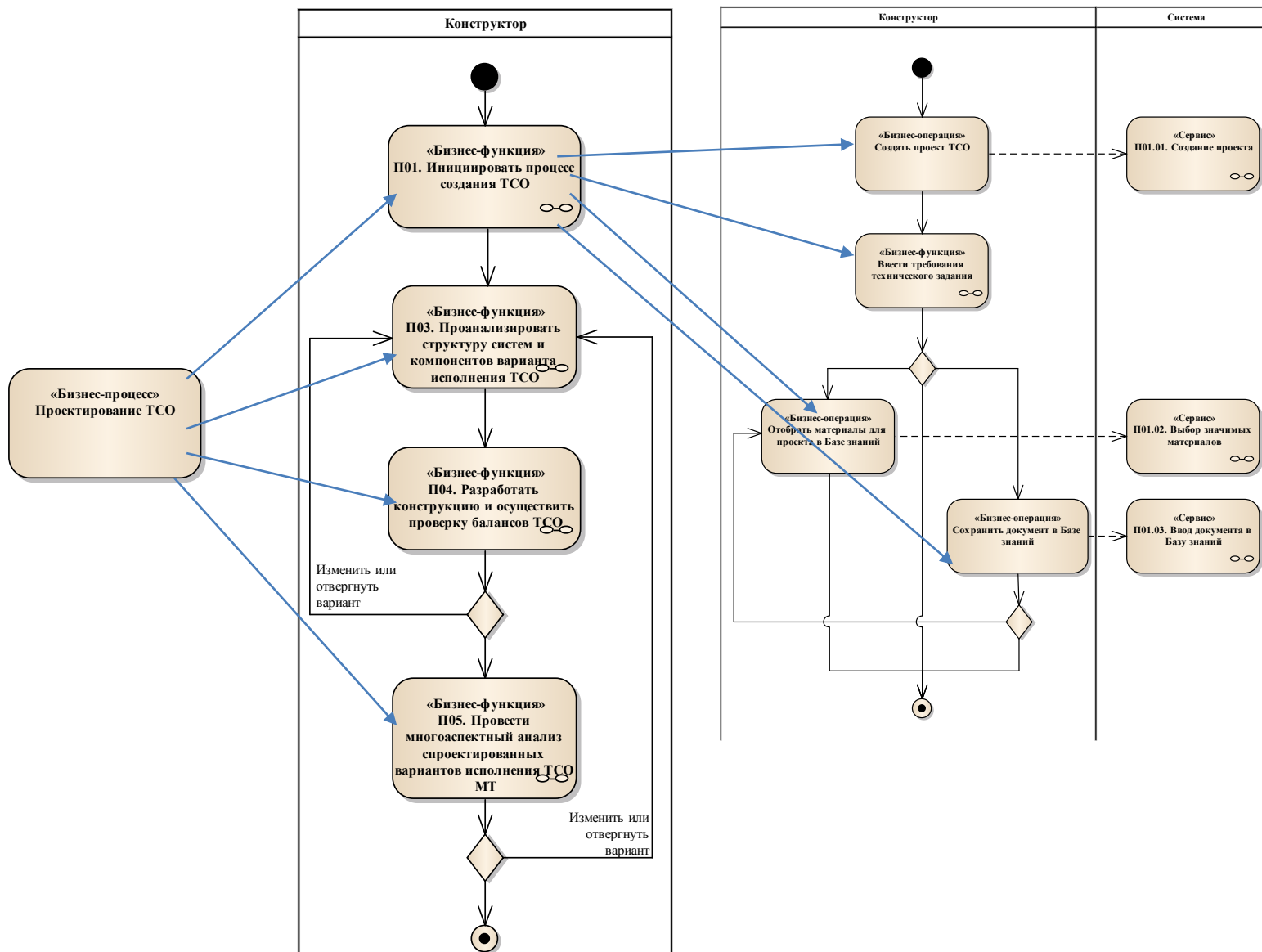


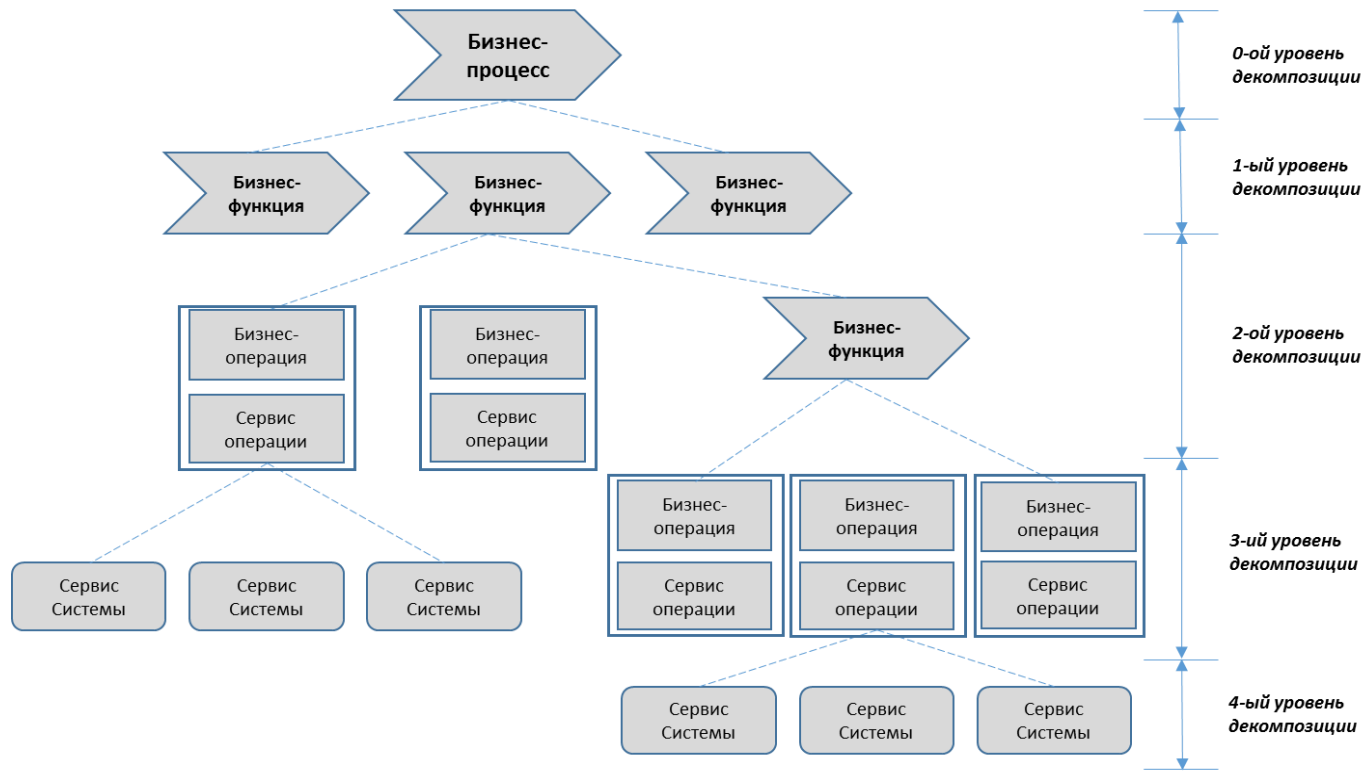
Стохастическая модель процесса идентификации сервисов информационной системы

V Международная конференция «Облачные вычисления.
Образование. Исследования. Разработка»
4-5 декабря 2014 года

Григорий Циперман
Советник директора



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



- Процесс декомпозиции бизнес-процесса рассматривается как ветвящийся случайный надкритический процесс Гальтона-Ватсона
- Проектирование сценариев моделируется как процесс построения связного случайного графа в модели Эрдеша-Реньи

■ Ветвящийся процесс

$$D_0 = D(0) = (f_0); D_1 = D(f_0) = (x_1, x_2, \dots, x_n)$$

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

$$D_{\underbrace{ki\dots j}_k} = D(\underbrace{x_{i\dots j}}_{k-1}) = (\underbrace{x_{i\dots j1}}_k, \dots, \underbrace{x_{i\dots jp}}_k)$$

■ Вероятностное пространство $(\Omega, \mathcal{F}, P)$

Пусть $\xi_{ki\dots j}$ – количество элементов декомпозиции $D_{ki\dots j}$. Элементарной реализацией процесса декомпозиции назовем набор $\omega = (\xi_0, \xi_{1i}, \xi_{2ip}, \dots, \xi_{kip\dots j})$

- ✓ $\Omega = \{\omega\}$
- ✓ σ -алгебра $\mathcal{F}$ порождается подмножествами пространства Ω
- ✓ Вероятностная мера P задается соотношениями

$$P\{\xi_{kip\dots j} = k\} = p_k$$

$$P\{(\xi_0, \xi_{1i}, \xi_{2ip}, \dots, \xi_{kip\dots j}) = (k_0, k_{1i}, k_{2ip}, \dots, k_{kip\dots j})\} = p_{k_0} p_{k_{1i}} p_{k_{2ip}} \dots p_{k_{kip\dots j}}$$

■ Условия задачи

- ✓ Количество элементов k -ого поколения: $Z(k) = |\{D_{ki\dots j}\}| = \sum_{i\dots j} \xi_{ki\dots j}$
- ✓ $Z(0) = 1, Z(1) = N_0$, где N_0 неслучайная величина
- ✓ Ненулевая вероятность вырождения $\alpha > 0$
- ✓ Условие надкритичности: $E[\xi] > 1$

■ Постановка задачи:

Требуется найти наиболее вероятное число элементов декомпозиции (ожидаемое число структурных компонент модели ИС):

- Число элементов декомпозиции, относящихся к k -ому поколению имеет распределение Пуассона

$$P\{\xi_k = x\} = \frac{\lambda^x}{x!} e^{-\lambda}$$

- Математическое ожидание распределения элементов декомпозиций $\lambda = E[\xi]$ не зависит от поколения процесса
- Детальность** модели $\gamma = 1 - \alpha$ связана с математическим ожиданием λ известным соотношением

$$f(\gamma) = e^{-\lambda\gamma} = 1 - \gamma$$

где $f(\gamma)$ – производящая функция. Отсюда

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

- Математическое ожидание количества элементов k -ого поколения

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

Тогда общее количество элементов декомпозиции до уровня вырождения K

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

- Количество элементов на уровне вырождения процесса

$$\Delta N_{deg} = N_0 \lambda^K \frac{1 - \gamma}{1 - \lambda(1 - \gamma)}$$

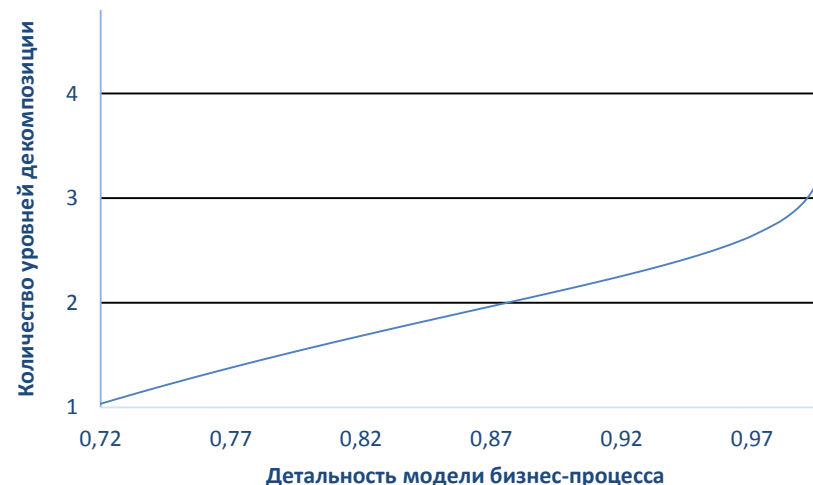
- Количество элементов декомпозиции

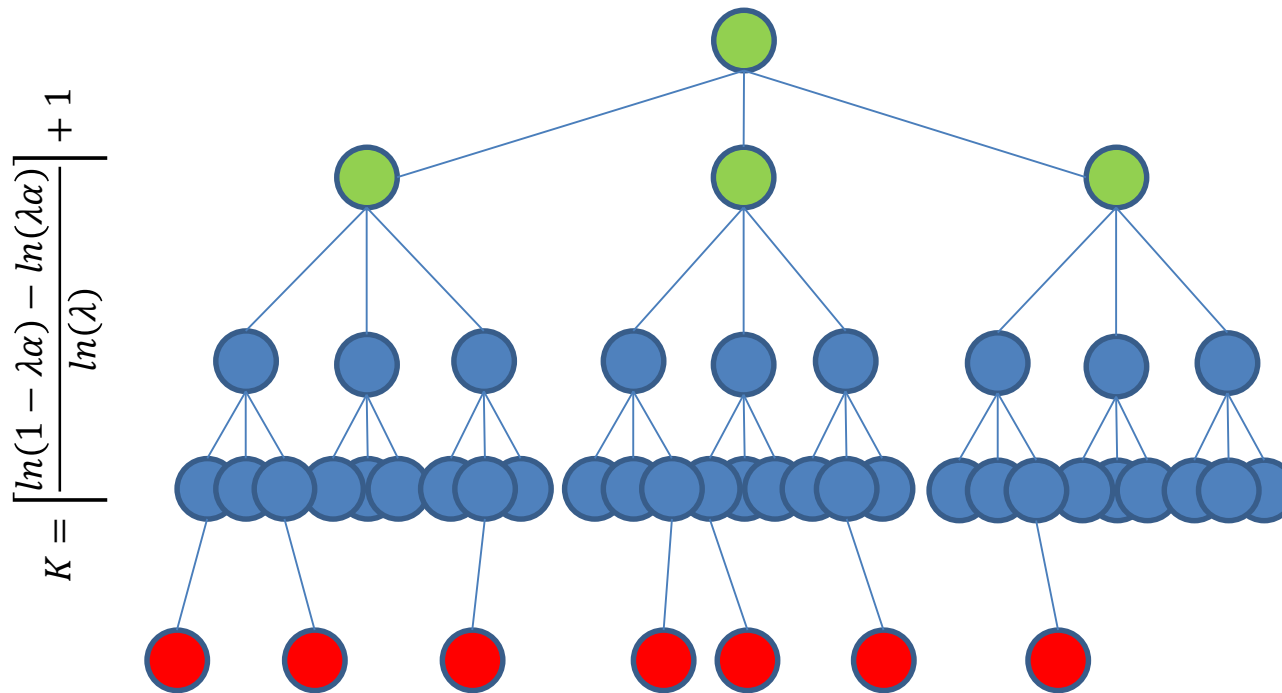
$$T_D = n + \Delta N_{deg}$$

- Число уровней декомпозиции (с точностью до +1) определяется соотношением

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

- Полученная оценка согласуется с реальной практикой проектирования, где в основном декомпозиция заканчивается на третьем уровне, а уровень 5 встречается достаточно редко
- Детальность модели γ определяется стадиями проектирования:
 - $0,72 < \gamma < 0,88$ – концептуальное проектирование (3-4 уровня декомпозиции)
 - $0,88 \leq \gamma < 0,992$ – техническое проектирование (4-5 уровней декомпозиции)
 - $\gamma \geq 0,992$ – рабочее проектирование (от 6 уровней декомпозиции)





$$Z(0) = 1$$

$$Z(1) = N_0$$

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

$$\Delta N_{deg} = N_0 \lambda^K \frac{1 - \gamma}{1 - \lambda(1 - \gamma)}$$

- Диапазон количества элементов декомпозиции на уровне вырождения

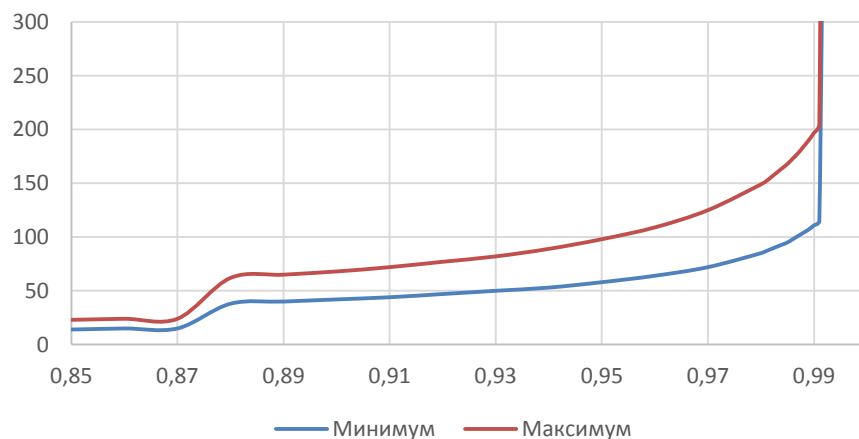
$$\underline{\Delta N_{deg}} = \left\lfloor N_0 \lambda^K \frac{1-\gamma}{1-\lambda(1-\gamma)} \right\rfloor \leq \Delta N_{deg} \leq \left\lceil N_0 \lambda^K \frac{1-\gamma}{1-\lambda(1-\gamma)} \right\rceil = \overline{\Delta N_{deg}}$$

- Диапазон общего количества элементов декомпозиции

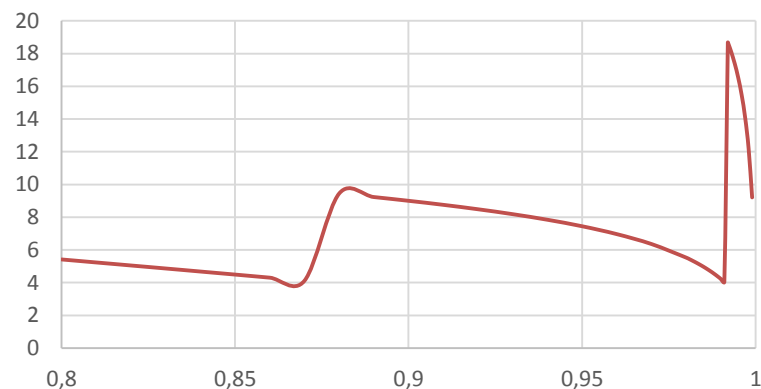
$$\underline{T_D} = \left\lfloor 1 + N_0 \frac{\lambda^K - 1}{\lambda - 1} + \underline{\Delta N_{deg}} \right\rfloor \leq T_D \leq \overline{T_D} = \left\lceil 1 + N_0 \frac{\lambda^K - 1}{\lambda - 1} + \overline{\Delta N_{deg}} \right\rceil$$

$$N_0 = 4$$

Количество элементов декомпозиции



Количество элементов декомпозиции на уровне вырождения



- Вероятностное пространство $\mathfrak{L}(n, M)$ всех графов порядка n и размера M
 - ✓ Случайный граф $G(n, M)$, где M – случайное множество ребер, $0 \leq M \leq N = \binom{n}{2}$
 - ✓ $\mathfrak{L}(n, M) = \{G(n, M)\}; |\mathfrak{L}(n, M)| = \binom{N}{M}$
 - ✓ $P\{G = G(n, M)\} = \binom{N}{M}^{-1}$
- Процесс на случайном графе $\tilde{G} = (G_t)_0^N$
 - ✓ $\tilde{G} = (G_t)_0^N$ – марковский процесс на $V = \{1, 2, \dots, n\}$
 - ✓ $(G_0)_0^N = 0$
 - ✓ Каждый G_t представляет собой граф на V
 - ✓ G_t имеет t ребер, где $t = 1, 2, \dots, N$
 - ✓ $G_0 \subset G_1 \subset \dots$
 - ✓ Состояние процесса G_t в момент времени $t = M$ отождествляется с элементом пространства $\mathfrak{L}(n, M)$:
$$(G_t)_0^N \rightarrow G(n, M)$$
- Постановка задачи:

Задачей исследования процесса построения сценария проектируемой ИС является определение необходимого количества времени T_S для формирования связного графа $G(n, M)$.

На практике это означает, что полностью определена связь бизнес-функций и входящих в них бизнес-операций и все состояния проектируемой ИС связаны между собой хотя бы одним маршрутом

■ Теорема Боллобаша

Для $\forall k \in \mathbb{N}, x \in \mathbb{R}$, если

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

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

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

■ Ожидаемое количество связей в односвязном графе ($k=1$)

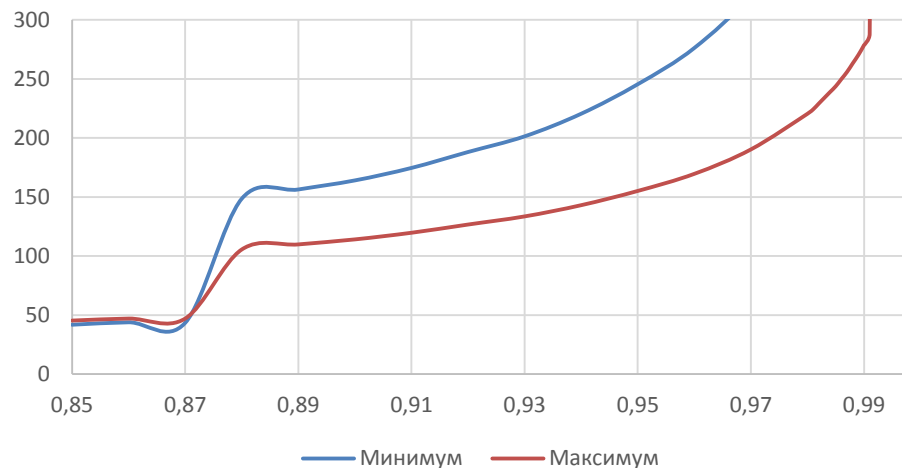
$$T_S = M(T_D) = \frac{T_D}{2} \left(\ln(T_D) + \ln(\ln(T_D)) - \ln(-\ln(1 - \gamma)) \right)$$

Диапазон для T_S соответственно определяется как

$$M(\underline{T_D}) \leq T_S \leq M(\overline{T_D})$$

Количество связей

$N_0 = 4$



Трудоемкость процесса идентификации сервисов ИС в СОА

$$T_P = k_d T_D + k_s T_S$$

Здесь k_d и k_s – экспертно определяемые коэффициенты.

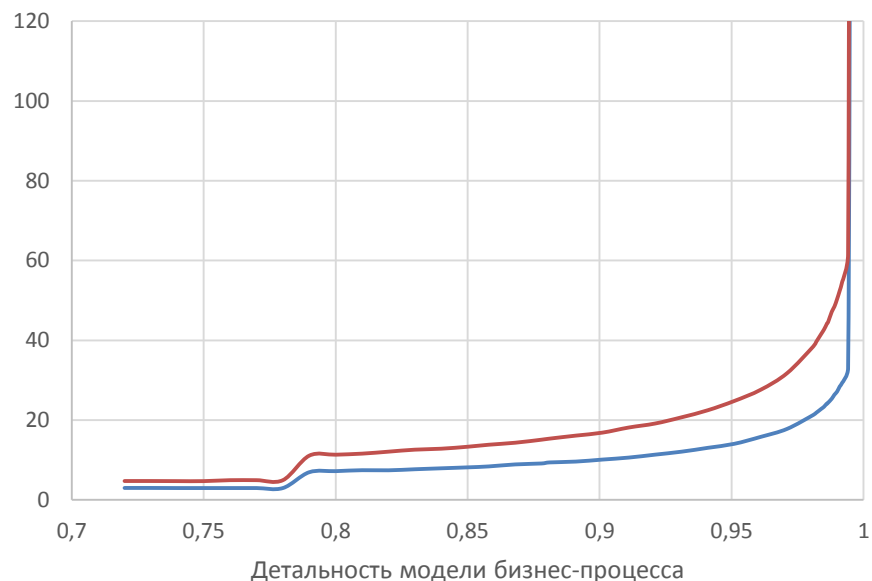
Эти коэффициенты определяются временем, необходимым для проектирования одним человеком одного элемента декомпозиции и одной связи сценария соответственно. Для расчетов мы будем исходить из следующей экспертной оценки:

$$k_d = 1,5 \div 2,0 \text{ чел} \cdot \text{час}$$

$$k_s = 0,1 \div 0,3 \text{ чел} \cdot \text{час}$$

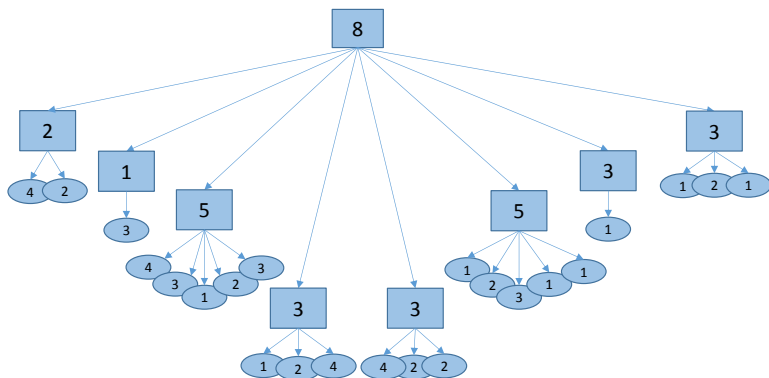
$$N_0 = 4$$

γ	κ	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



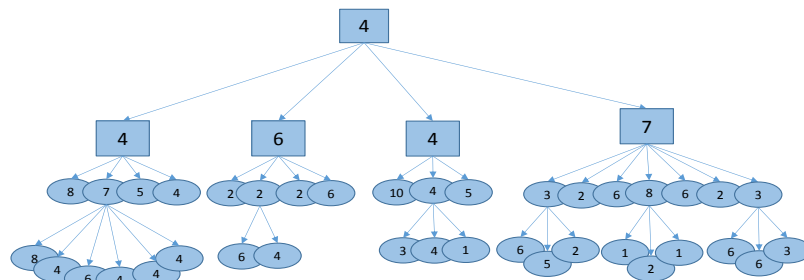
Проект OLAP-системы

- Исходные характеристики:
 - Количество элементов декомпозиции первого уровня $N_0 = 8$
 - Количество уровней декомпозиции $k=3$
 - Количество элементов декомпозиции $\xi_f = 83$
 - Количество связей в сценариях $M_f = 249$
 - Мат. ожидание количества элементов декомпозиции $\lambda = 2,42$
- Проверка по критерию Пирсона подтверждает гипотезу о пуассоновском распределении количества элементов декомпозиции для всех уровней модели
- Расчетные (модельные) значения метрик:
 - $\gamma = 0,881$
 - $K = 3$
 - $76 \leq T_D \leq 122$
 - $192 \leq T_S \leq 343$



Проект TCO MT

- Исходные характеристики:
 - Количество элементов декомпозиции первого уровня $N_0 = 4$
 - Количество уровней декомпозиции $k=4$
 - Количество элементов декомпозиции $\xi_f = 187$
 - Количество связей в сценариях $M_f = 345$
 - Мат. ожидание количества элементов декомпозиции $\lambda = 4,43$
- Проверка по критерию Пирсона подтверждает гипотезу о пуассоновском распределении количества элементов декомпозиции для всех уровней модели
- Расчетные (модельные) значения метрик:
 - $\gamma = 0,987$
 - $K = 3$
 - $102 \leq T_D \leq 181$
 - $239 \leq T_S \leq 487$



Спасибо за внимание