Информационные процессы, Том 11, № 2, 2011, стр. 291–303

Метод группового поллинга при независимой активности датчиков в сетях мониторинга

Е.Е. Маликова*, И.И. Цитович**

* *Московский технический университет связи и информатики, Министерство российской федерации по связи и информатизации, Москва, Россия*
** *Институт проблем передачи информации им. А.А. Харкевича РАН, Москва, Россия*
Поступила в редколлегию 17.06.2011

Аннотация—В работе исследуются свойства группового опроса датчиков для последующего решения задачи идентификации активных датчиков. Здесь будет рассматриваться случай, когда все датчики срабатывают независимо друг от друга.

1. ВЕДЕНИЕ

Повсеместное проникновение беспроводных сетей передачи данных (БСПД), построенных на базе сетей подвижной сотовой связи стандартов GSM 2,5+, CDMA, UMTS и LTE, спутниковых систем и беспроводных вычислительных сетей стандартов WiFi и WiMAX привело к качественному скачку в построении систем мониторинга и телеметрии. Эти системы все больше проникают во все сферы человеческой деятельности, что обусловлено также успехами микроэлектроники, приведшими к созданию недорогих датчиков, передающих информацию о состоянии объектов.

В настоящее время с применением БСПД строятся системы наблюдения за чрезвычайными ситуациями, состоянием технических объектов, датчиками напряжения в электросетях, экологический мониторинг, системы наблюдения за общегородским транспортом, системы дистанционного управления датчиками учета расхода воды, газа и электричества, системы управления платежными терминалами и т.д. Повсеместно внедряются различные корпоративные сети распределенного мониторинга, например, за состоянием газо- и нефтепроводов. Наряду со стационарными развиваются и динамические системы мониторинга и телеметрии. Различные типы датчиков устанавливаются на машины службы скорой помощи, транспортных средствах таксопарков, при этом они подключаются и к системам позиционирования, и могут передавать информацию о местонахождении и состоянии объекта.

Для предотвращения чрезвычайных ситуаций в масштабах страны разрабатываются и начинают применяться системы оперативного наблюдения на крупных территориально-распределенных объектах с большим числом датчиков, удаленных на значительные расстояния от центров обработки данных. При этом предъявляются жесткие требования к оперативности и достоверности доставки информации для предупреждения критических ситуаций. Передаваемый в этих системах трафик может быть мультисервисным, т.к. наряду с данными телеметрии, передающимися, как правило, с низкими скоростями, при возникновении чрезвычайных ситуаций могут передаваться сигналы видеонаблюдений, использующие максимальные скорости, например, но каналам WiMax или спутниковым каналам.

Для опроса датчиков систем мониторинга с целью выявления активных (тех, которые должны передавать информацию о чрезвычайной ситуации), применяются различные системы упорядоченного опроса элементов сети, получившие в иностранной литературе название поллинга. По математическим моделям они в общем виде сходны с приоритетными системами массового обслуживания, однако здесь приоритеты назначаются по определенному правилу. Параметры поллинга могут изменяться динамически.

Разработке и исследованию характеристик различных моделей систем БСПД посвящено множество работ отечественных и зарубежных исследователей (см., например, [1]). В большинстве работ в основном разработаны методы упорядоченного поллинга, в которых число активных датчиков сравнимо с их общим количеством. Для больших систем, состоящих из сотен и более датчиков, такой подход не является эффективным, поскольку время опроса будет пропорционально числу датчиков и будет слишком большим, а резервируемый для этого ресурс будет использован неэффективно, поскольку вероятность того, что какой-то датчик активен, мала. В связи с этим необходимо разработать методы опроса, основанные на групповом опросе, когда несколько датчиков одновременно на одной полосе передают сигнальную информацию, и тем самым существенно сократить время, необходимое для выявления активных датчиков ([2]–[4]).

Математической основой для решения этой задачи послужили работы по планированию отсеивающих экспериментов. Наиболее близкой к решаемой нами задаче является задача определения предельной скорости отсеивающего плана поиска s значимых переменных двоичной функции, измеряемой со случайными ошибками, [5]. Однако непосредственно перенести полученные результаты на рассматриваемую здесь задачу обнаружения активных датчиков невозможно. Как показывают исследования, здесь необходимо сочетать элементы статистического планирования экспериментов и последовательное планирование, поскольку в текущий момент времени не известно количество активных датчиков, а от этого существенно зависят свойства процедуры, поскольку параметры алгоритма поиска должны быть настроены на определенное количество активных датчиков.

В разделе 2 приведена постановка задачи. В разделе 3 приводится описание алгоритма “пофакторного”([5]) обнаружения активных датчиков. Это означает, что ищется не все множество активных датчиков S, а каждый из его элементов отдельно. Предлагаемый подход несколько снижает эффективность метода обнаружения активных датчиков, но позволяет получить простой с вычислительной точки зрения алгоритм поиска активных датчиков.

В следующем разделе проводится исследование свойств метода пофакторного обнаружения активных датчиков. Здесь получены формулы для вычисления вероятностей неправильной идентификации датчиков, основанные на применении оценок вероятностей с помощью техники, используемой в теории больших уклонений.

На основании полученных теоретических результатов был разработан алгоритм численного моделирования задачи обнаружения активных датчиков с помощью группового опроса. Результаты расчетов в зависимости от изменения таких параметров модели, как количество датчиков в сети, вероятность искажения сигнала при передаче на диспетчерский пункт, несовпадения предполагаемого и фактического числа активных датчиков, приведены в разделе 5.

При построении матрицы опроса обычно рекомендуется подбирать ее таким образом, чтобы число различий в участии в опросе для двух различных датчиков было примерно одинаковым. Такие матрицы опроса называют сбалансированными. В этом случае алгоритм обнаружения активных датчиков должен работать более стабильно, чем при чисто случайном выборе участия конкретного датчика в текущем опросе. Влияние сбалансированности матрицы опроса на результат идентификации активных датчиков рассмотрен в разделе 6.

Завершается работа выводами и рекомендациями об организации группового опроса датчиков.

2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть имеется телеметрическая сеть, содержащая t датчиков; требуется разработать стратегию их опроса с целью скорейшего выявления s датчиков, имеющих данные для передачи. Необходимо за наименьшее число шагов определить активные датчики таким образом, чтобы средняя вероятность неправильного определения хотя бы одного из них не превосходила заданный уровень (все усреднения производятся по равномерному априорному распределению передающих датчиков на множестве $T = \{1, \ldots, t\}$).

В данной задаче предполагается, что $s << t$, что соответствует случаям, когда число активных датчиков в сети мало. Такая ситуация является наиболее типичной для телеметрической сети, которая покрывает большую территорию, а вероятность чрезвычайных ситуаций на ее локальном участке мала. В этих условиях проведение опроса всех датчиков заняло бы значительное время, поскольку каждый цикл занимал бы не менее t опросов. Цель работы состоит в том, чтобы предложить групповой метод опроса, позволяющий выявлять активные датчики за $O(\log_2(t))$ опросов, что позволит в разы сократить время обнаружения критической ситуации.

Таким образом, имеется t датчиков, состояние которых описывается переменными $x_1, \ldots, x_t$, которые принимают значения 0 или 1; значение 0 означает, что соответствующий датчик пассивен, т.е. не должен передавать информацию, значение 1 — что соответствующий датчик активен, т.е. должен передавать информацию. Среди них лишь переменные с номерами $i_1, \ldots, i_s$ принимают значения 1, а остальные равны 0. Упорядоченное множество активных датчиков обозначим через S.

Групповой опрос состоит в том, что принимается одновременно сигнал от нескольких источников. Он задается с помощью вектора $a = (a_1, \ldots, a_t)$, где a_i принимает значения 0 или 1. Значение $a_i = 1$ означает, что i-й датчик участвует в опросе, а $a_i = 0$ означает, что i-й датчик не участвует в опросе. Если N — число опросов, то все опросы задаются булевой матрицей опросов $A = (a^1, \ldots, a^N)^{\mathrm{T}}$, где $a^j = (a_1^j, \ldots, a_t^j)$ — вектор j-го опроса. Если в группе опрашиваемых датчиков имеется хотя бы один активный, то мы наблюдаем наличие сигнала, который интерпретируем как 1. Если в группе нет ни одного активного источника, то от нее не поступает ни одного сигнала, что интерпретируется как 0. Таким образом, в качестве ответа датчиков j-ой группы сформируется результат

$$f_j = (a_1^j \wedge x_1) \vee \cdots \vee (a_t^j \wedge x_t), \qquad (1)$$

где $\wedge$ — булево произведение, $\vee$ — булева сумма.

Предполагается, что в сети возможны ошибки при передаче информации. Это означает, что значение функции f_j известно с некоторой ошибкой: при каждом опросе происходит искажение результата независимо от других опросов в соответствии со стохастической матрицей переходов

$$W = \begin{pmatrix} 1-\alpha_0, & \alpha_0 \\ \alpha_1, & 1-\alpha_1 \end{pmatrix},$$

где α_0 — вероятность искажения приема 0 (т.е. наблюдается значение 1 вместо 0), а α_1 — вероятность искажения приема 1 (т.е. наблюдается значение 0 вместо 1). Поэтому результат j-го опроса будет g_j, который принимает значения 0 или 1 в соответствии с матрицей W независимо от значений в других наблюдениях при условии, что значения f_j фиксированы.

Мы ограничимся случаем $\alpha_0 + \alpha_1 < \frac{1}{2}$ для упрощения обозначений. В действительности в большинстве случаев значения α_0 и α_1 близки к 0, поскольку искажения сигналов в сети малы.

Все результаты сохраняются и в случае $\alpha_0 + \alpha_1 > \frac{1}{2}$, если в наблюденях заменить 0 на 1 и наоборот.

Подход базируется на результатах, полученных в [5], однако там рассматривалось последовательное планирование “опросов” при построении матрицы A, которое неприемлемо в нашей ситуации, так как нашем случае мы заранее должны сообщить расписание выхода на датчика на связь.

Необходимо получить алгоритмы построения матрицы опроса A и определения множества S активных датчиков на основании наблюдений $g_1, \ldots, g_N$.

3. ОПИСАНИЕ АЛГОРИТМА ПОФАКТОРНОГО ОБНАРУЖЕНИЯ АКТИВНЫХ ДАТЧИКОВ

Как следует из [5], асимптотически оптимальный (при стремлении к 0 вероятности неправильного определения активных датчиков) план опроса A получается при использовании случайной матрицы, в которой значения 1 в матрице опроса выбираются независимо друг от друга с вероятностью p_0, где $1-(1-p_0)^{s_0}$ обеспечивает достижение пропускной способности C двоичного канала с переходной матрицей W ([5]);

$$C = 1 - p^0 h(\alpha_0) - (1-p^0) h(\alpha_1), \tag{2}$$

где

$$h(\alpha) = -(\alpha \log(\alpha) + (1-\alpha)\log(1-\alpha)), 0 < \alpha < 1,$$

$$p^0 = 1 - \sqrt[s_0]{\frac{\frac{1}{2} - \alpha_0}{1 - \alpha_0 - \alpha_1}}.$$

Здесь s_0 — предполагаемое число активных датчиков.

Поскольку ни само множество S, ни количество датчиков s не известны, то при описании алгоритма используются обозначения: s_0 — предполагаемое число активных датчиков, s_r — фактическое число активных датчиков, $\hat{s}$ — оценка числа активных датчиков

Для упрощения процедуры идентификации решение об активности конкретного датчика принимается на основе “пофакторного анализа” с использованием метода максимального правдоподобия. Для принятия решения о том, является датчик активным или нет, используются следующие данные: $x_{00}(i)$ — количество наблюдений, когда i-й датчик не опрашивался и результат опроса оказался $g = 0$, $x_{10}(i)$ — количество наблюдений, когда i-й датчик опрашивался, а результат опроса $g = 0$, $x_{01}(i)$ — количество наблюдений, когда i-й датчик не опрашивался, результат опроса $g = 1$ и $x_{11}(i)$ — количество наблюдений, когда i-й датчик опрашивался и результат опроса $g = 1$. Поскольку предполагается, что вероятности ошибок при передаче информации малы, то результаты наблюдений, когда датчик был активным, а результат наблюдения $g = 0$ говорит в пользу того, что датчик неактивен, поэтому относительно большое значение $x_{10}(i)$ говорит о том, что i-ый датчик неактивен. Если бы искажения результатов наблюдений не было, то положительное значение $x_{10}(i)$ однозначно указывало бы на то, что датчик неактивен. Если датчик активен, то вероятность такой ситуации равна α_1, поскольку значение функции f равно 1 и для получения результата $g = 0$ необходимо, чтобы произошло искажение сигнала, что происходит по сделанным предположениям с вероятностью α_1. Если же датчик был неактивным, то к такому результату приводят два варианта исходов: ни один активный датчик не попал в опрос, что происходит с вероятностью $(1-p_0)^{s_r}$, и результат $f = 0$ не исказился, что происходит с вероятностью $1 - \alpha_0$, или же хотя бы один активный датчик попал в опрос, что происходит с вероятностью $1-(1-p_0)^{s_r}$, и результат $f = 1$ исказился, что происходит с вероятностью α_1. Если p^* — вероятность того, что f равно 1, то $p^* = 1-(1-p_0)^{s_r}$, а вероятность события, что неактивный датчик участвовал в опросе и результат опроса 0, равна $(1-p^*)(1-\alpha_0) + p^*\alpha_1 = 1 - \alpha_0 - p^*(1-\alpha_0-\alpha_1)$. Вклад такого результата в логарифм отношения правдоподобия составит величину

$$a_{10} = \log \frac{\alpha_1}{1 - \alpha_0 - p^*(1 - \alpha_0 - \alpha_1)}. \tag{3}$$

Значение a_{10} — отрицательная величина, что и говорит о том, что результат наблюдения, когда датчик участвовал в опросе и результат опроса 0, говорит в пользу неактивности датчика.

Если датчик участвовал в опросе и результат опроса 1, то это говорит в пользу того, что датчик активен. Количество таких результатов $x_{11}(i)$, а их вес

$$a_{11} = \log \frac{1 - \alpha_1}{\alpha_0 + p^*(1 - \alpha_0 - \alpha_1)}. \tag{4}$$

Действительно, если датчик активен, то вероятность такой ситуации равна $1 - \alpha_1$, поскольку значение функции f равно 1 и для получения результата $g = 1$ необходимо, чтобы не произошло искажение сигнала, что происходит по сделанным предположениям с вероятностью $1 - \alpha_1$. Если же датчик был неактивным, то к такому результату приводят два варианта исходов: ни один активный датчик не попал в опрос, что происходит с вероятностью $(1 - p_0)^{s_r}$, и результат $f = 0$ исказился, что происходит с вероятностью α_0, или же хотя бы один активный датчик попал в опрос, что происходит с вероятностью $1 - (1 - p_0)^{s_r}$, и результат $f = 1$ не исказился, что происходит с вероятностью $1 - \alpha_1$. Суммарная вероятность равна $(1 - p^*)\alpha_0 + p^*(1 - \alpha_1) = \alpha_0 + p^*(1 - \alpha_0 - \alpha_1)$. Значение a_{10} — положительная величина, что и говорит о том, что результат наблюдения, когда датчик участвовал в опросе и результат опроса 1, говорит в пользу активности датчика.

Результаты наблюдений, когда датчик не участвовал в опросе, менее информативны: их веса отличаются от 0 незначительно. Если датчик не участвовал в опросе и результат опроса 1, то количество таких результатов $x_{01}(i)$, а их вес

$$a_{01} = \log \frac{\hat{p}(1 - \alpha_0 - \alpha_1) + \alpha_0}{p^*(1 - \alpha_0 - \alpha_1) + \alpha_0}. \tag{5}$$

Действительно, если датчик активен, то к такому результату приводят два варианта исходов: ни один из оставшихся активных датчиков не попал в опрос, что происходит с вероятностью $(1 - p_0)^{s_r - 1}$ (здесь показатель степени $s_r - 1$, поскольку теперь осталось активных датчиков на 1 меньше), и результат $f = 0$ исказился, что происходит с вероятностью α_0, или же хотя бы один активный датчик попал в опрос, что происходит с вероятностью $\hat{p} = 1 - (1 - p_0)^{s_r - 1}$, и результат $f = 1$ не исказился, что происходит с вероятностью $1 - \alpha_1$. Суммарная вероятность равна $(1 - \hat{p})\alpha_0 + \hat{p}(1 - \alpha_1) = \alpha_0 + \hat{p}(1 - \alpha_0 - \alpha_1)$. Если же датчик был неактивным, то вероятность будет вычисляться как и в предыдущем случае и равна $(1 - p^*)\alpha_0 + p^*(1 - \alpha_1) = \alpha_0 + p^*(1 - \alpha_0 - \alpha_1)$.

Если датчик не участвовал в опросе и результат опроса 0, то количество таких результатов $x_{00}(i)$, а их вес

$$a_{00} = \log \frac{1 - \alpha_0 - \hat{p}(1 - \alpha_0 - \alpha_1)}{1 - \alpha_0 - p^*(1 - \alpha_0 - \alpha_1)}. \tag{6}$$

Действительно, если датчик активен, то к такому результату приводят два варианта исходов: ни один из оставшихся активных датчиков не попал в опрос, что происходит с вероятностью $1 - \hat{p}$, и результат $f = 0$ не исказился, что происходит с вероятностью $1 - \alpha_0$, или же хотя бы один активный датчик попал в опрос, что происходит с вероятностью $\hat{p}$, и результат $f = 1$ исказился, что происходит с вероятностью α_1. Суммарная вероятность равна $1 - \alpha_0 - \hat{p}(1 - \alpha_0 - \alpha_1)$. Если же датчик был неактивным, то вероятность будет вычисляться как и в соответствующем случае, когда датчик участвовал в опросе и результат опроса был 0, и равна $1 - \alpha_0 - p^*(1 - \alpha_0 - \alpha_1)$.

Пусть β_0 — вероятность ошибки первого рода или вероятность пропуска цели. Она означает вероятность события, состоящего в том, что активный датчик будет идентифицирован как неактивный. В этом случае будет пропущен момент возникновения внештатной ситуации. Поскольку всего таких датчиков s, то вероятность пропуска по крайней мере одного активного датчика будет $P_1 \approx s\beta_0$. Эта вероятность должна быть максимально малой и должна задаваться как основной параметр, характеризующий качество решающего правила определения активных датчиков. Поскольку величина s мала, то величина ошибки первого рода при идентификации активного датчика при "пофакторном анализе" должна быть сравнима с P_1.

Пусть β_1 — вероятность ошибки второго рода или вероятность ложной тревоги. Она означает вероятность события, состоящего в том, что неактивный датчик будет идентифицирован как активный. В этом случае будет принято решение о приеме информации с данного датчика. Анализ информации, поступающей с датчика, позволит определить, что в данном случае никакая внештатная ситуация не наблюдается. Последствия такого решения будут состоять лишь в том, что будет ошибочно активизирован канал связи с этим датчиком. Поскольку в данном случае последствия ошибочного решения не столь существенны, как в предыдущем случае, то требования к такого рода ошибкам более слабые.

Однако, поскольку всего таких датчиков $t - s$, где $t \gg s$, то для среднего числа ошибочно обнаруженных активных датчиков M_1 получаем $M_1 \leq t\beta_1$. Параметр M_1 должен задаваться как основной параметр, характеризующий качество решающего правила определения активных датчиков. Как указывалось выше $M_1 = O(1)$, и, конечно, желательно, чтобы $M_1 < 1$. Следовательно,

$$\beta_1 < \frac{M_1}{t}. \tag{7}$$

Поскольку предполагается, что количество датчиков велико, то вероятность ошибки второго рода β_1 должна быть малой и для больших систем она окажется существенно меньше ошибки первого рода.

Таким образом, после проведения N опросов для каждого из t датчиков вычисляются величины $x_{00}(i)$, $x_{01}(i)$, $x_{10}(i)$ и $x_{11}(i)$, а на их основании логарифмы отношения правдоподобия

$$L(i) = a_{00}x_{00}(i) + a_{10}x_{10}(i) + a_{01}x_{01}(i) + a_{11}x_{11}(i). \tag{8}$$

На основании значения ошибки первого рода вычисляется значение порога L_0, при превышении которого логарифмом отношения правдоподобия $L(i)$ принимается решение, что i-й датчик является активным.

Алгоритм поиска активных датчиков выгладит следующим образом.

1. Исходными параметрами являются:

t — количество датчиков;

s_0 — количество предполагаемых активных датчиков;

α_0 и α_1 — вероятности ошибок искажения результата опроса;

N — количество опросов;

A — матрица опросов, которая формируется как матрица со случайными элементами 0 и 1, в которой N строк и t столбцов;

L_0 — порог принятия решения.

2. На основании результатов опросов $g_1, \ldots, g_N$ вычисляются величины $L(i)$ по формулам (8), (3)–(6), в которых величина s_r заменена на s_0.

3. Производится сравнение логарифмом отношения правдоподобия $L(i)$ с порогом L_0 для каждого i; при превышении порога для некоторого $L(\hat{i})$ считается, что $\hat{i}$-й датчик является активным, т.е. должен передавать информацию о внештатной ситуации.

4. Выходными параметрами алгоритма будут величины $\hat{i}_1, \ldots, \hat{i}_{\hat{s}}$, где $\hat{s}$ — количество обнаруженных активных датчиков, $\hat{i}_k$ — номера обнаруженных датчиков.

5. Характеристиками качества алгоритма являются вероятности правильной идентификации активных датчиков: пусть $\hat{S} = \{\hat{i}_1, \ldots, \hat{i}_{\hat{s}}\}$, тогда P_1 — вероятность того, что будет пропущен активный датчик, т.е. $S \nsubseteq \hat{S}$ (обнаружение лишних активных датчиков не является ошибкой) и P_2 — вероятность неправильного определения активности хотя бы одного датчика, т.е. $S \neq \hat{S}$. Вероятности вычисляются в предположении, что на множестве возможных значений S задано равномерное распределение.

Выбор характеристики P_1 обусловлен тем, что при объявлении активным датчика, не являющегося активным, не будет тяжелых последствий, поскольку при анализе информации, поступающей с датчика, будет обнаружено, что он неактивный. В этом случае ущерб состоит только в излишнем использовании ресурсов сети передачи информации с датчика. Поэтому при малых значениях P_1 и P_2 можно считать, что алгоритм дает удовлетворительное качество обнаружения активных датчиков, только он может приводить к появлению ложной активности датчиков.

Если $L(i) > L_0$, то i-й датчик считается активным. Порог $L_0 > 0$ выбирается таким образом, чтобы обеспечить заданную ошибку первого рода. Чем выше значение порога, тем меньше вероятность ошибки неправильной идентификации неактивного датчика, но, вместе с тем, возрастает вероятность пропуска активного датчика. Далее исследуются свойства рассматриваемого алгоритма для определения правила выбора порога принятия решения об активном датчике.

4. СВОЙСТВА АЛГОРИТМА

Из [5] следует, что нижняя граница для требуемого числа опросов $N_0 = s_r \frac{\log_2(t)}{C}$. В рассматриваемой ситуации поиска активных датчиков значение s_r не известно, а используется априорное предположение о его значении s_0. Поэтому важно установить, как ошибка в определении s_r влияет на качество предложенного алгоритма.

Если справедливо предположение, что i-й датчик активен, то из полученных при выводе формул (3)–(6) вероятностей следует, что

$$\begin{aligned}\mathsf{E}_{s_r} L(i) = N\Big(& a_{00}(1-p_0)(1-\alpha_0-\hat{p}(1-\alpha_0-\alpha_1)) + a_{10}p_0\alpha_1 + \\ &+ a_{01}(1-p_0)(\hat{p}(1-\alpha_0-\alpha_1)+\alpha_0) + a_{11}p_0(1-\alpha_1)\Big). \end{aligned} \tag{9}$$

Если же справедливо предположение, что i-й датчик неактивен, то

$$\begin{aligned}\mathsf{E}_{s_r} L(i) = N\Big(& a_{00}(1-p_0)(1-\alpha_0-p^*(1-\alpha_0-\alpha_1)) + \\ &+ a_{10}p_0(1-\alpha_0-p^*(1-\alpha_0-\alpha_1)) + a_{01}(1-p_0)(p^*(1-\alpha_0-\alpha_1)+\alpha_0) + \\ &+ a_{11}p_0(\alpha_0+p^*(1-\alpha_0-\alpha_1))\Big). \end{aligned} \tag{10}$$

Здесь знак E_{s_r} обозначает среднее значение, вычисленное в предположении, что имеется равномерное распределение на множестве возможных значений номеров активных датчиков, если их количество равно s_r, поэтому $\hat{p}$ p^* вычисляются в предположении, что $s_r = s_0$, а p_0 в соответствии со значением s_r.

Отметим, что в (9) величина $C_1(s_r, s_0) = \frac{\mathsf{E}_{s_r} L(i)}{N}$ положительная, а в (10) величина $C_2(s_r, s_0) = \frac{\mathsf{E}_{s_r} L(i)}{N}$ отрицательная. Из (9) следует, что должно выполняться ограничение на значение порога L_0:

$$L_0 < aN, \tag{11}$$

где a должно удовлетворять условию

$$a < C_1(s_r, s_0). \tag{12}$$

В противном случае, вероятность ошибки первого рода будет большой, поскольку порог будет установлен настолько высоко, что вероятность его достижения статистикой L_0 будет мала. Если выполняется условие (12), то при $N \to \infty$ вероятность ошибки первого рода будет стремиться к 0 экспоненциально быстро.

Необходимо, однако, заметить, что величина $C_1(s_r, s_0)$ не известна, поскольку не известно значение s_r. В связи с этим для параметра a нужно использовать другую границу

$$a < \min_{s_r} C_1(s_r, s_0). \tag{13}$$

Здесь минимум берется по такому множеству значений s_r, на которые рассчитан алгоритм поиска активных датчиков. Таким образом, возникает ограничение на допустимую область значений количества активных датчиков, если выбран некоторый порог L_0 и матрица опросов настроена на число активных датчиков s_0. Последнее обстоятельство приводит к тому, что в алгоритм поиска активных датчиков должен быть адаптивным. После каждого опроса необходимо оценивать s_r. Если $\hat{s}$ выходит за пределы множества, по которому искался минимум в (13), то нужно далее использовать алгоритм с новым значением s_0. Если $\hat{s}$ находится левее множества, по которому искался минимум в (13), то нужно уменьшить значение s_0, а если правее — то увеличить. Поскольку переключение режима опросов требует передачи на все датчики соответствующего управляющего сигнала, то желательно такие переключения делать редко; в качестве рекомендации можно предложить алгоритм постепенного удвоения значения s_0, если s_0 нужно увеличивать, или уменьшение в 2 раза, если нужно уменьшать.

Для вычисления вероятности второго рода нужно использовать большие уклонения. Пусть

$$\varphi(z) = \mathsf{E}_{s_r} \exp(zL(i)). \tag{14}$$

Поскольку все опросы проводятся независимо, то как в (10) для неактивного i-го датчика

$$\begin{aligned}\ln\varphi(z) = N\ln\Big(&\exp(za^0_{00})(1-p_0)(1-\alpha_0-p^*(1-\alpha_0-\alpha_1))+\\ &+\exp(za^0_{10})p_0(1-\alpha_0-p^*(1-\alpha_0-\alpha_1))+\\ &+\exp(za^0_{01})(1-p_0)(p^*(1-\alpha_0-\alpha_1)+\alpha_0)+\\ &+\exp(za^0_{11})p_0(\alpha_0+p^*(1-\alpha_0-\alpha_1))\Big).\end{aligned} \tag{15}$$

Из (14) следует, что

$$\mathsf{P}_{s_r}(L(i) > L_0) \le \frac{\varphi(z)}{\exp(zaN)},$$

т.е.

$$\mathsf{P}_{s_r}(L(i) > L_0) \le \exp(-k(a)N), \tag{16}$$

где

$$k(a) = \max_{z>0} \left(za - \frac{\ln \varphi(z)}{N} \right). \tag{17}$$

Как и в случае с границей (12), величина $k(a)$ зависит от значения s_r, которое не известно, поэтому нужно использовать $\min_{s_r} k(a)$.

Таким образом, неопределенности, связанные с тем обстоятельством, что значение s_0 не известно, приводят к тому, что при создании матрицы опросов A используется избыточное число опросов $N = \frac{s_0 \log_2(t)}{C}(1+\psi)$, где параметр $\psi > 0$ задает величину превышения числа опросов над минимально допустимым при сделанном предположении о числе активных датчиков.

5. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ

Необходимо получить алгоритмы построения матрицы опроса A и определения множества S активных датчиков на основании наблюдений $g_1, \ldots, g_N$.

Проводилось исследование эффективности предлагаемого метода группового поллинга в зависимости от изменения различных параметров модели: роста числа факторов t, величины ошибки в канале передачи данных $\alpha_0 = \alpha_1$, числа активных датчиков s_r. При моделировании предполагалось, что любой из датчиков может быть активным независимо от состояния других датчиков с вероятностью $p = \frac{s_r}{t}$. Для каждого варианта исходных данных проводилось 1000 численных экспериментов, в которых всякий раз изменялся и случайный план опроса A. В качестве параметров качества алгоритма обнаружения активных датчиков использовались две характеристики: P_1 — вероятность того, что будет пропущен активный датчик (обнаружение лишних активных датчиков не является ошибкой) и P_2 — вероятность неправильного определения активности хотя бы одного датчика, т.е. вероятность правильного решения задачи. Выбор характеристики P_1 обусловлен тем, что при объявлении активным датчика, не являющегося активным, не будет тяжелых последствий, поскольку при анализе информации, поступающей с датчика, будет обнаружено, что он неактивный. В этом случае ущерб состоит только в излишнем использовании ресурсов сети передачи информации с датчика. Поэтому при малых значениях P_1 и P_2 можно считать, что алгоритм дает удовлетворительное качество обнаружения активных датчиков, только он может приводить к появлению ложной активности датчиков. Кроме того, вычислялась величина $\hat{s}$ — среднее число обнаруженных активных датчиков.

В зависимости от значения s_0 рассчитывалось число опросов $N = 4\frac{s_0 \log_2(t)}{C}$, т.е. параметр избыточности числа опросов $\psi = 3$. В первых двух случаях предполагалось, что число активных датчиков s_r совпадает с числом предполагаемых активных датчиков s_0.

В таб. 1 приведены результаты расчетов при различных значениях числа датчиков в сети t.

Таблица 1.

Результаты вычислений при $\alpha_0 = \alpha_1 = 0,01$, $s_r = 2$, $s_0 = 2$.

t	N	P_1	P_2	$\hat{s}$
100	56	0,00	0,09	2,10
200	64	0,03	0,08	2,03
300	68	0,01	0,06	2,05
400	72	0,02	0,07	2,04
500	76	0,01	0,03	2,03

Полученные данные показывают, что при росте числа датчиков в сети эффективность метода группового поллинга возрастает, при этом вероятность ошибки падает. Последнее обстоятельство связано с тем, что при росте N возрастает эффект избыточности числа наблюдений, задаваемый параметром ψ.

Из приведенных данных при $tP0$ видно, что индивидуальный опрос всех датчиков дает сравнимые значения по количеству опросов, однако при индивидуальном опросе из-за наличия искажений в канале передачи информации $\hat{s} = 3$. Если при использовании группового опроса только в 10% случаев будет обнаружен лишний датчик, то в при индивидуальном опросе это будет примерно в трети опросов, а в некоторых случаях будет обнаружено 2, 3 и более ошибочно активных датчиков.

Из приведенных данных при $tP0$ видно, что при групповом опросе число опросов уменьшается более чем в 6 раз, а точность решения становится выше. Дополнительно отметим, что при индивидуальном опросе среднее число ошибочно определенных датчиков уже будет 5, а не 0,03. С ростом t указанный эффект только возрастает, поскольку для уменьшения вероятности ложной тревоги при приведенных данных $\alpha_0 = \alpha_1$ следует использовать повторный опрос идентифицированных активных датчиков, поскольку число неправильно идентифицированных датчиков растет линейно по t.

Далее в таб. 2 приведены результаты исследования влияния величины ошибки в канале передачи данных $\alpha_0 = \alpha_1$. Отметим, что с ростом ошибок передачи данных растет и вероятность неправильного определения активных датчиков при использовании индивидуального опроса; для тех данных, при которых производились расчеты, вероятность P_1 будет практически равна 1, а $\hat{s}$ будет изменяться от 7 до 27, изменяясь линейно с ростом α_0.

Приведенные в таб. 2 данные показывают устойчивость предлагаемого метода группового поллинга относительно вероятностей ошибок при передаче данных. Так, например, рост необходимого числа опросов растет относительно медленно, а вероятность пропуска активных датчиков практически нулевая даже при 5% уровне ошибок. Устойчивой является и величина $\hat{s}$. Полученные результаты позволяют сделать вывод, что метод группового опроса является более устойчивым к изменениям вероятностей ошибок при передаче информации, чем метод индивидуального опроса.

Таблица 2.

Результаты вычислений при $tP0$, $s_r = 2$, $s_0 = 2$.

α_0	N	P_1	P_2	$\hat{s}$
0,01	76	0,00	0,06	2,06
0,02	80	0,01	0,06	2,06
0,03	88	0,02	0,09	2,06
0,04	92	0,02	0,04	2,01
0,05	100	0,01	0,11	2,25

Во всех предыдущих случаях предполагалось, что предполагаемое число активных датчиков s_0 совпадает с фактическим числом s_r. В действительности значение s_r не известно и может отличаться от параметра s_0, который задает план опросов A. Поэтому важно знать, насколько устойчиво работает алгоритм идентификации активных датчиков при несовпадении s_0 и s_r. Результаты исследования влияния числа фактически активных датчиков приведены в таб. 3.

Анализ представленных данных показывает, что при $s_0 > s_r$ результаты обнаружения активных датчиков являются удовлетворительными. Это связано с тем, что при этом используемый план опроса не является оптимальным, поэтому опросы не предоставляют максимально возможную информацию об активных датчиках. Однако, поскольку в этом случае используется более избыточное число опросов, то проведенных опросов достаточно для достаточно правильной идентификации активных датчиков. Единственный недостаток опроса в этом случае состоит в том, что используется избыточное количество опросов.

Таблица 3.

Результаты вычислений при $\alpha_0 = \alpha_1 = 0,01$, $tP0$, $s_0 = 2$, Nv.

s_r	P_1	P_2	$\hat{s}$
1	0,02	0,02	0,98
2	0,01	0,05	2,06
3	0,01	0,59	6,52
4	0,02	0,99	35,7
5	0,04	1.00	128

Ситуация принципиально меняется при $s_0 < s_r$. В этом случае так же используемый план опроса не является оптимальным, но требуемое количество опросов теперь должно быть больше, чем количество проведенных. Поэтому получаемой информации не достаточно для правильной идентификации активных датчиков. Приведенные результаты показывают, что вероятность пропуска активных датчиков остается малой, однако количество предполагаемых активных датчиков стремительно нарастает. Отсюда следует вывод, что при нарастании числа обнаруживаемых активных датчиков больше предполагаемого следует переходить к опросу с удвоенным значением s_0, поскольку большое значение числа обнаруженных активных датчиков является, скорее всего, результатом неправильно выбранного значения s_0.

Обнаруженный эффект показывает, что применение группового поллинга возможно лишь при условии возможности изменения значения параметра s_0 и, следовательно, перехода на другой план опроса в зависимости от состояния сети мониторинга.

6. СБАЛАНСИРОВАННЫЕ МАТРИЦЫ ОПРОСА

Полученные в предыдущем разделе оценки вероятности неправильного обнаружения активных датчиков основаны на предположении, что все опросы независимы и одинаково устроены с точки зрения выбора группы датчиков, которые будут участвовать в очередном опросе. В действительности ошибка в определении активного датчика связана с тем, что какой-то неактивный датчик участвует практически в тех же опросах, что и один из активных. Если бы они участвовали или не участвовали бы во всех опросах одинаково, то они оказались бы неразличимы, т.е. или оба были признаны активными или неактивными. Поэтому нужно стремиться к тому, чтобы различие в участии в опросах для различных датчиков было как можно больше. В качестве такого параметра рассматривается

$$d = \min_{i,j,\, i\neq j} \#\{a_i^k \neq a_j^k\}, \tag{18}$$

где # обозначает количество элементов в множестве.

Далее приведены результаты имитационного моделирования: сравнивался результат расчетов, полученных с использованием чисто случайной матрицы опросов, и сбалансированной матрицы опросов, которая выбиралась из 100 случайных матриц как матрица с наибольшим значением параметра d. Вычисления проводились при тех же параметрах, что и предыдущем разделе, только избыточность при выборе числа опросов была уменьшена. Вычислялись характеристики P_1, P_2 и $\hat{s}$ как для стандартного способа обнаружения активных датчиков, так и с использованием сбалансированной матрицы (в последнем случае характеристики снабжены дополнительным индексом b).

В таб. 4 приведены результаты расчетов при различных значениях числа датчиков в сети t.

Таблица 4.

Результаты вычислений при $\alpha_0 = \alpha_1 = 0,01$, $s_r = 2$, $s_0 = 2$.

t	N	d	P_2	P_2^b	P_1	P_1^b	$\hat{s}$	$\hat{s}^b$
100	42	8	0,25	0,09	0,06	0,00	2,16	2,10
200	48	8	0,14	0,13	0,00	0,01	2,14	2,11
300	51	9	0,22	0,18	0,02	0,01	2,23	2,19
400	54	9	0,15	0,19	0,01	0,01	2,15	2,18
500	57	9	0,12	0,09	0,00	0,00	2,12	2,10

Приведенные результаты показывают, что положительный эффект достигается лишь при небольшой продолжительности опроса, поскольку при росте продолжительности опроса сбалансированность матрицы перестает быть существенным фактором при анализе результатов опроса. Связано это с тем, что при возрастании длины матрицы опросов ее сбалансированность возрастает автоматически за счет случайного выбора 0 или 1, что объясняется действием закона больших чисел.

7. ВЫВОДЫ

1. Для современных сетей мониторинга характерно большое количество датчиков информации, причем во многих случаях вероятность передачи полезной информации конкретным датчиком очень мала. В этом случае предоставлять датчикам постоянный канал связи с центральным диспетчерским пунктом экономически нецелесообразно.
2. Если число датчиков сети превосходит несколько сотен, то применение циклических опросов или других способов индивидуального опроса датчиков требует излишних ресурсов и может быть весьма продолжительным по времени.
3. Метод группового опроса датчиков оправдан для сетей мониторинга большой емкости. Чем больше размер сети, тем больше эффективность метода группового поллинга.
4. Метод группового опроса датчиков устойчив к вероятностям искажения сигнала в сетях передачи данных, включая потерю единичных пакетов.
5. Метод группового опроса датчиков предполагает наличие нескольких режимов опроса при фиксированном числе активных датчиков.
6. Метод группового опроса датчиков не устойчив к числу активных датчиков: в случае, когда число активных датчиков больше числа предполагаемых, может происходить значительное увеличение ложных сообщений об активности датчиков. Вместе с тем, вероятность пропуска активного датчика сохраняет свойство устойчивости.
7. В случае, когда число активных датчиков меньше числа предполагаемых, метод группового опроса датчиков обеспечивает заданные параметры качества обнаружения активных датчиков, однако использует излишнее число опросов.
8. Построение сбалансированных матриц не является существенной необходимостью при реализации рассматриваемого метода обнаружения активных датчиков, поскольку изначально предполагается существенное завышение продолжительности опроса по сравнению с оптимальной, поскольку метод должен стабильно работать при колебании числа активных датчиков в заданных пределах, которые должны устанавливаться для каждого из режимов опроса.

СПИСОК ЛИТЕРАТУРЫ

1. Вишневский В.М., Семенова О.В. *Системы поллинга: теория и применение в широкополосных беспроводных сетях.* М.: Техносфера. 2007.

2. Маликова Е.Е., Цитович И.И. Стратегия группового поллинга в широкополосных беспроводных сетях мониторинга. *Обозрение прикладной и промышленной математики.* 2010. Т.17, № 2. С. 284–285.

3. Маликова Е.Е. Метод повышения пропускной способности систем телеметрии и мониторинга на базе беспроводных сетей. *T-Comm — Телекоммуникации и транспорт.* 2010. № 7. С. 37–39.

4. Маликова Е.Е. Задача оценки параметров качества сетей сбора и обработки телеметрической информации. *Труды 64 научной сессии, посвященной дню радио. Российское научно–техническое общество радиотехники, электроники и связи имени А.С. Попова.* Москва. 2009. С. 363–365.

5. Малютов М.Б. Нижние границы для средней длины последовательного планирования экспериментов. *Известия вузов. Математика.* 1983. Т. 27, №11. С. 19–41.