

# Стохастическая модель процесса идентификации сервисов информационной системы

*Г.Н. Циперман <g.tsiperman@voskhod.ru>  
ФГУП НИИ «Восход», 119607, Россия, г. Москва,  
ул. Удальцова, дом 85.*

**Аннотация.** В статье рассматривается задача оценки объема трудозатрат на проектирование функциональных требований к информационной системе в сервис-ориентированной архитектуре. Для этого предлагается стохастическая модель процесса идентификации сервисов информационной системы, позволяющая при минимальных исходных данных дать оценку ожидаемого количества объектов проекта и связей между ними. Модель основана на представлении процесса декомпозиции автоматизируемого бизнес-процесса как ветвящегося случайного надкритического процесса Гальтона-Ватсона. Проектирование связей между элементами декомпозиции моделируется как процесс построения связного случайного графа в модели Эрдеша-Реньи. Предсказания модели подтверждаются проверкой на экспериментальных данных.

**Ключевые слова:** сервис, идентификация, стохастическая модель, ветвящийся процесс, случайный граф, оценка времени проектирования, трудозатраты, сервис-ориентированная архитектура.

## 1. Введение

Процесс идентификации сервисов информационной системы (ИС) в парадигме сервис-ориентированной архитектуры (СОА) представляет собой суть проектирования ИС, включающего стадии разработки концепции, технического и рабочего проектирования.

На раннем этапе проектирования трудно предсказать сложность проекта, характеризуемую количеством и структурой связей между фактами предметной области, которые порождают соответствующие требования к сервисам ИС. Однако, для практики проектирования ИС этот вопрос имеет особую актуальность, вызванную необходимостью обеспечения объективной стоимости проекта.

Поэтапная оценка такой сложности не актуальна, т.к. в сложившейся практике создания ИС отдельное финансирование стадий проекта, когда стоимость каждой последующей стадии оценивается на основе результатов предыдущей, является исключительно редким явлением. Заказчик хочет получить оценку стоимости проекта до того, как проект начнется.

Данная работа посвящена изучению общих свойств процесса идентификации сервисов, обеспечивающих прогнозирование трудоемкости стадий формирования функциональных требований к создаваемой ИС.

## 2. Модель процесса идентификации сервисов

Основным методом идентификации сервисов является декомпозиция автоматизируемого бизнес-процесса, в результате которой сначала определяется дерево бизнес-функций [1]. Каждая бизнес-функция представляет собой описание некоторого промежуточного результата, получаемого при выполнении основного бизнес-процесса.

Например, бизнес-процесс проектирования технического сложного объекта (ТСО) на первом уровне декомпозируется на следующие бизнес-функции:

- Инициация проектирования. Результатом является организация проекта по проектированию ТСО.
- Анализ структуры проектируемого ТСО, в результате которого формируются функциональная и логическая модель изделия.
- Разработка конструкции, предполагающая получение конструктивной модели ТСО и определение его комплектующих.
- Многоаспектный анализ предлагаемой конструкции ТСО, определяющий на основе технических и экономических расчетов соответствие характеристик проекта требованиям технического задания.
- Завершение проектирования, результатом которого является проект ТСО.

В свою очередь, каждая бизнес-функция может быть декомпозирована далее, исходя из более детальных промежуточных результатов ее выполнения.

Понятно, что для составления полной картины бизнес-процесса на каждом уровне декомпозиции должна быть определена последовательность выполнения бизнес-функций, т.е. объекты декомпозиции должны быть связаны управляющими потоками. Такая модель называется сценарием бизнес-функций и представляет собой связный граф, узлами которого являются объекты декомпозиции, а ребрами – потоки управления.

Конечный уровень декомпозиции бизнес-процесса соответствует описанию необходимых действий (операций), которые должен выполнить субъект

бизнес-процесса, чтобы получить требуемый промежуточный результат. Объекты этого уровня называются бизнес-операциями.

Таким образом, декомпозиция бизнес-функции бизнес-процесса включает в себя как бизнес-функции следующего уровня, так и бизнес-операции и описывается соответствующим сценарием.

Собственно, идентификация сервисов ИС начинается с определения сервисов операций. Под сервисом операции мы будем понимать спецификацию функций информационной системы, поддерживающих исполнение соответствующей бизнес-операции. Между бизнес-операцией и сервисом операции устанавливается однозначное соответствие. В связи с этим мы можем рассматривать их в связке как единый конструкт. Визуальное представление процесса декомпозиции бизнес-процесса приведено на рис. 1.

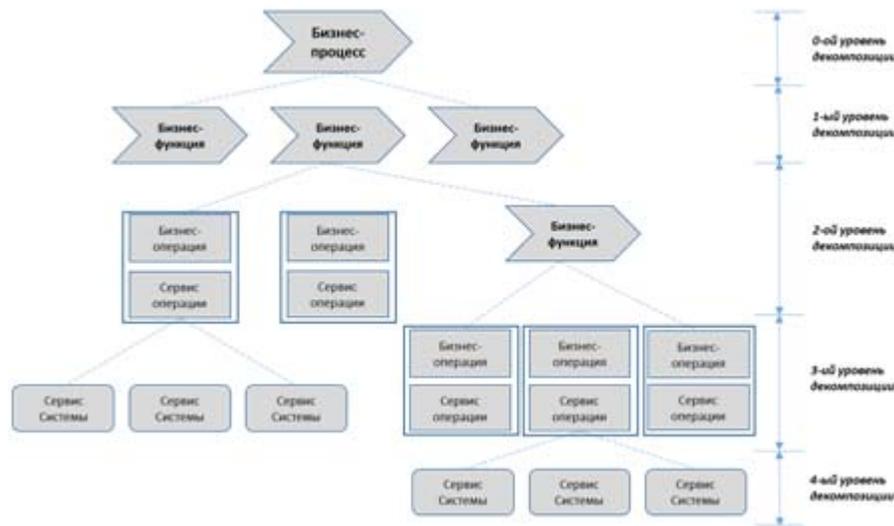


Рис. 1. Дерево декомпозиции бизнес-процесса.

Дадим формальное определение процессу декомпозиции бизнес-процесса.

Пусть  $B_{fnc} = \{f_i\}$  и  $B_{opr} = \{o_j\}$ , где  $i, j \in \mathbb{N}$  – множества бизнес-функций и бизнес-операций бизнес-процесса (вместе с сервисами операций). Определим отношение декомпозиции:

$$D^b: B_{fnc} \rightarrow B_{fnc} \cup B_{opr} \quad (1)$$

$$D_i^b = D^b(f_i), f_i \notin D_i^b \quad (2)$$

При этом

$$\left\{ \begin{array}{l} \forall \left( \begin{array}{l} D_i^b = (f_{i1}, \dots, f_{ik}, o_{i1}, \dots, o_{ip}), \\ D_j^b = (f_{j1}, \dots, f_{jl}, o_{j1}, \dots, o_{js}) \end{array} \right) \in D^b, \forall i, j, p, s, n \in \mathbb{N}: \\ D_i^b \cap D_j^b = \emptyset \\ D_1^b \cup D_2^b \cup \dots \cup D_n^b = (B_{fnc} \setminus f_0) \cup B_{opr} \end{array} \right. \quad (3)$$

В (3)  $f_0$  – исходный бизнес-процесс. Таким образом, каждая бизнес-функция и бизнес-операция присутствуют только в одной декомпозиции.

Определяем условие завершения декомпозиции, отражающее тот факт, что декомпозиция имеет смысл, если в результате порождается не менее одного элемента:

$$\forall i \in \mathbb{N}: |D_i^b| \geq 1 \quad (4)$$

Дадим формальное определение сценарию бизнес-функции. Рассмотрим множество пар элементов

$$S^b = \{(s_i, s_j); \forall i, j \in \mathbb{N}, i \neq j; s_{i,j} \in (B_{fnc} \setminus f_0) \cup B_{opr}\} \quad (5)$$

Сценарием бизнес-функции  $f_k$  с декомпозицией  $D_k^b$  называется подмножество

$$S_k^b \subset S^b$$

$$S_k^b = \{(s_i, s_j); D_k^b \subset \{s_{i,j}\}; \forall i, j \in \mathbb{N}, i \neq j, s_{i,j} \in (B_{fnc} \setminus f_0) \cup B_{opr}\} \quad (6)$$

Сценарий бизнес-функции в отличие от ее декомпозиции может включать бизнес-функции и бизнес-операции из других декомпозиций. Этот факт отражает повторное использование конструктов при проектировании информационной системы. Как видно из (6), в случае, если бизнес-функция не имеет собственной декомпозиции она может иметь сценарий, в который входят элементы иных декомпозиций.

Пара элементов сценария соответствует отношению передачи управления между элементами и, вообще говоря, является ориентированной. Однако, нас интересует оценка количества таких пар и вопросы ориентации этой связи мы рассматривать не будем.

Идентификация сервисов ИС представляет собой декомпозицию сервисов операций: сервисы системы обеспечивают выполнение функциональности, специфицированной сервисами операций. Аналогично последовательность вызова сервисов системы моделируется управляющими потоками, связывающими сервисы в соответствующий сценарий сервиса операции. Формально это определяется следующим образом.

Пусть  $C_{opr} = \{c_i\}$  множество сервисов операций. Связь сервисов операций модели с бизнес-операциями обеспечивается однозначным отношением зависимости

$$L: O_i \rightarrow c_i, O_i \in B_{opr}, c_i \in C_{opr} \quad (7)$$

Как отмечалось выше, пару  $(o_i, c_i)$  в силу характера однозначного отношения между ними мы будем рассматривать как единый объект.

Пусть  $C_{sys} = \{r_i\}$  множество системных сервисов модели. Аналогично (1) - (3) определяем отношение декомпозиции для сервисов:

$$D^s: C_{opr} \rightarrow C_{sys} \quad (8)$$

$$D_i^s = D^s(r_i) \quad (9)$$

$$\left\{ \begin{array}{l} \forall (D_i^s = (r_{i1}, \dots, r_{ik}), D_j^s = (r_{j1}, \dots, r_{jl})) \in D^s, \forall i, j, k, l, n \in \mathbb{N}: \\ D_i^s \cap D_j^s = \emptyset \\ D_1^s \cup D_2^s \cup \dots \cup D_n^s = C_{sys} \\ |D_i^s| \geq 1 \end{array} \right\} \quad (10)$$

Сервис операции, не имеющий самостоятельной декомпозиции, тем не менее, может иметь сценарий, включающий элементы декомпозиции других сервисов операций. Сценарий сервиса операции определяется аналогично сценарию бизнес-функции (5) – (6):

Рассматриваем множество пар элементов

$$S^s = \{(r_i, r_j); i \neq j; r_{i,j} \in C_{sys}\} \quad (11)$$

Сценарием сервиса операции  $c_k$  с декомпозицией  $D_k^s$  называется подмножество

$$\begin{aligned} S_k^s &\subset S^s \\ S_k^s &= \{(r_i, r_j); D_k^s \subset \{r_{i,j}\}; i \neq j, r_{i,j} \in C_{sys}\} \end{aligned} \quad (12)$$

## 2.1. Вероятностная модель процесса декомпозиции

Процесс декомпозиции модели информационной системы мы будем рассматривать как случайный ветвящийся процесс Гальтона-Ватсона[2]. Обобщим оператор декомпозиции:

$$D(x_i) = \begin{cases} D^b(x_i), x_i = f_i \in B_{fnc} \\ D^s(x_i), x_i = r_i \in C_{sys} \end{cases} \quad (13)$$

Множество элементов декомпозиции будем обозначать как

$$D_i = D(x_i) = (x_{i1}, x_{i2}, \dots, x_{in}) \quad (14)$$

В таких обозначениях ветвящийся процесс может быть описан следующим образом:

$$D_0 = D(0) = (f_0)$$

$$D_1 = D(f_0) = (x_1, x_2, \dots, x_n)$$

$$D_{21} = D(x_1) = (x_{11}, \dots, x_{1m}); \dots; D_{2n} = D(x_n) = (x_{n1}, \dots, x_{nl}) \quad (15)$$

$$D_{\frac{ki \dots j}{k}} = D\left(x_{\frac{i \dots j}{k-1}}\right) = \left(x_{\frac{i \dots j1}{k}}, \dots, x_{\frac{i \dots jp}{k}}\right)$$

Пусть  $\xi_{ki \dots j}$  – количество элементов декомпозиции  $D_{ki \dots j}$ . Элементарной реализацией процесса декомпозиции назовем набор

$$\omega = (\xi_0, \xi_{1i}, \xi_{2ip}, \dots, \xi_{kip \dots j}) \quad (16)$$

Вероятностное пространство  $(\Omega, \mathcal{F}, P)$  процесса декомпозиции определяется на пространстве событий, представляющем собой множество всех элементарных реализаций

$$\Omega = \{\omega\} \quad (17)$$

При этом,  $\sigma$ -алгебра  $\mathcal{F}$  порождается подмножествами пространства  $\Omega$ , а вероятностная мера  $P$  задается соотношениями:

$$\begin{aligned} P\{\xi_{kip \dots j} = k\} &= p_k \\ P\{(\xi_0, \xi_{1i}, \xi_{2ip}, \dots, \xi_{kip \dots j}) = (k_0, k_{1i}, k_{2ip}, \dots, k_{kip \dots j})\} &= p_{k_0} p_{k_{1i}} p_{k_{2ip}} \dots p_{k_{kip \dots j}} \end{aligned} \quad (18)$$

Процесс декомпозиции, как и соответствующий ему процесс Гальтона-Ватсона, можно представить в виде плоского дерева, корнем которого является бизнес-процесс  $f_0$ . Будем называть  $k$ -ым поколением процесса декомпозиции или уровнем дерева подмножество декомпозиций процесса

$$I_k = \{D_{ki \dots j}\} \quad (19)$$

Количество элементов  $k$ -ого поколения будем обозначать

$$Z(k) = |\{D_{ki \dots j}\}| = \sum_{i \dots j} \xi_{ki \dots j} \quad (20)$$

Сформулируем основную задачу исследования этапа определения структурных элементов информационной системы. Для процесса декомпозиции, представляющего основной механизм определения структуры,  $Z(0) = 1$  и, положим  $Z(1) = N_0$ , где  $N_0$  – неслучайная величина, соответствующая количеству бизнес-функций первого уровня декомпозиции бизнес-процесса, специфицированных при предварительном изучении темы проекта создания ИС.

Будем рассматривать процесс декомпозиции  $\tilde{D}_0$  как  $N_0$  подпроцессов  $\tilde{D}_{1i}$ , представляющих собой надкритические процессы Гальтона-Ватсона с ненулевой вероятностью вырождения  $\alpha$  (соответственно, вероятность невырождения процесса  $\gamma = 1 - \alpha$ ). Надкритичность процесса означает, что математическое ожидание числа элементов декомпозиции

$$E\{\xi\} > 1 \quad (21)$$

Требуется найти наиболее вероятное число элементов декомпозиции (ожидаемое число структурных компонент модели ИС) при указанных условиях.

## 2.2. Вероятностная модель построения сценария

Положим мы определили структуру модели информационной системы, спроектировав множество ее компонент, представляющее собой множество декомпозиций

$$D = \{D_i\}, |D| = n \quad (22)$$

На следующем этапе проектирования мы должны построить сценарии бизнес-функций и сервисов операций, определяющих использование ИС в автоматизируемом бизнес-процессе. Мы будем рассматривать структуру системы перед началом построения сценариев как несвязный граф  $G(n)$ , состоящий из  $n$  изолированных вершин. При таком подходе, исходя из смысла проектирования информационной системы, результатом проектирования сценариев должен явиться сценарий использования системы, представляющий собой связный граф  $G(n, M)$ , где  $M$  – количество ребер этого графа. Формально, используя определения (6) и (12):

$$M = |S| = \sum_k |S_k|$$

$$S = \{S_k\} = \begin{cases} S_k^b = \{(s_i, s_j); D_k^b \subset \{s_{i,j}\}; \forall i, j \in \mathbb{N}, i \neq j, s_{i,j} \in (B_{fnc} \setminus f_0) \cup B_{opr}\} \\ S_k^s = \{(r_i, r_j); D_k^s \subset \{r_{i,j}\}; i \neq j, r_{i,j} \in C_{sys}\} \end{cases} \quad (23)$$

Задачей исследования процесса построения сценария использования проектируемой информационной системы является определение необходимого количества времени для получения сценария  $S$  или, иными словами, формирования связного графа  $G(n, M)$ .

Будем рассматривать граф  $G(n, M)$  как случайный граф в модели Эрдеша-Реньи. Формальное описание этой модели приведем, следуя изложению из монографии [3].

Определим вероятностное пространство всех графов  $\mathfrak{L}(n, M)$  порядка  $n$  и размера  $M$ , каждый элемент которого равновероятен. Очевидно

$$0 \leq M \leq N = \binom{n}{2}, |\mathfrak{L}(n, M)| = \binom{N}{M} \quad (24)$$

Тогда вероятность формирования любого графа  $G(n, M)$  из этого пространства

$$P\{G = G(n, M)\} = \binom{N}{M}^{-1} \quad (25)$$

В общем случае  $M$  функционально зависит от  $n$ :  $M = M(n)$ .

Введем понятие процесса на случайном графе.

Пусть  $V = \{1, 2, \dots, n\}$ . Процессом  $\tilde{G} = (G_t)_0^N$ , где  $N = \binom{n}{2}$  называется марковский процесс на  $V$  такой, что:

- 1)  $(G_0)_0^N = 0$
- 2) Каждый  $G_t$  представляет собой граф на  $V$
- 3)  $G_t$  имеет  $t$  ребер, где  $t = 1, 2, \dots, N$
- 4)  $G_0 \subset G_1 \subset \dots$

Пусть  $\tilde{\mathfrak{L}} = \{\tilde{G}\}$ ,  $|\tilde{\mathfrak{L}}| = N!$  вероятностное пространство, элементами которого являются равновероятные процессы  $\tilde{G}$ . Отображение  $\tilde{\mathfrak{L}} \rightarrow \mathfrak{L}(n, M)$  определяется как

$$\tilde{G} = (G_t)_0^N \rightarrow G(n, M) \quad (26)$$

Иначе: состояние процесса  $G_t$  в момент времени  $t = M$  отождествляется с элементом пространства  $\mathfrak{L}(n, M)$ .

Мы рассмотрим процесс формирования сценария использования информационной системы как процесс на случайном графе. Момент завершения формирования сценария соответствует моменту  $T_S$ , когда граф  $G(n, M)$  становится связным. Под связностью мы будем понимать вершинную односвязность ( $k(G) = 1$ ). На практике это означает, что полностью определена связь бизнес-функций и входящих в них бизнес-операций и все состояния проектируемой информационной системы связаны между собой хотя бы одним маршрутом.

## 3. Метрики процесса идентификации сервисов

В качестве основной метрики процесса идентификации сервисов ИС, отражающей его стохастические закономерности, мы рассмотрим общее

количество элементов функциональной модели бизнес-процесса: бизнес-функций, бизнес-операций и системных сервисов

$$T_D = |D| = \sum_{i=0}^k Z(i) \quad (27)$$

Определение процесса декомпозиции (15) служит основой для рассмотрения всех элементов в единой вероятностной модели, несмотря на различие их природы. Другие метрики процесса (общее время проектирования и его эффективность) определяются количеством элементов модели.

Модель процесса проектирования в части определения структуры функциональной модели автоматизируемого бизнес-процесса определяется следующими исходными предположениями:

- 1) В основе определения структурных элементов функциональной модели лежит процесс декомпозиции автоматизируемого бизнес-процесса.
- 2) Процесс декомпозиции рассматривается как ветвящийся процесс Гальтона-Ватсона с граничными условиями  $Z(0) = 1$ ;  $Z(1) = N_0$ , где  $N_0$  – неслучайная величина.
- 3) Рассматриваемый ветвящийся процесс является надкритическим с ненулевой вероятностью вырождения. Процесс вырождается в случае, если каждая декомпозиция  $k$ -ого поколения включает 0 или 1 элемент.
- 4) Число элементов декомпозиций, относящихся к  $k$ -ому поколению имеет распределение Пуассона

$$P\{\xi_k = x\} = \frac{\lambda^x}{x!} e^{-\lambda} \quad (28)$$

- 5) Математическое ожидание распределения элементов декомпозиций  $\lambda = E[\xi]$  не зависит от поколения процесса.

Другой метрикой, зависящей от общего количества элементов функциональной модели, является количество связей между элементами. Это количество мы будем определять, как минимально необходимое для связности графа функциональной модели бизнес-процесса.

### 3.1. Ожидаемое количество элементов функциональной модели

Для количества элементов  $k$ -ого поколения ветвящегося процесса математическое ожидание определяется как [4]

$$Z(i) = \lambda^i, i = (0, 1, 2, \dots) \quad (29)$$

Соответственно оценка для общего количества элементов модели, учитывая независимость  $\lambda$  от номера поколения, может быть определено по формуле суммы геометрической прогрессии

$$n = \sum_{i=0}^k Z(i) = \frac{\lambda^{k+1} - 1}{\lambda - 1} \quad (30)$$

Учитывая граничные условия рассматриваемого процесса декомпозиции, мы будем рассматривать общий процесс как  $N_0$  независимых процессов, начинающихся со второго уровня модели (считаем с нулевого уровня). Тогда выражение (30) примет вид

$$n = 1 + N_0 \left( \frac{\lambda^k - 1}{\lambda - 1} \right) \quad (31)$$

Для того, чтобы сделать оценку (31) применимой необходимо оценить математическое ожидание элементов единичной декомпозиции и ожидаемое число уровней модели.

Обсудим вероятности вырождения ветвящегося процесса  $\alpha$  и, соответственно, невырождения  $\gamma = 1 - \alpha$ . Интуитивно понятно, что чем выше вероятность невырождения процесса, тем больше структурных элементов декомпозиции удастся получить. В практике проектирования информационных систем это соответствует глубине понимания проектировщиком автоматизируемого бизнес-процесса: чем больше деталей предметной области известно, тем больше фактов (элементов) может быть отражено в функциональной модели бизнес-процесса. В связи с этим мы будем называть вероятность  $\gamma$  детальностью модели бизнес-процесса, обратную вероятность  $\alpha$  – ее неопределенностью.

Детальность проектируемой модели может служить входным параметром для оценки основных метрик процесса проектирования. Чем выше детальность, тем подробнее должна быть модель и, соответственно, больше трудозатрат такая работа потребует.

Известным фактом теории ветвящихся процессов является зависимость вероятности вырождения надкритического ветвящегося процесса от математического ожидания  $\lambda$  [4]. Неопределенность модели является наименьшим неотрицательным корнем уравнения

$$f(\alpha) = \alpha \quad (32)$$

Здесь  $f(\alpha)$  вероятностная производящая функция, задающая распределение случайной величины  $\xi$ . Для пуассоновского распределения [5]

$$f(\alpha) = \sum_{i=0}^{\infty} P\{\xi = i\} \alpha^i = \sum_{i=0}^{\infty} \frac{(\lambda\alpha)^i}{i!} e^{-\lambda} = e^{\lambda(\alpha-1)} = e^{-\lambda\gamma} \quad (33)$$

Подставив (33) в (32), легко видеть

$$\lambda = -\frac{\ln(1-\gamma)}{\gamma} = -\frac{\ln(\alpha)}{1-\alpha} \quad (34)$$

Уравнение (34) определяет зависимость ожидаемого количества элементов единичной декомпозиции от детальности (неопределенности) функциональной модели бизнес-процесса.

Для оценки ожидаемого количества уровней декомпозиции воспользуемся следующими соображениями. Пусть вырождение процесса декомпозиции начинается с уровня  $K$ . Мы рассмотрим каждый элемент этого уровня  $n_{Ki}$  как начальный для отдельного процесса  $\tilde{D}_{Ki}$ . В этом случае для каждого такого процесса декомпозиции общее количество элементов определяется выражением [4]

$$n_{Ki} = \frac{1}{1-f'(\alpha)} \quad (35)$$

Учитывая (32) и (33) выражение (35) принимает вид

$$n_{Ki} = \frac{1}{1-\lambda\alpha} \quad (36)$$

Как видно из (36) число элементов, возникающих при вырождении любого из процессов  $\tilde{D}_{Ki}$ , определяется выражением

$$\Delta n_{deg} = \frac{1}{1-\lambda\alpha} - 1 = \frac{\lambda\alpha}{1-\lambda\alpha} \quad (37)$$

Диапазон, в котором лежит  $\Delta n_{Ki}$  определяется неравенством<sup>1</sup>

$$\underline{\Delta n_{deg}} = \left\lfloor \frac{\lambda\alpha}{1-\lambda\alpha} \right\rfloor \leq \Delta n_{deg} \leq \left\lceil \frac{\lambda\alpha}{1-\lambda\alpha} \right\rceil = \overline{\Delta n_{deg}} \quad (38)$$

Общее число процессов  $\tilde{D}_{Ki}$  равно  $\lambda^{K-1}$  – количеству элементов на  $K$ -ом уровне декомпозиции бизнес-процесса, относящихся к подпроцессу  $\tilde{D}_{1i}$ . Тогда

<sup>1</sup>Мы будем обозначать символами  $\lfloor a \rfloor$  и  $\lceil a \rceil$  значения округления числа вниз и вверх соответственно.

количество элементов, входящих в следующие за  $K$ -ым уровни (уровни вырождения), будет определяться как  $\Delta N_{deg} = \lambda^{K-1} \Delta n_{deg}$  и лежит в диапазоне

$$\underline{\Delta N_{deg}} = \lambda^{K-1} \underline{\Delta n_{deg}} \leq \Delta N_{deg} \leq \overline{\Delta N_{deg}} = \lambda^{K-1} \overline{\Delta n_{deg}} \quad (39)$$

Значение количества элементов декомпозиции на уровне вырождения процесса  $\tilde{D}_{1i}$ , определяемое выражением (39), мы будем относить к последнему  $K+1$ -ому поколению декомпозиции этого процесса. Для того, чтобы оценить значение  $K$ , положим, что уровень  $K$  является последним, т.е. на уровне вырождения нет элементов. Это означает, что

$$\Delta N_{Ki} = \lambda^{K-1} \frac{\lambda\alpha}{1-\lambda\alpha} < 1 \quad (40)$$

Разрешая неравенство (40) относительно  $K$ , получаем оценку

$$K = \left\lceil \frac{\ln(1-\lambda\alpha) - \ln(\lambda\alpha)}{\ln(\lambda)} \right\rceil + 1 \quad (41)$$

На рис. 2 приведен график зависимости количества уровней декомпозиции  $K$  функциональной модели бизнес-процесса от детальности  $\lambda$ .

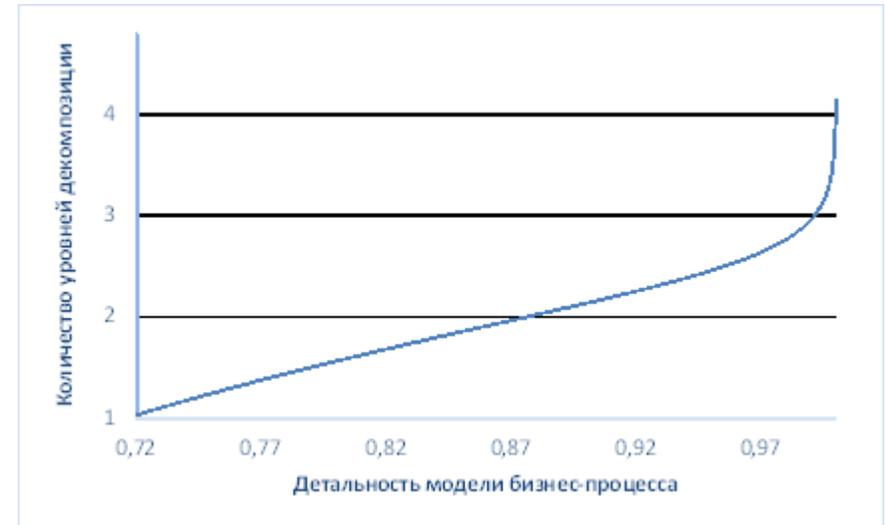


Рис. 2. График зависимости количества уровней декомпозиции функциональной модели бизнес-процесса от детальности.

Определение уровня начала вырождения декомпозиции основывалось на предположении, что следующий уровень модели, соответствующий вырождению, не содержит элементов декомпозиции. На самом деле, это не так: мы должны учесть элементы на уровне вырождения (на практике их вполне заметное количество). Однако, полученные значения уровня начала декомпозиции в предположении, с ростом числа элементов уровня вырождения меняются медленно. Можно показать, что  $K$  растет медленнее логарифма числа элементов на уровне вырождения декомпозиции. В связи с этим оценка, определяемая уравнением (41), является вполне приемлемой и, что важно, согласуется с реальной практикой проектирования, где в основном декомпозиция заканчивается на третьем уровне, а уровень 5 встречается достаточно редко.

Теперь мы имеем все необходимые данные для оценки общего числа элементов декомпозиции исходного бизнес-процесса:

$$T_D = n + N_0 \Delta N_{deg} = 1 + N_0 \left( \frac{\lambda^K - 1}{\lambda - 1} + \Delta N_{deg} \right) \quad (42)$$

Учитывая неравенство (39) ожидаемое количество элементов декомпозиции определяется диапазоном

$$\begin{aligned} \underline{T}_D &= \left\lfloor 1 + N_0 \left( \frac{\lambda^K - 1}{\lambda - 1} + \underline{\Delta N_{deg}} \right) \right\rfloor \leq T_D \leq \overline{T}_D = \\ &= \left\lceil 1 + N_0 \left( \frac{\lambda^K - 1}{\lambda - 1} + \overline{\Delta N_{deg}} \right) \right\rceil \end{aligned} \quad (43)$$

### 3.2. Время проектирования сценариев

Как отмечалось, момент завершения проектирования сценариев модели определяется наступлением в случайном процессе  $\tilde{G}$  связности графа  $G(n, M)$ , вершинами которого являются элементы декомпозиции бизнес-процесса (бизнес-функции, бизнес-операции и системные сервисы). При этом  $T_S$  определяется количеством ребер  $M$  в модели.

В работе [3] показано, что для  $\forall k \in \mathbb{N}, x \in \mathbb{R}$ , если

$$M(n) = \frac{n}{2} \left( \ln(n) + k \ln(\ln(n)) + x + o(1) \right) \quad (44)$$

то вероятность того, что состояние процесса  $\tilde{G}$  является  $k$ -связным определяется предельным соотношением

$$P\{k(G(n, M)) = k\} \rightarrow 1 - e^{-e^{-x/k}} \quad (45)$$

Используем это соотношение следующим образом. Положим, что требуемая вероятность связности сценариев определяется детальностью модели, т.е.

$$P\{k(G(n, M)) = 1\} = \gamma \quad (46)$$

Используя требование (45), из (46) получаем

$$x = -\ln(-\ln(1 - \gamma)) \quad (47)$$

Тогда, с учетом (47), время  $T_S$ , необходимое для того, чтобы граф  $G(n, M)$  стал связным, на основании выражения (44) можно оценить как

$$T_S = M(T_D) = \frac{T_D}{2} \left( \ln(T_D) + \ln(\ln(T_D)) - \ln(-\ln(1 - \gamma)) \right) \quad (48)$$

Здесь  $T_D$  определяется соотношением (43). Диапазон для  $T_S$  соответственно определяется как

$$M(\underline{T}_D) \leq T_S \leq M(\overline{T}_D) \quad (49)$$

### 3.3. Оценка трудоемкости проектирования

На основании оценки количества элементов декомпозиции модели бизнес-процесса мы можем оценить ожидаемую трудоемкость проектирования. Исходим из того, что трудоемкость, необходимая для декомпозиции пропорционально количеству элементов декомпозиции (43), а трудоемкость построения сценариев бизнес-функций и сервисов операций пропорциональна количеству ребер связного графа модели, определяемому выражением (48):

$$T_P = k_d T_D + k_s T_S \quad (50)$$

Здесь  $k_d$  и  $k_s$  – экспертно определяемые коэффициенты. Эти коэффициенты определяются временем, необходимым для проектирования одним человеком одного элемента декомпозиции и одной связи сценария соответственно. Для расчетов мы будем исходить из следующей экспертной оценки:

$$\begin{aligned} k_d &= 1,5 \div 2,0 \text{ чел} \cdot \text{час} \\ k_s &= 0,1 \div 0,3 \text{ чел} \cdot \text{час} \end{aligned} \quad (51)$$

Отметим, что эти экспертные оценки отражают опыт автора и, вообще говоря, зависят от квалификации специалиста. В каждом конкретном случае представленные коэффициенты должны уточняться с учетом специфики проектной организации.

В качестве примера прогнозную оценку трудоемкости проектирования в зависимости от детальности модели бизнес-процесса представляют табл. 1 для  $N_0 = 4$  ирис.3 для  $N_0 = 8$ .

Табл. 1. Оценка трудоемкости проектирования для  $N_0 = 4$ .

$\gamma$	К	$T_D$		$T_S$		Трудоемкость (чел*час)	
		min	max	min	max	min	max
0,83	2	14	23	21	43	34	59
0,86	2	15	24	23	44	37	61
0,87	2	15	24	22	43	37	61
0,88	3	38	62	79	149	100	169
0,9	3	42	68	89	164	111	185
0,93	3	50	82	107	201	132	224
0,96	3	64	109	141	276	170	301
0,98	3	85	149	194	391	228	415
0,985	3	95	168	220	447	256	470
0,99	3	111	197	263	534	301	554
0,991	3	115	205	273	558	312	577
0,992	4	581	1042	1929	3810	1741	3227
0,993	4	624	1123	2089	4139	1875	3488



Рис 3. График зависимости трудоемкости от детальности модели при  $N_0 = 8$ .

Скачки на графике зависимости времени проектирования от детальности модели соответствуют изменению уровня начала вырождения модели. Интерпретация таких переходов связана со стадиями проектирования информационной системы. Уровень начала вырождения 2 соответствует стадии разработки концепции системы и предполагает определение автоматизируемых операций. Третий или четвертый уровень соответствует стадии технического проектирования, на котором специфицируются требования к системным сервисам. Наконец, пятый и последующие уровни предполагают детальную разработку требований к программным компонентам системных сервисов и соответствуют стадии рабочего проектирования.

#### 4. Проверка модели на экспериментальный данных

Для проверки предложенной стохастической модели процесса идентификации сервисов информационной системы мы рассмотрим два выполненных автором технических проекта информационных систем. Первый проект связан с проектированием аналитической (OLAP) системы, а второй касается информационной системы поддержки концептуального проектирования технически сложных объектов морской техники (проект ТСО МТ).

Проверка будет осуществлена по следующим направлениям:

- 1) Проверка распределения количества элементов декомпозиции каждого уровня на соответствие распределению Пуассона<sup>2</sup>
- 2) Попадание общего количества элементов декомпозиции проектной модели  $\xi_f$  в рамки теоретически предсказанного интервала
- 3) Соответствие общего количества связей в сценариях проектной модели  $M_f$  предсказаниям стохастической модели процесса проектирования.

#### 4.1. Проект OLAP-системы

Распределение количества элементов декомпозиции  $\xi_{(k_i \dots j)}$  в модели OLAP-системы представляет рис. 4.

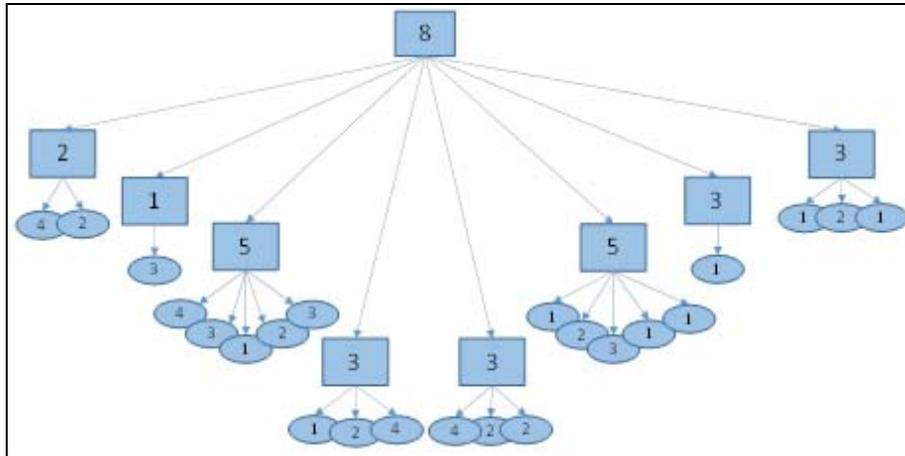


Рис. 4. Структура декомпозиции модели OLAP-системы.

Исходные характеристики модели:

- Количество элементов декомпозиции первого уровня  $N_0 = 8$
- Количество уровней декомпозиции  $k=3$  (начиная с нулевого)
- Количество элементов декомпозиции  $\xi_f = 83$

<sup>2</sup> Проверка гипотезы о том, что распределение количества элементов в декомпозиции соответствует распределению Пуассона, выполнено при помощи интернет-ресурса <http://math.semestr.ru/group/poisson-examples.php>.

- Количество связей в сценариях  $M_f = 249$ .

Распределение количества элементов декомпозиции для уровней 2 и 3, а также для всей модели, полученное на основе анализа структуры модели, представляет табл. 2.

Табл. 2. Распределение количества элементов декомпозиции в модели OLAP-системы.

$i$	$\xi_{2i}$	$\xi_{3i}$	$\xi_t$
1	1	8	9
2	1	7	8
3	4	4	8
4	0	4	4
5	2	0	2
$E[\xi]$	3,13	2,17	2,42

Проверка гипотезы о соответствии распределения количества элементов декомпозиции распределению Пуассона дает следующие результаты:

- 1) Для уровня 2. При уровне значимости 0,4 критическое значение статистики Пирсона  $K_{кр} = 7,38$ , а наблюдаемое значение  $K_{набл} = 0,87$ .
- 2) Для уровня 3. При уровне значимости 0,4 критическое значение статистики Пирсона  $K_{кр} = 7,38$ , а наблюдаемое значение  $K_{набл} = 3,72$ .
- 3) Для всей модели. При уровне значимости 0,4 критическое значение статистики Пирсона  $K_{кр} = 9,35$ , а наблюдаемое значение  $K_{набл} = 3,91$ .

Во всех случаях наблюдаемое значение статистики Пирсона не попадает в критическую область ( $K_{набл} < K_{кр}$ ), поэтому справедливо предположение о том, что данные выборки имеют распределение Пуассона.

Расчет определяет для  $\lambda = \xi_t = 2,42$  ожидаемые значения:

$$\gamma = 0,881$$

$$K = 3$$

$$76 \leq T_D \leq 122$$

$$192 \leq T_S \leq 343$$

Как видно, экспериментальное значение  $\xi_f$  близко к нижней границе диапазона. Это объясняется тем, что уровень вырождения в модели отсутствует, т.е. последний уровень декомпозиции равен теоретическому уровню начала вырождения  $K=3$ .

#### 4.2. Проект ТСО МТ

На рис. 5 представлено распределение количества элементов декомпозиции  $\xi_{ki...j}$  в модели информационной системы поддержки концептуального проектирования ТСО МТ.

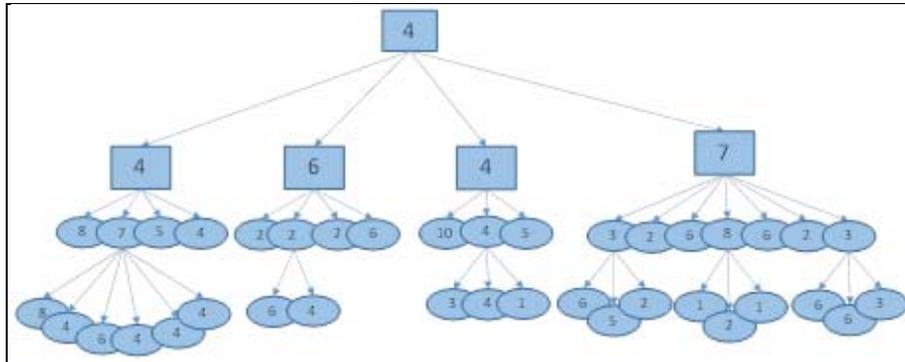


Рис. 5. Структура декомпозиции модели информационной системы поддержки концептуального проектирования ТСО МТ.

Исходные характеристики модели:

- Количество элементов декомпозиции первого уровня  $N_0 = 4$
- Количество уровней декомпозиции  $k = 4$  (начиная с нулевого)
- Количество элементов декомпозиции  $\xi_f = 187$
- Количество связей в сценариях  $M_f = 345$ .

Распределение количества элементов декомпозиции для уровней 2 и 3, а также для всей модели, полученное на основе анализа структуры модели, представляет табл. 3.

Табл. 3. Распределение количества элементов декомпозиции в модели проекта ТСО МТ

$i$	$\xi_{2i}$	$\xi_{3i}$	$\xi_{3i}$	$\xi_t$
1	0	5	3	7
2	0	2	2	4
3	2	2	2	10
4	0	2	6	3
5	1	3	1	9
6	1	1	5	2
7	0	2	0	3
8	0	0	1	0
9	0	1	0	1
$E[\xi]$	5,25	4,72	4,00	4,43

Проверка гипотезы о соответствии распределения количества элементов декомпозиции распределению Пуассона дает следующие результаты:

- 1) Для уровня 2. При уровне значимости 0,05 критическое значение статистики Пирсона  $K_{кр} = 3,84$ , а наблюдаемое значение  $K_{набл} = 2,63$ .
- 2) Для уровня 3. При уровне значимости 0,05 критическое значение статистики Пирсона  $K_{кр} = 12,59$ , а наблюдаемое значение  $K_{набл} = 9,01$ .
- 3) Для уровня 4. При уровне значимости 0,05 критическое значение статистики Пирсона  $K_{кр} = 11,07$ , а наблюдаемое значение  $K_{набл} = 6,68$ .
- 4) Для всей модели. При уровне значимости 0,05 критическое значение статистики Пирсона  $K_{кр} = 14,07$ , а наблюдаемое значение  $K_{набл} = 9,21$ .

Во всех случаях наблюдаемое значение статистики Пирсона не попадает в критическую область ( $K_{набл} < K_{кр}$ ), поэтому справедливо предположение о том, что данные выборки имеют распределение Пуассона.

Расчет определяет для  $\lambda = \xi_t = 4,43$  ожидаемые значения:

$$\gamma = 0,987$$

$$K = 3$$

$$102 \leq T_D \leq 181$$

$$239 \leq T_S \leq 487$$

Экспериментальное значение  $\xi_f \approx \overline{T_D}$  близко к верхней границе диапазона (расхождение составляет 3%). Близость к верхней границе ожидаемого диапазона объясняется тем, что в модели присутствует уровень вырождения, т.е. последний уровень декомпозиции на единицу превосходит теоретический уровень начала вырождения  $K=3$ .

Таким образом, можно сделать заключение, что предлагаемая стохастическая модель процесса идентификации сервисов информационной системы дает хорошее приближение к экспериментальным данным.

## Список литературы

- [1]. Г.Н. Циперман. Применение метода адаптивной кластеризации для проектирования сложных информационных систем. III Научно-практическая конференция «Актуальные проблемы системной и программной инженерии». Сборник научных трудов. Москва, МЭСИ, 6 -7 июня 2013 г.
- [2]. В.А. Ватутин, А.М. Зубков. Ветвящиеся процессы. I. Сборник «Итоги науки и техники». Серия «Теория вероятностей. Математическая статистика. Теоретическая кибернетика». Москва, ВИНТИ, 1985, том 23, стр. 3-67.
- [3]. BelaBollobas. RandomGraphs. Secondedition, CambridgeUniversityPress 2001
- [4]. В.А. Ватутин. Ветвящиеся процессы Беллмана-Харриса. Лекционные курсы НОЦ. Математический институт им. В.А. Стеклова. Москва, МИАН, 2009.
- [5]. Н. Алон, Дж. Спенсер. Вероятностный метод. Москва, БИНОМ. Лаборатория знаний, 2011.

## The stochastic model of information system services identification process

G. Tsiperman <[g.tsiperman@voskhod.ru](mailto:g.tsiperman@voskhod.ru)>

Federal State Unitary Enterprise Organization "VOSKHOD", 85 Udaltsova Str., Moscow, 119607, Russian Federation

**Abstract.** This article refers to labor costs estimation of information system functional requirements design problem in a service-oriented architecture. Stochastic model of information system services identification process is offered for this purpose, which allows providing the estimation of expected project objects and interrelations amount using minimal initial data. This model is based on automated business-process decomposition representation as a branching random supercritical Galton-Watson process. Decomposition elements relation design is modeled as connected random graph construction process by Erdős-Rényi model. Model predictions have been confirmed through verification by experimental data.

**Keywords:** service, identification, stochastic model, branching process, random graph, design time estimation, labor costs, service-oriented architecture

## References

- [1]. G.N.Tsiperman. Primenenie metoda adaptivnoi klasterizatsii dly aproekturovaniya slozhnyh informatsionnyh system [Application of a method of adaptive clustering for complex information systems design] III Nauchno-prakticheskaya konferentsiya «Aktualnye problemy sistemnoii programnoi inzhenerii» [III Scientific-practical conference "Actual problems of system and software engineering"]. Sbornik nauchnyh trudov [Collection of proceedings]. Moscow, MESI, 6 -7 June 2013 г. (in Russian).
- [2]. V.A. Vatutin, A.M. Zubkov. Vetyaschiesya protsessy. I [Branchingprocess. I]. Sbornik «Itogi nauki I tekhniki». [The book "The Results of science and technology"]. Moscow, VINITI, 1985, vol. 23, 3-67 p.(in Russian).
- [3]. BelaBollobas. RandomGraphs. Secondedition, CambridgeUniversityPress 2001
- [4]. V.A. Vatutin. Vetyaschiesya protsessy Bellman-Harris [Branching processes Bellman-Harris]. LektsionnyekursyNOTS [Lecture courses]. Mathematical instituteV.A. Steklov. Moscow, MIAN, 2009.(in Russian).
- [5]. N. Alon, J.H. Spencer. The Probabilistic Method. Third Edition. JohnWiley&Sons, 2008.