

## Исследование эффективности группового поллинга при зависимой 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 объектов, на которых возникла чрезвычайная ситуация. Здесь будет исследоваться влияние ситуации, когда датчики, расположенные на одном объекте, могут срабатывать одновременно.

### 1. ВЕДЕНИЕ

Проведенное в [3] исследование показало, что разрабатываемый метод группового поллинга чувствителен к реальному количеству активных датчиков  $s_r$  и плохо работает, если их предполагаемое число меньше фактического. Как правило, на технических объектах устанавливается несколько сигнальных датчиков и в случае возникновения внештатных ситуаций все датчики объекта или несколько из них могут сработать одновременно. В этом случае неопределенность с количеством активных датчиков возрастает, поскольку при возникновении внештатной ситуации на объекте сработает заранее неизвестное число датчиков  $z$ ; значение  $z$  зависит от многих неконтролируемых факторов: расположения датчиков на объекте, места возникновения внештатной ситуации, ее типа и т.п. Таким образом, на объекте может срабатывать одновременно, вообще говоря, случайное число датчиков, при этом вероятность срабатывания нескольких датчиков как реакция на одно явление существенно больше рассмотренной в [3] вероятности активности датчиков  $p_d = \frac{s_0}{B}$ , где  $s_0$  теперь — предполагаемое максимальное число внештатных ситуаций в контролируемой сети, а  $B$  — общее число объектов сети. Поэтому необходимо провести дополнительное исследование эффективности предложенного метода в ситуации, когда возможно одновременное срабатывание нескольких датчиков. Это и является задачей данной работы.

В разделе 2 приведена математическая модель задачи, принципиальное отличие которой от модели, рассмотренной в [3], состоит в том, что теперь основной единицей сети является не датчик, а объект. Далее, в разделе 3 приведена модификация алгоритма поиска активных датчиков для пофакторного обнаружения объектов с чрезвычайными ситуациями. Некоторые результаты численных исследований предложенного алгоритма приведены в разделе 4. Приведенные результаты показывают, что применение алгоритма, учитывающего структуру сети, существенно повышает эффективность алгоритма группового опроса по сравнению со стандартным алгоритмом, рассмотренным в [3]. Раздел 5 посвящен некоторым уточнениям алгоритма, связанным с тем, что структура сети может быть более неоднородной, чем предполагается в основной части работы.

## 2. МАТЕМАТИЧЕСКАЯ МОДЕЛЬ

Пусть имеется телеметрическая сеть, содержащая  $B$  объектов; на  $j$ -ом объекте установлено  $n_j$  датчиков,  $n_j \leq n^*$ ,  $j = 1, \dots, B$ . Предполагается, что  $B$  велико, а максимальное число одновременно срабатывающих датчиков при возникновении чрезвычайной ситуации  $n^*$  относительно невелико. Таким образом, имеется  $t = \sum_{j=1}^B n_j$  датчиков и требуется разработать стратегию их опроса с целью скорейшего выявления объектов  $j_1, \dots, j_r$ , на которых есть датчики, имеющие данные для передачи. По аналогии с [3] обозначим:  $S = \{j_1, \dots, j_r\}$ . В отличие от рассматриваемой в [3] задачи сейчас предполагается, что при обнаружении на объекте чрезвычайной ситуации будет сниматься информация со всех датчиков. Такой подход объясняется тем, что анализ информации со всех датчиков позволяет получить более точное представление о характере происшествия. Передача информации с датчиков, которые не обнаружили признаков происшествия, приводит лишь к некоторому дополнительному использованию каналов связи и может рассматриваться как несущественное удорожание стоимости обнаружения происшествия.

В данной задаче предполагается, что вероятность возникновения внештатных ситуаций в сети  $p_d$  не известна, однако известно, что она достаточно мала:  $p_d \leq p_d^0$ , где  $p_d^0$  — некоторая заданная вероятность, аналог величины  $p_d = \frac{s_0}{t}$ , используемой в [3]. Вместе с тем, использование в алгоритме управления опросами верхней границы этой вероятности для расчета допустимого числа активных датчиков не эффективно, поскольку, как показано в [3], используемый метод группового поллинга чувствителен к предполагаемому количеству датчиков. Это связано с тем, что при неправильном задании предполагаемого числа активных датчиков значительно падает информативность данных опроса. Такая ситуация является наиболее типичной для телеметрической сети, которая покрывает большую территорию, а вероятность чрезвычайных ситуаций на ее локальном участке мала ([2], [1]).

Таким образом, предполагается, что при возникновении внештатных ситуаций на  $j$ -ом объекте одновременно срабатывает  $z_j$  датчиков, где  $z_j$  — дискретная случайная величина с распределением на множестве  $\{1, \dots, n_j\}$ . Далее возможны различные постановки задачи в зависимости от априорных предположений о распределении величин  $z_j$ .

Распределение величины  $z_j$  зависит от соответствующего объекта, причины возникновения чрезвычайной ситуации, расположения датчиков и т. п. Все эти обстоятельства обычно неизвестны или их трудно учесть. Поэтому далее рассматривается предположение, что вероятность срабатывания всех датчиков на одном объекте достаточно велика. Как показали исследования [3], самая трудная для идентификации ситуация будет в том случае, когда срабатывает большее число датчиков, поэтому алгоритм идентификации, который хорошо работает в предположении  $z_j \equiv n_j$ , должен хорошо работать и в ситуации  $z_j < n_j$ .

В остальном примем, что выполняются те же предположения, что и в [3]. Состояние датчиков описывается переменными  $x_1, \dots, x_t$ , которые принимают значение 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  $i_1, \dots, i_{s_r}$ , принимают значение 1, а остальные — 0. Упорядоченное множество активных датчиков обозначим через  $S_d$ .

Групповой опрос, как и ранее, задается с помощью вектора  $a = (a_1, \dots, a_t)$ , 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  $A = (a^1, \dots, a^N)^T$ , где  $a^l = (a_1^l, \dots, a_t^l)$  — вектор  $l$ -го опроса. Если в группе опрашиваемых датчиков имеется хотя бы один активный, то мы наблюдаем наличие сигнала, который интерпретируем как 1. Если в группе, состоящей из опрашиваемых датчиков, нет ни одного активного источника, то от нее не поступает ни одного сигнала, что интерпретируется

как 0. Таким образом, в качестве ответа датчиков  $l$ -го опроса сформируется результат

$$f_l = (a_1^l \wedge x_1) \vee \dots \vee (a_t^l \wedge x_t), \quad (1)$$

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

Так же предполагается, что в сети возможны ошибки при передаче информации. Это означает, что значение  $f_l$  известно с некоторой ошибкой: при каждом опросе происходит искажение результат 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  $\alpha_0$  — вероятность искажения приема 0 (т.е. наблюдается значение 1 вместо 0), а  $\alpha_1$  — вероятность искажения приема 1 (т.е. наблюдается значение 0 вместо 1). Поэтому результат  $l$ -го опроса будет  $g_l$ , который принимает значения 0 или 1 в соответствии с матрицей  $W$  независимо от значений в других наблюдениях при условии, что значения  $f_l$  фиксированы.

Таким образом, отличие от рассмотренной в [3] модели состоит в том, что теперь величины  $x_1, \dots, x_t$  не являются независимыми. Необходимо получить алгоритмы построения матрицы опроса  $A$  и определить множество  $S$  объектов, на которых наблюдается чрезвычайная ситуация, на основании результатов опросов  $g_1, \dots, g_N$ .

### 3. ОПИСАНИЕ АЛГОРИТМА ПОФАКТОРНОГО ОБНАРУЖЕНИЯ АКТИВНЫХ ДАТЧИКОВ

Отличие данного алгоритма состоит в построении матрицы опросов  $A$ . Она должна учитывать структуру сети мониторинга, поскольку теперь датчики срабатывают зависимо.

Начнем с простейшего случая  $n_j \equiv n^*$ , т.е. количество датчиков на всех объектах одинаково. В этом случае сначала выбираются объекты, на которых будут опрашиваться датчики, а затем на каждом из этих объектов, выбирается опрашиваемый датчик. Следовательно, в каждом опросе на одном объекте может участвовать не более чем 1 датчик. Для создания такой матрицы поступают следующим образом. Сначала строится матрица опросов объектов, как и в предыдущем случае (т.е. при  $t = B$ ). Далее нужно на каждом объекте выбрать опрашиваемый датчик. Делается это следующим образом: если объект участвует в опросе, то среди датчиков этого объекта опрашиваемый датчик выбирается случайным образом независимо от участия в предыдущих опросах.

Для упрощения процедуры идентификации решение о наличии чрезвычайной ситуации на объекте принимается на основе лпофакторного анализа [4] с использованием метода максимального правдоподобия, но при этом происходит не поиск активного датчика, а поиск объекта, на котором имеются активные датчики.

Для принятия решения о том, имеется ли на объекте чрезвычайная ситуация или нет, используются следующие данные:  $x_{00}(j)$  — количество наблюдений, когда на  $j$ -ом объекте не опрашивался ни один датчик и результат опроса оказался  $g = 0$ ,  $x_{10}(j)$  — количество наблюдений, когда на  $j$ -ом объекте опрашивался хотя бы один датчик, а результат опроса  $g = 0$ ,  $x_{01}(j)$  — количество наблюдений, когда на  $j$ -ом объекте не опрашивался ни один датчик, а результат опроса  $g = 1$  и  $x_{11}(j)$  — количество наблюдений, когда на  $j$ -ом объекте опрашивался хотя бы один датчик и результат опроса  $g = 1$ . Поскольку результаты наблюдений предполагаются независимыми, то  $x_{00}(j)$ ,  $x_{01}(j)$ ,  $x_{10}(j)$  и  $x_{11}(j)$  образуют достаточную статистику для пофакторного определения наличия чрезвычайной ситуации на  $j$ -ом объекте.

В предположении, что на объекте срабатывают одновременно все датчики, получим для отношения правдоподобия поиска активных объектов те же результаты, что и в предыдущем случае.

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

$$L(j) = a_{00}x_{00}(j) + a_{10}x_{10}(j) + a_{01}x_{01}(j) + a_{11}x_{11}(j). \tag{2}$$

Поскольку значение  $r$  неизвестно, то здесь вместо него также используется оценка  $r_0$ . Поэтому используются величины

$$a_{00} = \log \frac{1 - \alpha_0 - \hat{p}(1 - \alpha_0 - \alpha_1)}{1 - \alpha_0 - p^*(1 - \alpha_0 - \alpha_1)}, \tag{3}$$

$$a_{01} = \log \frac{\hat{p}(1 - \alpha_0 - \alpha_1) + \alpha_0}{p^*(1 - \alpha_0 - \alpha_1) + \alpha_0}, \tag{4}$$

$$a_{10} = \log \frac{\alpha_1}{1 - \alpha_0 - p^*(1 - \alpha_0 - \alpha_1)}, \tag{5}$$

$$a_{11} = \log \frac{1 - \alpha_1}{\alpha_0 + p^*(1 - \alpha_0 - \alpha_1)}. \tag{6}$$

где  $p_0 = 1 - (1 - p^0)^{s_0}$ ,  $p_0^* = 1 - (1 - p^0)^{s_0-1}$ ,

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

На основании значения ошибки первого рода вычисляются значения порога  $L_0$ , при превышении которого логарифмом отношения правдоподобия  $L(j)$  принимается решение, что на  $j$ -ом объекте наблюдается чрезвычайная ситуация.

Алгоритм поиска объектов выглядит следующим образом. Исходными параметрами являются:

$B$  — количество объектов;

$s_0$  — количество предполагаемых чрезвычайных ситуаций;

$n^*$  — количество датчиков на объекте;

$\alpha_0$  и  $\alpha_1$  — вероятности ошибок искажения результат 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.

На основании результатов опросов вычисляются величины  $L(j)$  по формулам (2)–(6). Далее производится сравнение  $L(j)$  с порогом  $L_0$  для каждого  $j$ ; при превышении порога для некоторого  $L(\hat{j})$  считается, что на  $\hat{j}$ -ом объекте возникла внештатная ситуация.

Выходными характеристиками алгоритма будут величины  $\hat{j}_1, \dots, \hat{j}_{\hat{s}}$  — номера объектов, на которых зафиксированы чрезвычайные ситуации, где  $\hat{s}$  — количество обнаруженных чрезвычайных ситуаций.

Характеристиками качества алгоритма являются вероятности правильной идентификации чрезвычайных ситуации на объектах: пусть  $\hat{S} = \{\hat{j}_1, \dots, \hat{j}_{\hat{s}}\}$ , тогда  $P_1$  — вероятность того, что будет пропущен объект, где имеется чрезвычайная ситуация, т. е.  $S \not\subseteq \hat{S}$  и  $P_2$  — вероятность неправильного определения наличия чрезвычайной ситуации хотя бы на одном объекте, т. е.  $S \neq \hat{S}$ . Вероятности вычисляются в предположении, что на множестве возможных значений  $S$  задано равномерное распределение.

#### 4. РЕЗУЛЬТАТЫ ЧИСЛЕННОГО МОДЕЛИРОВАНИЯ, ЕСЛИ ВСЕ ДАТЧИКИ НА ОБЪЕКТЕ СРАБАТЫВАЮТ ОДНОВРЕМЕННО

Для проведения численных экспериментов была написана программа имитационного моделирования. Обычно проводилось 1000 численных экспериментов, когда на каждом эксперименте последовательность псевдослучайных чисел, с помощью которых разыгрываются случайные события, была различной. Объем вычислений позволяет гарантировать точность второго знака после запятой для оценок вероятностей, являющихся характеристиками алгоритма, если использовать доверительные интервалы, вычисленные по правилу  $3\sigma$ .

На основании исходных данных вычислялись исходные параметры алгоритма: продолжительность опроса, коэффициенты для вычисления отношения правдоподобия, порог принятия решения. Для каждого эксперимента случайным образом определялись объекты, на которых предполагается наличие чрезвычайной ситуации. Далее строилась случайным образом матрица опросов. На основании матрицы опросов и выбранных номеров объектов с чрезвычайными ситуациями определялись значения результатов опроса с учетом возможного искажения результата наблюдения. После этого для каждого объекта вычислялась статистика  $L(j)$ : определялось, участвовал ли хотя бы один из датчиков данного объекта в текущем опросе и в зависимости от этого и результата опроса определялся тот из коэффициентов  $a_{ij}$ , который следует добавить в статистику  $L(j)$ . После вычисления всех статистик определялись: множество объектов с чрезвычайными ситуациями и количество таких объектов. Далее определялось, является ли принятое решение правильным с точки зрения правильного определения всех объектов или не пропуска ни одного объекта с чрезвычайной ситуацией. Эти результаты фиксировались в специальных счетчиках и на основании их значений вычислялись оценки характеристик  $P_1$ ,  $P_2$  и  $\bar{s}$  как среднее значение количества обнаруженных объектов.

Были проведены численные исследования эффективности предлагаемого метода группового опроса, учитывающие структуру сети. В таб. 1 приведены результаты вычислений, когда  $B_0$ ,  $r = 1$ ,  $s_0 = 1$ .

Таблица 1.

Результаты обнаружения внештатных ситуаций при моделировании зависимого срабатывания датчиков.

$n^*$	$N$	$P_2$	$P_1$	$\bar{s}$
1	16	0,02	0,00	1,02
2	18	0,00	0,00	1,00
3	20	0,03	0,02	0,99

Приведенные данные показывают, что предлагаемый метод поиска одного объекта, на котором наблюдается внештатная ситуация, обладает хорошими свойствами. В первом случае с небольшой вероятностью был обнаружен лишний объект, а в третьем — в трех случаях их 100 объект не был обнаружен.

Применение непосредственно метода, рассмотренного в [3], дало следующие результаты, приведенные в таб. 2.

Таблица 2.

Результаты обнаружения внештатных ситуаций с использованием метода, не учитывающего структуру сети при моделировании зависимого срабатывания датчиков.

$t$	$n^*$	$N$	$P_2$	$P_1$	$\bar{s}$
200	1	32	0,00	0,00	1,00
400	2	72	0,03	0,00	2,03
600	3	120	0,03	0,01	3,01

Приведенные результаты показывают, что они практически такие же по точности определения чрезвычайной ситуации, как и в предыдущем случае, только необходимо существенно увеличить число опросов по сравнению с алгоритмом, когда учитывается структура сети. При этом число опросов растет линейно с ростом числа устанавливаемых на объектах датчиков.

Ниже приведены результаты расчетов, когда использовалось то же число опросов, что и при использовании алгоритма, учитывающего структуру сети мониторинга.

Таблица 3.

Результаты обнаружения активных датчиков с использованием метода, не учитывающего наличие зависимости при моделировании зависимого срабатывания датчиков.

$t$	$n^*$	$N$	$P_2$	$P_1$	$\bar{s}$
200	1	16	0,29	0,00	1,36
400	2	18	1,00	0,24	7,61
600	3	20	1,00	0,96	4,90

Приведенные результаты показывают, что применение алгоритма, учитывающего структуру сети, существенно повышает эффективность алгоритма группового опроса.

### 5. ИССЛЕДОВАНИЕ СВОЙСТВ АЛГОРИТМА ПОФАКТОРНОГО ОБНАРУЖЕНИЯ АКТИВНЫХ ДАТЧИКОВ ПРИ БОЛЕЕ ОБЩИХ ПРЕДПОЛОЖЕНИЯХ О СТРУКТУРЕ СЕТИ

Сначала рассмотрим случай, когда существенно нарушается условие  $z_j \equiv n_j$ . В такой ситуации применение описанного ранее алгоритма может быть неэффективным, поскольку теперь значение вероятности участия датчика в опросе не будет обеспечивать максимальную информативность наблюдений, поскольку вероятность получения в результате опроса значение 1 оказывается заниженной.

Были проведены численные исследования применения описанного выше алгоритма в ситуации, когда при возникновении чрезвычайной ситуации всегда срабатывает один датчик, а остальные датчики срабатывают независимо с вероятностью  $p_d$ . В этом случае среднее значение  $z_j$  окажется равным  $1 + p_d(n^* + 1)$ . Это приводит к тому, что при малых значениях  $p_d$  будет уменьшаться соответственно и вероятность наблюдения значения 1 как результата опроса, поэтому результат наблюдения может интерпретироваться неправильно.

Таблица 4.

Результаты обнаружения внештатных ситуаций при моделировании зависимого срабатывания датчиков.  $B0, r = 2, s_0 = 2$ .

$P_d$	$P_2$	$P_1$	$\bar{s}$
1,0	0,06	0,05	1,97
0,9	0,23	0,22	1,78
0,8	0,39	0,37	1,61
0,7	0,52	0,51	1,41
0,6	0,63	0,62	1,26

Приведенные данные показывают, что точность работы алгоритма падает с ростом вероятности того, что другие датчики на объекте не срабатывают. Связано это с тем, что при попадании опроса на датчик, который не сработал, происходит неправильная интерпретация результата наблюдений: считается в этом случае, что на объекте нет чрезвычайной ситуации. Анализ значений логарифма отношения правдоподобия для таких датчиков показывает, что

в этом случае наблюдается большой разброс их значений, что и приводит к дополнительному пропуску чрезвычайных ситуаций. Поэтому при малых значениях  $p_d$  целесообразно использовать метод индивидуального поиска датчиков из [3].

Необходимо учитывать возможность присутствия неоднородных объектов в сети, т.е. когда на различных объектах установлено различное количество датчиков. В этом случае необходимо разбить всю сеть на подсети, в которых на объектах число датчиков, которые могут срабатывать одновременно, примерно одинаково. В таком случае вероятность участия объекта в опросе при построении матрицы  $A$  будет прямо пропорциональна числу установленных на нем датчиков.

## 6. ВЫВОДЫ

1. При выборе алгоритма обнаружения чрезвычайных ситуаций важно учитывать структуру ее объектов. При этом игнорирование возможности срабатывания сразу нескольких датчиков на объекте может существенно ухудшить качество принимаемых решений.
2. Показано, что алгоритм обнаружения объектов с чрезвычайными ситуациями чувствителен к распределению величины  $z$ . Следовательно, необходимо проводить дополнительное исследование влияния распределения этой величины на решающее правило.
3. Для разбиения множества всех датчиков на объекты целесообразно использовать пофакторный способ обработки результатов наблюдений, или более сложные методы, учитывающие корреляцию срабатывания всех датчиков, расположенных на соответствующем объекте, при влиянии различных критических ситуаций.

## СПИСОК ЛИТЕРАТУРЫ

1. Маликова Е.Е., Цитович И.И. Задача обеспечения эффективности группового поллинга в сетях мониторинга. *Проектирование и технология электронных средств*. 2009. № 2. С. 35–38.
2. Маликова Е.Е. Модель группового поллинга для систем мониторинга и телеметрии, учитывающая зависимость между срабатыванием датчиков. *Тезисы докладов Всероссийской конференции “Информационно-телекоммуникационные технологии и математическое моделирование высокотехнологичных систем”–2011*. М.: РУДН. 2011. С. 40–42.
3. Маликова Е.Е., Цитович И.И. Метод группового поллинга при независимой активности датчиков в сетях мониторинга. *Информационные процессы*. 2011. Т. 11, № 2. С. 291–303.
4. Малютов М.Б. Нижние границы для средней длины последовательного планирования экспериментов. *Известия вузов. Математика*. 1983. Т. 27, № 11. С. 19–41.