

Кондратов Дмитрий Александрович

Магистрант

Направление: Информатика и вычислительная техника

Магистерская программа: Информационные системы

**Экономически обоснованный выбор порога упреждающей реконфигурации
распределенной автоматизированной системы по прогнозу состояния узлов**

Аннотация. В статье рассматривается задача выбора порога принятия решения Θ в системе прогнозирования технического состояния узлов распределенной автоматизированной системы, по которому запускается упреждающая миграция нагрузки. Показано, что критерии F-меры и максимума точности здесь неприменимы: цена ошибок первого и второго рода различается более чем в 300 раз, а влияние порога на коэффициент готовности проявляется лишь в шестом знаке после запятой.

Ключевые слова: упреждающая реконфигурация, порог принятия решения, прогнозирование отказов, предиктивный мониторинг, коэффициент готовности, время восстановления, распределенные автоматизированные системы, цена ошибок классификации.

Непрерывность управления в распределенных автоматизированных системах (РАС), отнесенных к значимым объектам критической информационной инфраструктуры [6], определяется не столько частотой отказов узлов, сколько длительностью перерыва управления после каждого отказа. При реактивной схеме этот перерыв складывается из обнаружения по таймауту, переизбрания ведущего узла, переноса нагрузки и стабилизации и для трехузлового кластера «2 из 3» составляет 60,00 с. Штатные контуры самовосстановления сред оркестрации позволяют перенести нагрузку заранее – по прогнозу состояния узла, – сокращая перерыв до 8,00 с [4]. Однако

выигрыш реализуется только при верно настроенном пороге: низкий порог порождает поток ложных миграций, высокий — пропускает предотказные состояния.

Задача настройки порога в системах предиктивного обслуживания обычно решается по F-мере или по адаптивным статистическим правилам [3; 8], тогда как работы по упреждающему управлению ресурсами показывают, что выигрыш возникает лишь при согласовании условий срабатывания с ценой миграции [7].

Цель: обосновать критерий выбора порога упреждающей реконфигурации, исходящий не из метрик классификации, а из соотношения стоимости неупрежденного отказа и стоимости плановой миграции.

Архитектура контура упреждающей реконфигурации

Контур строится как шестиуровневая конвейерная структура: сбор телеметрии с периодом 10 с, гарантированная передача, хранение временных рядов, выделение признаков, прогнозирование и принятие решения. Состояние узла в момент t описывается вектором агрегатов по скользящему окну длительностью W :

$$x_t = (x_t^1, x_t^2, \dots, x_t^n)^T, \quad x_t^k = g^k(\{m(\tau) : \tau \in [t - W, t]\}), \quad (1)$$

где $m(\tau)$ – сырые отсчеты телеметрии; g^k – агрегирующая функция; $W = 15$ мин. Использованы 12 диагностических признаков, восемь из которых относятся к программному, сетевому и ресурсному слою: более 60% нарушений вызваны программным обеспечением, конфигурациями и истощением ресурсов [3; 4]. Решающее правило задается пороговым сравнением:

$$p_t = f(x_t) \in [0, 1], \quad d_t = 1[p_t \geq \Theta], \quad (2)$$

где f – обученная модель прогнозирования; $d_t = 1$ означает запуск упреждающей миграции нагрузки на заведомо работоспособный узел средствами оркестратора без участия оператора. Адаптация такого контура к дрейфу концепции в многомерных потоках исследована в [8], где достигнут AUC 89,71% при сохранении масштабируемости; для промышленного оборудования аналогичный конвейер строится на базе цифровых двойников [5].

Асимметрия цены ошибок распознавания

Качество решающего правила (2) оценивается парой условных вероятностей $TPR(\Theta) = P\{d = 1 \mid \text{предотказ}\}$ и $FPR(\Theta) = P\{d = 1 \mid \text{норма}\}$. На контрольной выборке из 2000 окон наблюдения, из которых 120 предшествуют зарегистрированным отказам или неплановым перезапускам служб, при $\Theta = 0,50$ получены $TPR = 0,85$ и $FPR = 0,05$ (таблица 1).

Таблица 1

Матрица ошибок и метрики качества прогнозирования при $\Theta = 0,50$

Показатель	Значение	Показатель	Значение
Верно распознанные предотказы (TP)	102	Precision	0,520
Пропуски (FN)	18	TPR (Recall)	0,850
Ложные срабатывания (FP)	94	FPR	0,050
Верные отрицания (TN)	1786	F ₁ -мера	0,645

Точность $Precision = 0,520$ означает, что примерно каждое второе срабатывание ложно. Для классической диагностики это неудовлетворительно, однако цена ошибок асимметрична: пропуск предотвращения оборачивается реактивной реконфигурацией длительностью 60,00 с с ущербом $s_{ав} = 180$ тыс. руб., а ложное срабатывание – плановой миграцией стоимостью $s_{мигр} = 0,5$ тыс. руб. без прерывания управления. Отношение цен превышает 300, поэтому симметричные по своей природе критерии F_1 и максимума точности для настройки порога непригодны.

Влияние порога на время восстановления и готовность

Порог влияет на эффективное время восстановления управления через долю упрежденных отказов:

$$\tau_{p}^{ef}(\Theta) = TPR(\Theta) \cdot \tau_{пред} + [1 - TPR(\Theta)] \cdot \tau_{p}, \quad (3)$$

где $\tau_{пред} = 8,00$ с – время плановой миграции на прогретый экземпляр службы; $\tau_p = 60,00$ с – время реактивного восстановления. Введение состояния реконфигурации в граф состояний трехузлового кластера с критерием «2 из 3» дает

$$U(\tau_p) = 3\lambda\tau_p + 3\rho^2 + \rho^3, \quad K_T = 1 - U(\tau_p), \quad \rho = \lambda/\mu. \quad (4)$$

При наработке на отказ 10 000 ч изменение τ во всем достижимом диапазоне от 8 до 60 с сдвигает коэффициент готовности лишь в шестом знаке: K_g растет с 0,999994970 до 0,999998653, а годовой простой сокращается с 0,0441 до 0,0118 ч. Следовательно, порог нельзя обосновать и по критериям надежности: разница лежит за пределами точности исходных данных о потоках отказов [1; 2].

Экономический критерий выбора порога

Единственным содержательным критерием остается годовой экономический эффект контура мониторинга, учитывающий одновременно предотвращенный ущерб и издержки ложных миграций:

$$E_{\text{мон}}(\Theta) = \text{TPR}(\Theta) \cdot \text{потк} \cdot (\text{сав} - \text{смигр}) - \text{FPR}(\Theta) \cdot \text{Нокон} \cdot \text{смигр} \rightarrow \max, \quad (5)$$

где $\text{потк} = 3\lambda T_g = 2,628$ отк./год – ожидаемое число отказов узлов кластера; $\text{Нокон} = 1095$ – число агрегированных решений в год по кластеру после сглаживания и подавления повторных срабатываний. Результаты расчета для трех значений порога приведены в таблице 2.

Таблица 2.

Обоснование выбора порога принятия решения

Θ	TPR	FPR	$\tau^{\text{ef}}, \text{с}$	Издержки ложных срабатываний, тыс. руб./год	Емон, тыс. руб./год
0,30	0,95	0,20	10,60	109,5	338,7
0,50	0,85	0,05	15,80	27,4	373,6
0,70	0,60	0,012	28,80	6,6	276,5

Максимум критерия (5) достигается при $\Theta^* = 0,50$. Снижение порога до 0,30 повышает чувствительность до $\text{TPR} = 0,95$, но увеличивает число ложных миграций более чем вчетверо; повышение до 0,70 сокращает издержки до 6,6 тыс. руб., но теряет треть предотказных событий и увеличивает τ^{ef} до 28,80 с. Оптимум смещается вместе с отношением $\text{сав}/\text{смигр}$: чем дороже простой, тем ниже должен быть порог. Сходный вывод получен в задачах упреждающего управления виртуальными машинами, где

срабатывание по совокупности текущего и прогнозируемого состояния сокращает число миграций на 56% и нарушений SLA на 79% [7].

Экспериментальная проверка

Аналитические оценки получены при идеализирующих допущениях, поэтому выполнена проверка методом управляемого внесения отказов: пять сценариев по 200 прогонов. Для трех сценариев расхождение расчета и эксперимента не превысило 8%, а сокращение времени восстановления составило 3,80 раза. Наибольшее отклонение (+37,3%) дал сценарий частичной сетевой изоляции: узел отвечает на часть запросов, обнаружение по таймауту срабатывает позже, а кворумная группа несколько раз пересматривает состав активных узлов [6]. Отсюда рекомендация назначать сетевым признакам повышенный вес.

Заключение

Показано, что порог принятия решения в контуре упреждающей реконфигурации не может быть обоснован ни метриками качества классификации, ни показателями надежности: первые симметричны по отношению к ошибкам, цена которых различается более чем в 300 раз, вторые изменяются лишь в шестом знаке после запятой. Критерий (5) устраняет это противоречие и дает для типовой РАС оптимум $\Theta^* = 0,50$, эффективное время восстановления 15,80 с и годовой эффект 373,6 тыс. руб. при издержках ложных срабатываний 27,4 тыс. руб.

Критерий сформулирован в величинах, доступных службе эксплуатации (стоимость простоя и миграции, поток отказов, число решений в год), и потому применим к другим объектам автоматизации без переобучения модели. Дальнейшая работа – адаптивная подстройка порога при дрейфе распределения телеметрии.

Литература

1. ГОСТ Р 27.102–2021. Надежность в технике. Надежность объекта. Термины и определения. – М.: Стандартинформ, 2021. 24 с.

2. ГОСТ Р 27.013–2019. Надежность в технике. Методы оценки показателей безотказности и ремонтпригодности. – М.: Стандартинформ, 2019. 40 с.

3. Panahandeh M., Hamou-Lhadj A., Hamdaqa M., Miller J. ServiceAnomaly: An anomaly detection approach in microservices using distributed traces and profiling metrics // Journal of Systems and Software. 2024. Vol. 209. Art. 111917. DOI 10.1016/j.jss.2023.111917.

4. Anemogiannis V., Andreou B., Myrtollari K., Panagidi K., Hadjiefthymiades S. Enhancing Kubernetes Resilience through Anomaly Detection and Prediction. 2025. arXiv:2503.14114. DOI 10.48550/arXiv.2503.14114.

5. Abdullahi I., Longo S., Samie M. Towards a Distributed Digital Twin Framework for Predictive Maintenance in Industrial Internet of Things // Sensors. 2024. Vol. 24, № 8. Art. 2663. DOI 10.3390/s24082663.

6. Федеральный закон «О безопасности критической информационной инфраструктуры Российской Федерации» от 26.07.2017 № 187-ФЗ (последняя редакция) / [Электронный ресурс] / URL: <https://www.consultant.ru/document/>

7. Maiyza A. I., Hassan H. A., Sheta W. M., Abd-Elminaam D. S., Ali H. A. VTGAN based proactive VM consolidation in cloud data centers using value and trend approaches // Scientific Reports. 2025. Vol. 15. Art. 20133. DOI 10.1038/s41598-025-04757-z.

8. Raeiszadeh M., Ebrahimzadeh A., Glitho R. H., Eker J., Mini R. A. F. Real-Time Adaptive Anomaly Detection in Industrial IoT Environments. 2026. arXiv:2601.03085. DOI 10.48550/arXiv.2601.03085.