

АЛГОРИТМ ТОЧЕЧНОГО ПРОГНОЗИРОВАНИЯ СЛУЧАЙНЫХ ПРОЦЕССОВ В АВИАЦИОННЫХ ИНФОКОММУНИКАЦИОННЫХ СЕТЯХ

Д.В. Чирков, В.Г. Липовский

Национальный авиационный университет,
просп. Космонавта Комарова, 1, Киев, 03680, Украина

В работе рассмотрена общая методика прогнозирующего контроля современных мультисервисных сетей, которая в качестве основного компонента включает алгоритмы структурной идентификации моделей регулярных составляющих (трендов) наблюдаемых процессов. Приведена общая структура методики прогнозирующего контроля авиационных инфокоммуникационных систем. Представлен общий вид алгоритма структурной идентификации.

Ключевые слова: инфокоммуникационные сети, авиационные сети, случайные процессы, точечное прогнозирование, алгоритм прогнозирования, мультисервисные сети

Введение

Решение задач прогнозирующего контроля сводится к принятию решения по предупреждению прогнозируемого отказа.

Поскольку в решаемой задаче информация о том, какие решающие правила являются оптимальными, рассматриваются алгоритмы, позволяющие получать практически приемлемые результаты. Важнейшими условиями обеспечения этого являются:

- построение моделей прогнозирования для конкретных подсетей на основе реальных данных об их функционировании;
- приемлемая глубина контроля, позволяющая в реальных условиях собирать и анализировать измерительную информацию, не сводя деятельность оператора или провайдера исключительно к решению задач контроля;
- привлечение квалифицированных экспертов для формирования прогностических моделей;
- использование алгоритмов прогнозирования, позволяющих решать задачи структурной идентификации с приемлемой точностью, за приемлемое время и отличающихся робастностью – устойчивостью к выбросам реализаций наблюдаемых процессов;
- обеспечение высокой точности прогнозирования, для чего необходимо выполнение принципа максимума воспроизводимости, который ниже и будет рассматриваться как альтернатива принципу максимума правдоподобия.

Таким образом, задачей является синтез единой методики прогнозирующего контроля, включающей организационные, технические, алгоритмические и программные средства, позволяющие решать задачи прогнозирующего контроля с приемлемым качеством для авиационных инфокоммуникационных систем (АИС).

Основная часть

Общая структура методики прогнозирующего контроля приведена на рис. 1.

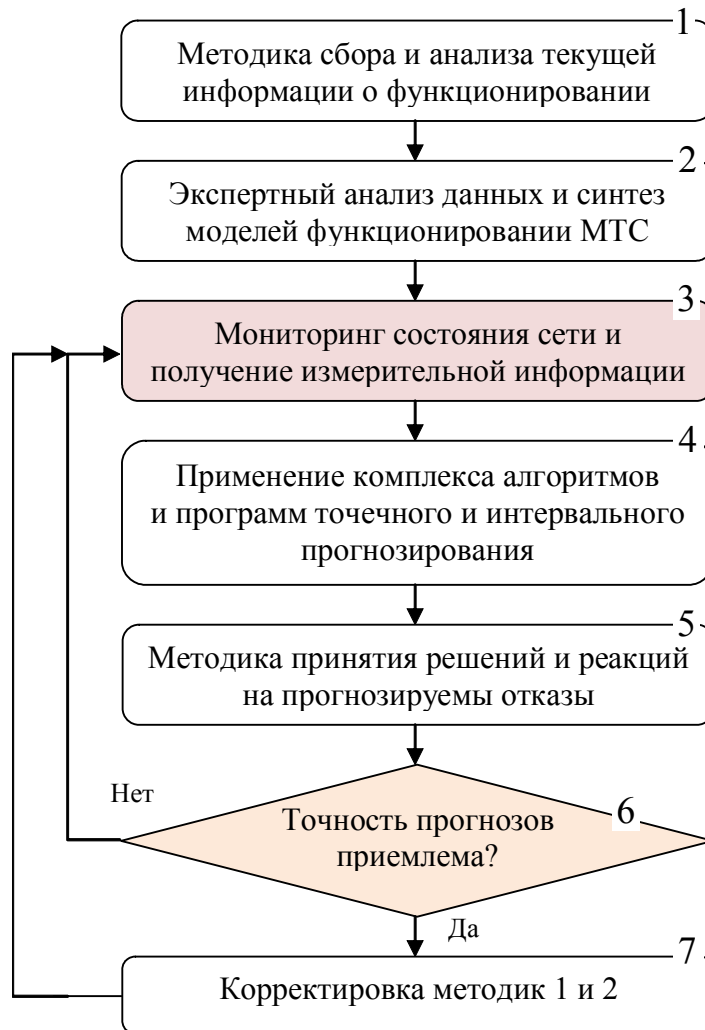


Рис. 1. Общая структура методики прогнозирующего контроля АИС

Разрабатываемая методика, как видим, содержит блоки, декомпозирующие решение общей задачи прогнозирующего контроля на подзадачи, логически вытекающие одна из другой. При этом сама методика представляет собой замкнутую структуру, включающую автоматизируемые и неавтоматизируемые («экспертные») составляющие. При этом должен быть разработан комплекс организационно-технических мероприятий, позволяющий решать задачу прогнозирования отказов с приемлемой точностью.

Рассмотрим по отдельности составляющие методики, приведенной на рис. 1.

Методика сбора и анализа текущей информации о функционировании АИС включает:

1) Методики определения контрольных точек сети, в которой учитывается возможность выполнения прямых или косвенных измерений параметров.

Данная группа методов и средств зависит от конкретной топологии сети. Каждый провайдер (оператор) строит свой центр управления сетью – *network operation centre* (NOC) и выбирает множество контрольных точек сети, исходя из следующих основных положений:

- в обязательном порядке контролируется загрузка портов абонентов выделенных каналов передачи данных;
- в обязательном порядке контролируется загрузка внешних портов, обращенных к портам вышестоящего провайдера;
- данные об отказах типа «непредоставление» услуги фиксируется в диспетчерских журналах.

2) Методики анализа результатов измерений параметров сети, причин возникновения и способов предупреждения отказов.

Экспертный анализ данных и синтез моделей функционирования АИС. Синтезированные в методике модели носят достаточной общий характер и могут применяться для моделирования и прогнозирования параметров реальных сетей с учетом необходимости применения методов структурной идентификации.

Мониторинг состояния сети и получение измерительной информации выполняется на NOC, а также методами периодического контроля и анализа журналов диспетчерских и технических служб.

Применение комплекса алгоритмов и программ точечного и интервального прогнозирования. Разработке такого алгоритмического и программного обеспечения посвящена данная работа.

Методика принятия решений и реакций на прогнозируемые отказы зависит от технической и маркетинговой политики конкретного провайдера. В частности, при планировании внедрения новой услуги, обязательным является оценивания класса QoS, обеспечиваемого данной сетью, прогнозирование возможных отказов с учетом вновь предлагаемого качества и состава услуг, принятие решений о модернизации сети, расширении полос пропускания внешних портов и т.д.

Назначение и смысл фрагментов 6 и 7 общей методики прогностического контроля (рис. 1) очевидны.

Таким образом, рассмотрена общая структура методики прогнозирующего контроля, которая содержит наиболее сложный с научной точки зрения структурный элемент – комплекс алгоритмов и программ точечного и интервального оценивания.

Процедура прогнозирования в общем виде сводится к следующим операциям:

- 1) Выбор класса моделей и получение данных измерений;
- 2) Решение задачи структурной идентификации (алгоритм решения этой задачи в общем виде показан на рис. 2);
- 3) Экстраполяция оптимальной по критерию воспроизводимости модели на будущие периоды времени.
- 4) Определение момента первого выхода прогноза за пороговое значение.

Рассмотрим эти операции применительно к классу решаемых задач.

1. *Выбор класса моделей.* Модели регулярных составляющих исследуемых процессов двух основных видов – в виде степенного полинома.

Поскольку каждый из факторов $\varphi_k(t)$ допускает многократные измерения, можно считать его квазинепрерывной функцией и для его аппроксимации использовать разложения в ряд вида [1], который в данном случае вырождается в степенной ряд. Для решения задач идентификации с заданной точностью ограничимся моделями вида [2]:

$$\varphi_k(t) = a_{k0} + a_{k1}t + a_{k2}t^2 + \dots + a_{kN}t^N + \varepsilon_k, \quad (1)$$

где

N — количество членов степенного ряда обеспечивающих точность и устойчивость модели фактора у вариациям данных;

$k = \overline{1, K}$ — условный номер фактора в фактор-модели;

ε_k — ошибка модели или случайная составляющая процесса. И модели [3]

$$\begin{aligned} \varphi_k(t) = & \varepsilon(t) + a_0 + a_1 t + a_2 t^2 + \dots + a_N t^N + \\ & + a_{N+1} \cos(\omega_1 t + \beta_1) + a_{N+2} \cos(\omega_2 t + \beta_2) + \dots + a_{N+3} \cos(\omega_3 t + \beta_3), \end{aligned} \quad (2)$$

где ε_k — нерегулярная (случайная) составляющая наблюдаемого процесса, содержащей гармонические компоненты.

2. *Решение задачи структурной идентификации* (рис. 2). В качестве базового метода для решения этой задачи определен метод максимума компактности.

Во-первых, в качестве метода параметрической идентификации используется метод среднего [1].

Применительно к задаче параметрической идентификации моделей вида (1) алгоритм, реализующий метод среднего сводится к следующим операциям.

1) Для очередной структуры модели

$$(s_{l,1}, s_{l,2}, \dots, s_{l,N}), \quad \forall l, n : s_{l,n} \in \{0,1\}, \quad n = \overline{1, N}, \quad (3)$$

определяется количество ее свободных параметров:

$$M_l = \sum_{n=1}^N s_{l,n}. \quad (4)$$

2) Пробная выборка $Y_{\text{пробн}}(t_k) = Y_k$, $k = \overline{1, K}$ объема K делится последовательно на M_l подвыборок приблизительно одинакового объема. Не снижая в значительной мере общности алгоритмов будем полагать, что все подвыборки имеют одинаковый объем q .

3) Для каждой из подвыборок формулируется требование: модель должна в среднем совпадать с данными подвыборки, что дает систему M_l линейных уравнений с M_l неизвестными:

$$\left\{ \begin{aligned} s_{l,1} b_0 q + s_{l,2} b_1 \sum_{k=1}^q t_k + \dots + s_{l,N} b_N \sum_{k=1}^q t_k^N &= \sum_{k=1}^q y_k \\ s_{l,1} b_0 q + s_{l,2} b_1 \sum_{k=q+1}^{2q} t_k + \dots + s_{l,N} b_N \sum_{k=q+1}^{2q} t_k^N &= \sum_{k=q+1}^{2q} y_k \\ s_{l,1} b_0 q + s_{l,2} b_1 \sum_{k=(M_l-1)q+1}^{qM_l} t_k + \dots + s_{l,N} b_N \sum_{k=(M_l-1)q+1}^{qM_l} t_k^N &= \sum_{k=(M_l-1)q+1}^{qM_l} y_k \end{aligned} \right. \quad (2)$$

где b — свободные параметры модели.

Как показано, в [4] оценки метода среднего являются несмещенными и состоятельными. Они уступают по эффективности оценкам метода наименьших квадратов приблизительно в $\sqrt{M_l}$ раз, где M_l — количество свободных параметров модели. Например, для 4-х параметрической модели эти оценки теоретически дают вдвое большую дисперсию погрешности, чем оценки МНК. Вместе с тем, они являются более устойчивыми с вычислительной точки зрения и существенно более робастными — более устойчивыми к отдельным выбросам реализаций случайных процессов. Последнее свойство с учетом особенностей наблюдаемых реализаций является в условиях решаемых задач более важным.



Рис. 2. Общий вид алгоритма структурной идентификации

Система уравнений (5) решается одним из известных методов численного решения систем линейных алгебраических уравнений, например методом исключения с прямым и обратным ходом.

Сделаем ряд замечаний о выборе показателя воспроизводимости Q_l . При обосновании применимости метода максимума компактности рассматривался показатель K_f , основанный на сравнении плотностей распределения ошибок модели на пробной и контрольной выборках.

Наиболее правдоподобному варианту плотности распределения вероятностей $f(X)$ переменной X из числа рассматриваемых вариантов соответствует максимум показателя воспроизводимости:

$$K_f = \int_{-\infty}^{\infty} \inf_f \{f_{\text{пробн}}(x) - f_{\text{контр}}(x)\} dx,$$

где $f_{\text{пробн}}(x)$, $f_{\text{контр}}(x)$ — оценки плотности вероятности по пробной и контрольной выборкам.

Этот показатель имеет некоторые недостатки, применительно к решаемым задачам:

- результаты измерений представляются дискретными рядами. Для дискретных величин понятие плотности распределения, вообще говоря, не определено;
- даже аппроксимация неизвестных распределений в условиях начальной задачи статистики плотностями превращается в существенную проблему.

В настоящей работе предлагается использовать показатели, основанные на сравнении интегральных функций распределения величин. Такие показатели имеют следующие преимущества:

- функцию распределения имеют любые величины – дискретные, непрерывные, комбинированные;
- имеется простая статистика, позволяющая непосредственно по данным измерений $Y_{\text{изм}}(t_k) = Y_k$, $k = \overline{1, N}$ строить эмпирические функции распределения:

$$F_{\text{изм}}(y) = \frac{1}{K} \sum_{k=1}^K h(y - Y_k), \quad h(y) = \begin{cases} 0, & y < 0 \\ 1, & y \geq 0 \end{cases}. \quad (6)$$

Для сравнения пар функций распределения можно использовать различные метрики [5]. Конкретный показатель Q_l , основанный на квантильной метрике используемой далее.

Для вычисления показателя воспроизводимости по пробной и контрольной выборке одинакового объема K , $2K=N$, где N — общий объем выборки применяется следующий алгоритм, для модели с номером 1 вычисляются ее невязки на участках пробной и контрольной выборок:

$$\delta_{l,k}^{\text{пр}} = Y_k - y_l(t_k), \quad k = \overline{1, K}, \quad (7)$$

$$\delta_{l,k}^{\text{конт}} = Y_k - y_l(t_k), \quad k = K+1, K+2, \dots, 2K. \quad (8)$$

Из полученных выборок невязок (7) и (8) формируются ранжированные выборки, содержащие те же самые числа, но расположенные в порядке возрастания. Далее эти выборки обозначаются так же, как и выборки (7) и (8) с учетом того, что в ранжированных выборках $\delta_{l,k}^{\text{пр,конт}} \leq \delta_{l,k+1}^{\text{пр,конт}}$. Очевидно, что члены ранжированных выборок составляют последовательность квантилей порядка $1/K$ эмпирических распределений. Показатель воспроизводимости вычисляется после этого как средний

модуль отклонения квантилей распределений невязок на пробной и контрольной выборках:

$$Q_i = \frac{1}{K} \sum_{k=1}^K |\delta_{i,k}^{\text{пр}} - \delta_{i,k}^{\text{конт}}|. \quad (9)$$

Показатель Q_i вычисляется для моделей всевозможных структур и в качестве оптимальной выбирается та из них, для которой он принимает минимальное значение (блок 6 алгоритма рис. 2).

Рассмотрим вычисление показателя (9) для линейной модели. Из табл. 1 следует, что наиболее сложной операцией при вычислении показателя 9 является операция ранжирования выборок. Такая операция выполняется многократным применением алгоритма поиска минимального значения массива чисел и не представляет принципиальных сложностей. В то же время, очевидно, что требованию линейности критериев метода максимума компактности показатель (9) удовлетворяет.

Таблица 1.

Вычисление показателя воспроизводимости для линейной модели

Номер недели, t_k	$Y_{\text{пробн}}(t_k)$	$Y_{\text{контр}}(t_k)$	$Y_1(t_k)$	σ_k	σ_k ранжир	$ \delta_{i,k}^{\text{пр}} - \delta_{i,k}^{\text{конт}} $
1	29.3		22.9	6.4	-7.1	
2	19.1		26.2	-7.1	-5.1	
3	26.2		29.5	-3.3	-3.3	
4	39.3		32.8	6.5	-1.4	
5	31.0		36.1	-5.1	1.3	
6	42.2		39.4	2.8	2.8	
7	41.4		42.8	-1.4	6.4	
8	47.4		46.1	1.3		
9		50.9	49.4	1.5	-13.6	6.5
10		58.5	52.7	5.8	-6.5	1.4
11		61.6	56.0	5.6	-6.0	2.7
12		53.3	59.3	-6.0	0.3	1.7
13		56.1	62.6	-6.5	1.5	0.2
14		52.3	65.9	-13.6	2.2	0.6
15		71.4	69.2	2.2	5.6	0.8
16		72.8	72.5	0.3	5.8	5.8
Сумма модулей межквантильных отклонений						19.7
Значение показателя воспроизводимости Q						2.5

Для сравнения моделей различных структур аналогичные вычисления показателя Q_i проводятся по тому же алгоритму. Результаты вычислений приведены в табл. 2. Анализ этой таблицы показывает, что модель в виде полинома первой степени является оптимальной по выбранному критерию, что согласуется с физической стороной решаемой задачи.

Отдельную проблему представляет параметрическая идентификация моделей вида (2) с гармоническими составляющими. Даже при известных периодах гармонических составляющих (сутки, неделя, год), соответствующие фазы входят в уравнения нелинейно. Точных методов решения систем нелинейных уравнений не существует, что может приводить к неоднозначности прогнозирования.

Таблица 2.

Значения показателей воспроизводимости для моделей различных структур

№ модели	Вектор структуры			Вид модели	Q
	S ₁	S ₂	S ₃		
1	1	0	0	b ₀	7.2
2	0	1	0	b ₁ t	8.3
3	1	1	0	b ₀ + b ₁ t	2.5
4	0	0	1	b ₂ t ²	18.6
5	1	0	1	b ₀ + b ₂ t ²	11.3
6	0	1	1	b ₁ + b ₂ t ²	12.7
7	1	1	1	b ₀ + b ₁ + b ₂ t ²	8.4

В рассматриваемых случаях, однако, имеется возможность привязки по фазе к характерным особенностям графиков загрузки портов. А именно, в суточных и недельных графиках имеются характерные особенности – час наибольшей нагрузки (ЧНН), день наибольшей нагрузки в неделю (ДНН), час минимальной нагрузки (ЧМН), день минимальной нагрузки (ДМН).

Из графика загрузки одного из портов (рис. 3) магистрального свитча условного провайдера 2 за двое суток видно, что падение загрузки в ЧМН является весьма устойчивым и соответствует примерно 5-6 часам утра.

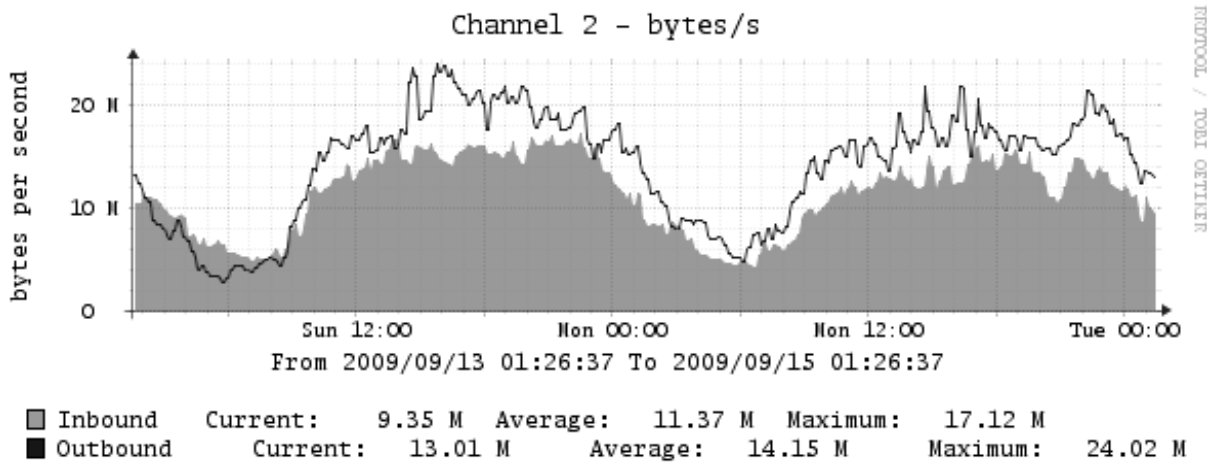


Рис. 3. График загрузки порта 2 условного провайдера за 2 суток

С учетом устойчивости фазы ЧМН применим квазиоптимальный метод ее определения. Как видно суточные колебания нагрузки составляют приблизительно от 40 Мбит/с в ЧМН до 160 Мбит/с в ЧНН. За этот же период линейная долгосрочная составляющая процесса изменения нагрузки изменится приблизительно на 3–30 Мбит/с. На таком коротком временном интервале, таким образом, можно ограничиться простой моделью тренда:

$$y(t) = a_0 + a_1 \cos(\omega_1 t + \beta_1), \tag{10}$$

где ω_1 — известная круговая частота, соответствующая суточному периоду; ее численное значение зависит от единицы измерения времени (секунда, минута, обычно – час).

Поскольку за период сумма отсчетов гармонической составляющей, выполненных через равные интервалы времени, равно нулю, то оценка постоянной составляющей a_0 в зависимости (10) получается так:

$$a_0 = \frac{1}{N} \sum_{n=1}^N Y_n . \quad (11)$$

После этого в модели (10) остается 2 свободных параметра – амплитуда суточных колебаний a_1 и фаза β_1 . С учетом высокого быстродействия современных компьютеров и невысокой частоты измерений задача определения этих параметром может быть решена комбинированным алгоритмом – полного перебора по фазе и поиска оптимума по амплитуде, а именно:

- осуществляется выбор очередного возможного значения фазы $\beta_{1,k} = t_k$, $k = \overline{1, N}$;
- при заданном значении фазы $\beta_{1,k}$ методом половинного деления или другим быстро сходящимся методом минимизации функции одной переменной осуществляется поиск минимума по a_1 функционала

$$\Phi_k = \sum_{n=1}^N |a_1 \cos(\omega_1 t_n + \beta_k) - (Y_n - a_0)| \xrightarrow{a_1} \min ; \quad (12)$$

- в качестве оптимальной выбирается фаза β_k , для которой значение функционала Φ_k минимально.

Для определяются фазы недельной составляющей выполняется идентификация моделей вида (10) для суточных составляющих 1-х, 2-х, ..., 7-х суток недели. Фаза выбирается так, чтобы ДНН соответствовал суткам с максимальным значением параметра a_1 .

Как видим, после определения фаз модели вида (2) идентификация ее свободных параметров сводится к составлению и решению систем линейных уравнений, аналогичных системе уравнений

$$\omega_i = \omega_i : \{z_i \notin [z_{i\min}, z_{i\max}]\}$$

Таким образом, в настоящей работе рассмотрены алгоритмы идентификации моделей, позволяющие осуществлять точечные прогнозы.

Выводы

В работе рассмотрена общая методика прогнозирующего контроля современных мультисервисных сетей. Эта методика в качестве основного компонента включает алгоритмы структурной идентификации моделей регулярных составляющих (трендов) наблюдаемых процессов. Полученные результаты позволяют сформулировать ряд выводов.

1) Предложенный для решения задач параметрической идентификации метод среднего доставляет устойчивые с вычислительной точки зрения и робастные оценки.

2) Предложенный показатель воспроизводимости моделей по данным контрольной выборки удобен с вычислительной точки зрения и не требует априорного знания законов распределения наблюдаемых величин.

3) В зависимостях наблюдаемых факторов, влияющих на отказоустойчивость, имеются выраженные периодичности. Для решения задач идентификации целесообразно применять квазиоптимальные методы решения нелинейных уравнений, когда неизвестная фаза определяется по характерным точкам. При этом общая задача декомпозируется на подзадачу определения фазы гармонической составляющей и последующему решению системы линейных уравнений.

Список литературы

1. Волков, Е.А. Численные методы [Текст] : учеб. пособие [для инж.-техн. спец. вузов] / Е.А. Волков. — Изд. 5-е, стер. — СПб. : Лань, 2008. — 248 с.
2. Казакова, Н.Ф. Анализ принципиальной задачи факторизации модели отказа предоставления услуги в сети NGN на уровне управления сетью / Н.Ф. Казакова, В.И. Гура // Информационная безопасность. — 2010. — № 1(3). — С. 127–130.
3. Сергеев, В.В. Процедура получения функций изменения факторов, влияющих на качество предоставления услуг в мультисервисных сетях / В.В. Сергеев, А.М. Мухин // Информационная безопасность. — 2010. — № 2(4). — С. 139–145.
4. Альтшуллер, Г.С. Найти идею [Текст] : введ. в теорию решения изобрет. задач / Г.С. Альтшуллер; отв.ред. А.К. Дюнин; АН СССР, Сиб. отд-ние. — 2-е изд., доп. — Новосибирск : Наука. Сиб. отд-ние, 1991. — 223 с.
5. Робастность в статистике [Текст] : поход на основе функций влияния / Ф. Хампель, Э. Рончетти [и др.] ; пер. с англ. под ред. В.М. Золотарева. — М. : Мир, 1989. — 512 с.

АЛГОРИТМ ТОЧКОВОГО ПРОГНОЗУВАННЯ ВИПАДКОВИХ ПРОЦЕСІВ В АВІАЦІЙНИХ ІНФОКОМУНІКАЦІЙНИХ МЕРЕЖАХ

Д.В. Чирков, В.Г. Липовський

Національний авіаційний університет,
просп. Космонавта Комарова, 1, Київ, 03680, Україна

У роботі розглянута загальна методика прогнозуючого контролю сучасних мультисервісних мереж, яка в якості основного компоненту включає алгоритми структурної ідентифікації моделей регулярних складових (трендів) процесів. Наведена загальна структура методики прогнозуючого контролю авіаційних інфокомунікаційних систем. Представлений загальний вигляд алгоритму структурної ідентифікації.

Ключові слова: інфокомунікаційні мережі, авіаційні мережі, випадкові процеси, точкове прогнозування, алгоритм прогнозування, мультисервісні мережі

POINT PREDICTION ALGORITHM FOR STOCHASTIC PROCESSES IN AIRCRAFT COMMUNICATION NETWORKS

Dmytro V. Chirkov, Valery G. Lipovsky

National Aviation University,
1 Kosmonavta Komarova Ave., Kyiv, 03680, Ukraine

The paper focuses on general method of predictive control of modern multi-service networks, which is a major component includes algorithms for the identification of structural models of regular components (trends) of the observed processes. The general structure for predictive control of aircraft communication systems is described. The general view of the structural identification algorithm is discussed.

Keywords: information and communication networks, airline networks, stochastic processes, point forecasting, prediction algorithm, multi-service networks