ходе визуального анализа определим, линейны ли графики наблюдаемых процессов на интервалах $\Delta T_c^{(i)}$. В результате, считаем, что DiskTransfersPersec – нелинейный процесс ($l_p = 0$), а ConnectionsEstablished – линейный ($l_p = 1$).

Стационарность процесса. Будем рассматривать стационарность процесса в широком смысле, т. е. среднее и дисперсия являются постоянными величинами, а корреляционная функция зависит только от разницы времени. В результате, считаем, что DiskTransfersPersec не стационарен ($s_p = 0$), а ConnectionsEstablished стационарен ($s_p = 1$).

Время жизни процесса. Для расчета этой характеристики необходимо оценить колебания значений на графиках. Если для 80 % интервалов $\Delta T_c^{(i)}$ колебания значений отсутствуют, то считаем процесс короткожи-

вущим. Будем считать оба процесса являются долговременными ($t_p = 1$).

Полученные характеристики запишем в виде модели C_{-1} (табл. 3).

Таблица 3

Параметры априорной модели наблюдаемых процессов

Процесс	$\{d_p, l_p, s_p, t_p\}$
DiskTransfersPersec	{1,0,0,1}
ConnectionsEstablished	{0,1,1,1}

Выберем план проведения эксперимента Е. Согласно спецификации мониторинга М известно, что основной период утилизации сети приходится на офисное рабочее время будних дней, характер поведения процессов может незначительно изменяться в течении дня, объекты мониторинга не являются высоконагруженными. На основании моделей C_{-1} можно утверждать, что DiskTransfersPersec является слабо предсказуемым процессом, следовательно интервал измерений и интервал актуализации должны быть короткими (1–3 секунды и 5–9 значений соответственно). Для более предсказуемого ConnectionsEstablished можно выбрать больший интервал измерений и интервал актуализации (3–6 секунды и 10–15 значений соответственно).

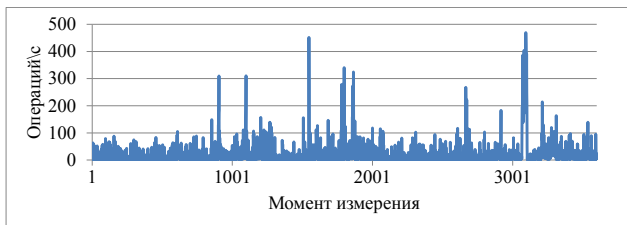
Исходя из полученной информации, выберем следующие параметры испытаний. Для каждого наблюдаемого процесса проведем 4 испытания, распределенных в течении рабочего дня. Для всех испытаний DiskTransfersPersec выберем следующую спецификацию: T_i – 15 мин., T_{01} – 11:00, T_{02} – 14:00, T_{03} – 15:00, T_{04} – 17:00, ΔT_c – 5 значений, Δt_c – 1 сек., R_c – округление до 10 знаков после запятой, сглаживание с помощью простого скользящего среднего с окном в 10 значений. Для всех испытаний ConnectionsEstablished выберем следующую спецификацию: T_i – 30 мин., T_{01} – 11:00, T_{02} – 14:00, T_{03} – 15:00, T_{04} – 17:00, ΔT_c – 15 значений, Δt_c – 3 сек., R_c – сглаживание с помощью простого скользящего среднего с окном в 5 значений.

6. 3. Этап экспериментального оценивания (апостериорная модель)

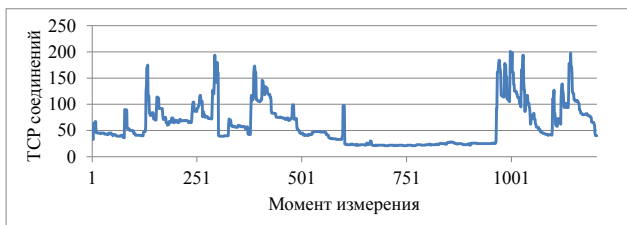
Проведем измерения значений наблюдаемых процессов в соответствии с выбранным планом эксперимента Е. Для каждого наблюдаемого процесса выборки всех испытаний объединяем в одну X_{ii} (рис. 4, а, б).

Проведем постобработку X_{ii} в соответствии с R_c : отформатируем значения и сгладим с помощью простого скользящего среднего. Результаты приведены на рис. 5, а, б.

Рассчитаем параметры апостериорной модели C_1 с помощью правил R_{-1} и R_1 . Для этого разобьем исходную выборку значений X_{ii} на интервалы в соответствии с ΔT_c и вычислим их распределение в соответствии со значениями априорных и апостериорных характеристик.

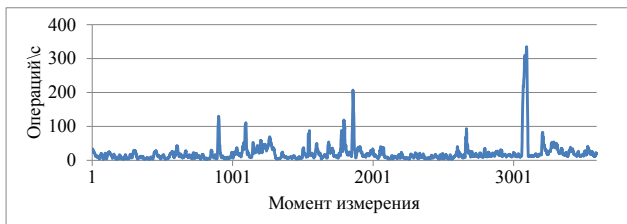


а

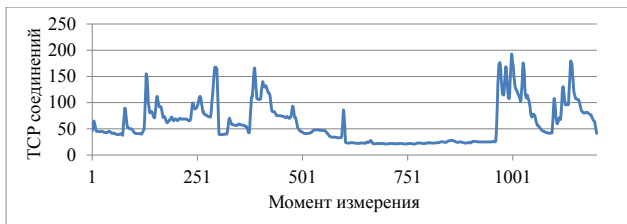


б

Рис. 4. Выборки значений наблюдаемых процессов: а – DiskTransfersPersec; б – ConnectionsEstablished



а



б

Рис. 5. Сглаженные значения наблюдаемых процессов: а – DiskTransfersPersec; б – ConnectionsEstablished

В табл. 4 приведены результаты распределения интервалов $\Delta T_c^{(i)}$ по значениям характеристик $d_\phi, l_\phi, t_\phi, l_f, m_f, u_f$.

Таблица 4

Распределение интервалов $\Delta T_c^{(i)}$ по значениям характеристик

Переменная	$d_\phi, \%$	$l_\phi, \%$	$t_\phi, \%$	$l_f, \%$	$m_f, \%$	$u_f, \%$
DiskTransfersPersec	0: 92,20 1: 7,80	0: 69,78 1: 30,22	0: 0 1: 100	0: 0 1: 0 2: 100	0: 73,82 1: 12,95 -1: 13,23	0: 61,98 1: 18,52 -1: 19,50
ConnectionsEstablished	0: 99,16 1: 0,84	0: 49,37 1: 50,63	0: 9,62 1: 90,38	0: 9,62 1: 14,23 2: 76,15	0: 35,57 1: 23,01 -1: 41,42	0: 30,54 1: 24,27 -1: 45,19

Проанализируем данные приведенные в табл. 4 и выберем окончательные значения рассмотренных характеристик для каждого наблюдаемого процесса.

Динамичность процесса. Сглаживание DiskTransfersPersec привело к значительному снижению его динамичности – только для 7,8 % интервалов $\Delta T_c^{(i)}$ процесс является высокодинамичным. Следовательно, в целом процесс является низкодинамичным ($d_\phi = 0$). Для ConnectionsEstablished подтверждается априорная оценка – для подавляющего количества интервалов актуализации данных $\Delta T_c^{(i)}$ процесс является низкодинамичным ($d_\phi = 0$).

Линейность процесса. Для 69,78 % интервалов $\Delta T_c^{(i)}$ DiskTransfersPersec характеризуется как нелинейный ($l_\phi = 0$). Для ConnectionsEstablished наблюдается приблизительно одинаковое количество интервалов $\Delta T_c^{(i)}$, на которых процесс линейный или нелинейный (курочно линейный процесс). В целом охарактеризуем его как линейный процесс ($l_\phi = 1$).

Время жизни процесса. Оба процесса являются долговременными ($t_\phi = 1$).

Степень нелинейности процесса. DiskTransfersPersec имеет третью степень нелинейности ($l_f = 2$). ConnectionsEstablished в целом также относится к третьей степени нелинейности ($l_f = 2$), однако для большого количества интервалов $\Delta T_c^{(i)}$ (23,85 %) процесс относится к первой или второй степени нелинейности.

Монотонность процесса. Для DiskTransfersPersec условие монотонности не выполняется для подавляющего количества интервалов $\Delta T_c^{(i)}$ ($m_f = 0$). ConnectionsEstablished убывает для 41,42 % интервалов $\Delta T_c^{(i)}$ ($m_f = -1$).

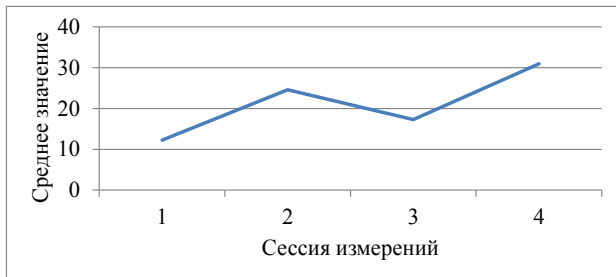
Выпуклость/вогнутость. Так как, DiskTransfersPersec не монотонный, то и условие выпуклости/вогнутости не выполняется ($u_f = 0$). ConnectionsEstablished является выпуклым для 45,19 % интервалов $\Delta T_c^{(i)}$ ($u_f = -1$).

Стационарность. Проверим соблюдение условий стационарности в широком смысле. Первое условие стационарности – постоянство математического ожидания. Рассчитаем средние значения наблюдаемых процессов для каждой из 4 проведенных сессий измерений (рис. 6).

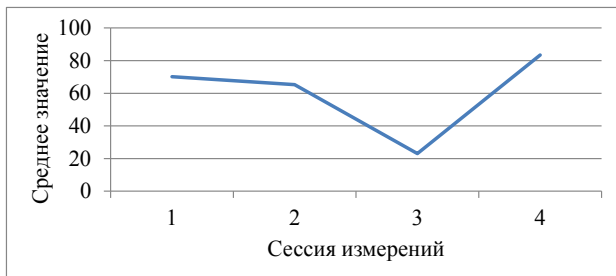
Условие постоянства математического ожидания не соблюдается для обоих наблюдаемых процессов. Следовательно, DiskTransfersPersec и ConnectionsEstablished не стационарны ($s_\phi = 0$).

Непрерывность. Эту характеристику считаем как экспертную оценку с помощью визуального анализа графиков функций наблюдаемых процессов. DiskTransfersPersec: высокая динамичность процесса и высокая степень нелинейности характеризует его функцию как непрерывную ($c_f = 1$). ConnectionsEstablished: на графиках процесса можно выделить уровни и перепады. Будем утверждать, что в точках перепадов функция процесса терпит разрыв первого рода ($c_f = 0$).

На основании полученных данных запишем параметры апостериорной модели S_t (табл. 5).



а



б

Рис. 6. Средние значения наблюдаемых процессов: а – DiskTransfersPersec; б – ConnectionsEstablished

Таблица 5

Апостериорные информационные модели наблюдаемых процессов

Процесс	$\{d_\phi, l_\phi, s_\phi, t_\phi, l_f, c_f, m_f, u_f\}$
DiskTransfersPersec	{0,0,0,1,2,1,0,0}
ConnectionsEstablished	{0,1,0,1,2,0,-1,-1}

На основании вычисленных апостериорных моделей видно, что значения некоторых характеристик наблюдаемых процессов изменились по сравнению с предварительным оцениванием. Это связано с тем, что на этапе экспериментального оценивания можно бо-

лее точно вычислить эти характеристики. Но следует заметить, что разница в значениях может быть также обусловлена трансформацией экспериментальных данных.

После вычисления апостериорных моделей следует выбрать аппроксимирующие функции наблюдаемых процессов и приступить к этапу проведения измерений. Эти вопросы в данной работе не рассматриваются.

7. Выводы

В результате решения поставленной задачи разработан метод оценивания информационных моделей наблюдаемых процессов в компьютерной сети. Данный метод позволяет оценивать информационные модели, предложенные в статье [14]. Шаги метода сформированы в строгом соответствии с этапами реализации мониторинга наблюдаемых процессов. На каждом этапе осуществляется расчет математических и статистических характеристик наблюдаемых процессов с помощью правил, описанных в [14]. Эффективность метода обуславливается его простотой и не высокими вычислительными затратами. Адекватность метода подтверждается примером его использования.

Дальнейшая работа может вестись над разработкой метода мониторинга процессов в компьютерной сети, который будет основываться на разработанных моделях и технологиях WBEM-CIM.

Научная новизна состоит в том, что предложен метод оценивания информационных моделей наблюдаемых процессов в компьютерной сети. Метод отличается тем, что позволяет в строгой соответствии с этапами реализации мониторинга сети, рассчитать заранее выбранные статистические и математические характеристики наблюдаемых процессов и оценить предложенные в [14] информационные модели.

Практическая ценность метода заключается в повышении эффективности решения задачи аппроксимации наблюдаемых процессов в компьютерной сети для снижения уровня служебного трафика систем мониторинга.

Литература

- Hoeksema, F. W. A methodical approach to performance measurement experiments: measure and measurement specification [Text] / F. W. Hoeksema, J. T. van der Veen, B. J. van Beijnum. – University of Twente, Centre for Telematics and Information Technology (CTIT), 1997. – 113 p.
- Sidibé, M. QoS monitoring for end-to-end heterogeneous networks configurations management [Text] / M. Sidibé, A. Mehaoua // IEEE International Conference on Automation, Quality and Testing, Robotics AQTR. – 2008. – Vol. 1. – P. 364–368. doi: 10.1109/aqtr.2008.4588769
- Lahmadi, A. Efficient distributed monitoring in 6LoWPAN networks [Text] / A. Lahmadi, A. Boeglin, O. Festor // In proceedings of the 9th international conference on network and service management, 2013. – P. 268–276.
- Yu, C. FlowSense: Monitoring Network Utilization with Zero Measurement Cost [Text] / C. Yu, C. Lumezanu, Y. Zhang, V. Singh, G. Jiang, H. V. Madhyasth // In proceedings of the 14th international conference on Passive and Active Measurement, 2012. – P. 31–41.
- Ballard, J. R. Extensible and Scalable Network Monitoring Using OpenSAFE [Text] / J. R. Ballard, I. Rae, A. Akella // In proceedings of the 2010 internet network management conference on Research on enterprise networking, 2010. – P. 1–6.
- Narayanan, H. T. S. Feasibility of SNMP OID compression [Text] / H. T. S. Narayanan, I. Geetha, N. Sumitra // Journal of King Saud University – Computer and Information Sciences. – 2013. – Vol. 25, Issue 1. – P. 35–42. doi: 10.1016/j.jksuci.2012.05.006
- Саенко, В. И. Метод выбора моментов измерений для процессов непрерывного мониторинга [Текст] / В. И. Саенко, А. И. Гриценко // Радиоэлектроника и информатика. – 2007. – № 4. – С. 119–122.

8. Hernandez, E. Adaptive Sampling for Network Management [Text] / E. Hernandez, M. Chidester, A. George // Journal of Network and Systems Management. – 2001. – Vol. 9, Issue 4. – P. 409–434.
9. Andrey, L. Survey of SNMP performance analysis studies [Text] / L. Andrey, O. Festor, A. Lahmadi, A. Pras, J. Schönwälder // International journal of network management. – 2009. – Vol. 19, Issue 6. – P. 527–548. doi: 10.1002/nem.729
10. Pras, A. Comparing the Performance of SNMP and Web Services-Based Management [Text] / A. Pras, T. Drevers, R. van de Meent, D. Quartel // IEEE: Transactions on network and service management. – 2004. – Vol. 1, Issue 2. – P. 72–82. doi: 10.1109/tnsn.2004.4798292
11. Айвазян, С. А. Прикладная статистика и основы эконометрики [Текст] / С. А. Айвазян, В. С. Мхитарян. – Юнити. – 1998. – 1005 с.
12. Стрижов, В. В. Методы выбора регрессионных моделей [Текст] / В. В. Стрижов, Е. А. Крымова. – Вычислительный центр РАН, 2010. – 60 с.
13. Карабутов, Н. Н. Выбор структуры модели при обработке результатов измерений в системах управления [Текст] / Н. Н. Карабутов // Измерительная техника. – 2008. – № 9. – С. 29–33.
14. Саенко, В. И. Информационные модели наблюдаемых процессов для мониторинга компьютерных сетей [Текст] / В. И. Саенко, А. И. Гриценко // Вестник национального технического университета «ХПИ». Серия: «Новые решения в современных технологиях». – 2014. – № 48. – С. 55–66.
15. Lavy, M. Windows Management Instrumentation (WMI) [Text] / M. Lavy, A. Meggitt. – New Riders Publishing, 2001. – 432 p.
16. DMTF Std. DSP0004, Specification Version 2.6.0. Common Information Model (CIM) Infrastructure [Text] / March 2010. – 186 p.

У даній роботі розглянуто процес побудови мультибазових сховищ даних на основі властивостей даних, для яких воно будується, і запитів, які виконуються до цих даних. Проаналізовано показники швидкодії побудованого сховища, зокрема на різних етапах побудови. Досліджено вплив параметрів двофазного алгоритму, зокрема порогу адаптації процесу пошуку на основі генів
Ключові слова: мультибазові сховища даних, структурованість даних, модифікований генетичний алгоритм з адаптацією

В данной работе рассмотрен процесс построения мультибазовых хранилищ данных на основе свойств данных, для которых оно строится, и запросов, которые выполняются с этими данными. Проанализированы показатели быстродействия построенного хранилища, в том числе на различных этапах построения. Исследовано влияние параметров двухфазного алгоритма, в частности порога адаптации процесса поиска на основе генов

Ключевые слова: мультибазовые хранилища данных, структурированность данных, модифицированный генетический алгоритмы с адаптацией

УДК 004.652
 DOI: 10.15587/1729-4061.2015.36646

ПОБУДОВА МУЛЬТИБАЗОВИХ СХОВИЩ ДАНИХ НА ОСНОВІ СТРУКТУРОВАНОСТІ ДАНИХ ТА ЗАПИТІВ

А. Ю. Яцишин
 Провідний прикладний програміст
 ДП «Головфінтех»
 вул. Дегтярівська, 38-44, м. Київ,
 Україна, 04119
 E-mail: andrew.yatsyshyn@hotmail.com

1. Вступ

На сьогодні існують різні роботи та літературні джерела, що описують сховища даних, їх проектування та інтеграцію даних. Зокрема, запропоновані підходи є у авторів У. Інмона [1], Р. Кімболла [2] та Д. Хекні [3]. Для сховищ даних формалізовано низку моделей що визначають сховище даних і дозволяє описувати процес їх побудови. Ці моделі будуються на основі схеми «сутність-зв'язок» або похідної від неї схеми «зірка». Також представлена низка робіт, що описують різні методи оптиміза-

ції баз даних – побудову індексів, вибір матеріалізованих представлень, горизонтальну та вертикальну фрагментацію.

Автором у роботі [4] було введено мультибазові сховища даних, які включають не тільки реляційну і багатовимірну базу даних, а також бази даних з відсутністю жорсткої схеми – XML та NoSQL, і файлове сховище.

Такі сховища даних проектуються для оптимального зберігання даних різного рівня структурованості, спрямованого на підвищення швидкодії виконання запитів.