О ГРАНИЦАХ ВЕРОЯТНОСТНЫХ АРГУМЕНТОВ ПРИ СИНТЕЗЕ ЛИНЕЙНЫХ СИГНАТУР И СТАТИСТИЧЕСКИХ ОЦЕНОК

Кобяк И. П.
Кафедра ЭВМ, Белорусский государственный университет информатики и радиоэлектороники
Минск, Республика Беларусь
E-mail: ipkobyak2012@mail.ru

В работе исследованы границы вероятностных аргументов, в рамках которых определены преимущества метода счета векторов состояний перед сигнатурным анализом и наоборот. Рассмотрены многомерные последовательности и соответствующие им приложения выполненных расчетов для определения границ аргумента r, удовлетворяющих требованию интегральной квазиоптимальности вероятности ошибки.

Целью данной работы является исследование границ аргументов при формировании оценок методом счета векторов состояний (СВС) и синтезе сигнатур, в пределах которых тот или иной алгоритм имеет меньшее значение суммы частичных составляющих производящей функции.

Пусть вероятность пропуска ошибки для метода СВС определяется равенством:

$$P_{cvc} = \frac{1}{m^n}\left[C_n^k(m-1)^{n-k} - 1\right], \qquad (1)$$

где cvc — *condition vector count*, k — число наблюдаемых событий заданного вида, $m = 2^r$ — общее число видов векторов в последовательности длиной n.

Соотношение, аналогичное (1), при формировании линейной свертки многоразрядных последовательностей имеет вид:

$$P_{msa} = \frac{2^{rn-l} - 1}{m^n}, \qquad (2)$$

где $l = \mathrm{intlog}_2 rn$, msa — *multichannel signature analysis*, а значение 2^{rn-l} — фактически число последовательностей, дающих одинаковую сигнатуру.

Определим границы интервалов значений k (где $k_0 = \frac{n}{m}$ — точка экстремума функции (1) при $n = \infty$), формируя решение как функцию от статистических аргументов $k = k_0 + \Delta k$, то есть с учетом отклонений аргумента от математического ожидания.

Математическая постановка задачи с использованием соотношений (1) и (2) при этом будет сводиться к анализу гипотетического равенства:

$$C_n^{k_0+\Delta k}(m-1)^{n-(k_0+\Delta k)} = 2^{rn-l} \qquad (3)$$

Совместное расположение графиков функций (1) и (2) определяет два случая формирования общих точек на плоскости пропуска ошибки. Так, при достаточно большом n и $k_0 = \frac{n}{m}$ для малых значений Δk, графики функций имеют две общие точки, которые могут быть определены из равенства (3). При значительных отклонениях Δk от k_0 существует одна общая точка, если $r > 1$.

Для решения поставленной задачи преобразуем левую часть соотношения (3), используя сокращенный вариант формулы Стирлинга. Тогда:

$$C_n^{k_0+\Delta k}(m-1)^{n-(k_0+\Delta k)} \approx$$
$$\approx \frac{(m-1)^{n-k_0-\Delta k}}{\left(\frac{1}{m}+\frac{\Delta k}{n}\right)^{k_0+\Delta k}\left(1-\frac{1}{m}-\frac{\Delta k}{n}\right)^{n-k_0-\Delta k}} \times$$
$$\times \sqrt{\frac{n}{2\pi\left(k_0+\Delta k\right)\left(n-k_0-\Delta k\right)}}$$

Прологарифмируем полученное равенство и заменим асимптотическое значение k_0 частным $\frac{n}{m}$. При этом, введя обозначение

$$\ln\left(\frac{n}{m}+\Delta k\right) = \ln\frac{n}{m} + \ln\left(1+\frac{m\Delta k}{n}\right)$$

с учетом допущения $\Delta k << n$, $n \to \infty$ получим:

$$n\ln m - \frac{1}{2}\ln 2\pi + \ln m - \frac{1}{2}\ln n - \frac{1}{2}\ln\left(1+\frac{m\Delta k}{n}\right) -$$
$$-\frac{1}{2}\ln\left(m-1-\frac{m\Delta k}{n}\right) \approx \ln\frac{m^{n+1}}{\sqrt{2\pi n}} - \frac{1}{2}\ln\left(m-1\right) \approx$$
$$\approx \ln\frac{m^{n+1}}{\sqrt{2\pi n\left(m-1\right)}}$$

Множители вида $\ln\left(1 \pm \frac{m\,\Delta k}{n}\right)$, $\frac{m\,\Delta k}{n} < 1$ заменим соответствующими рядами. Тогда после достаточно несложных преобразований получим:

$$C_n^{k_0\pm\Delta k}(m-1)^{n-(k_0\pm\Delta k)} =$$
$$= \frac{m^{n+1}}{\sqrt{2\pi n\left(m-1\right)}}\exp\left[-(\pm\Delta k)^2\frac{m}{2n}\frac{m}{(m-1)}\right]$$

С учетом полученного равенства, постановка задачи поиска границ вероятностных аргументов в (3) принимает вид:

$$\frac{m^{n+1}}{\sqrt{2\pi n\left(m-1\right)}}\exp\left[-\Delta k^2\frac{m}{2n}\frac{m}{(m-1)}\right] =$$
$$= 2^{rn-l} \qquad (4)$$

Определим теперь точки пересечения двух графиков как отклонение от экстремума $\frac{n}{m}$, то есть рассмотрим случай взаимного расположения графиков на плоскости при достаточно большом, но не асимптотическом n.

Разделим гипотетическое равенство (4) на m^n и экспоненту и прологарифмируем полученное соотношение. Тогда:

$$\ln \frac{m}{\sqrt{2\pi n\,(m-1)}} = \Delta k^2 \frac{m}{2n} \frac{m}{(m-1)} - l\ln 2$$

Отсюда:

$$\Delta k^2 = \frac{2n\,(m-1)}{m^2} \ln \frac{2^l m}{\sqrt{2\pi n\,(m-1)}}$$

Решениями же общего уравнения будут являться точки:

$$k = \frac{n}{m} \pm \sqrt{\frac{2n\,(m-1)}{m^2} \ln \frac{2^l m}{\sqrt{2\pi n\,(m-1)}}} \qquad (5)$$

Иными словами, в диапазоне изменения аргумента в пределах

$$\frac{n}{m} - \Delta k \le k \le \frac{n}{m} + \Delta k \qquad (6)$$

использование линейной свертки предпочтительнее метода СВС. Для всех остальных вероятностей статистический метод точнее. Заметим, однако, что в (5) и (6) должно выполняться неравенство вида $\Delta k < \frac{n}{m}$, в противном случае графики функций имеют только одну общую точку.

Преобразуем левую часть равенства (4) и определим значение функции в точке $k_0 - \Delta k = 0$. Подставляя значение $\Delta k = \frac{n}{m}$, получаем:

$$\frac{m^{n+1}}{\sqrt{2\pi n\,(m-1)}} \exp\left[-\frac{n}{2\,(m-1)}\right] = 2^{rn-l}$$

При $k = k_0 + \Delta k = \frac{2n}{m}$ значение в левой части соотношения имеет вид:

$$\frac{m^{n+1}}{\sqrt{2\pi n\,(m-1)}} \exp\left[-\frac{2n}{m-1}\right] = 2^{rn-l}$$

Иными словами, при данном уровне аппроксимации график биномиального распределения в диапазоне изменения аргумента $0 \le k \le 2k_0$ несимметричен относительно точки математического ожидания.

Общий случай для Δk в соотношении (5) может быть использован в качестве базы для анализа полученных результатов с точки зрения суммарного числа реализаций в классах перестановок, формируемых в различных поддиапазонах аргумента k.

Теорема. При биномиальном распределении последовательностей для

$$n \ge \frac{2\pi\,(m-1)}{r^2 m^2} e^9, r \le 4 \qquad (7)$$

суммарная вероятность пропуска ошибки в пределах интервала значений (6) больше, чем сумма вероятностей, определенных за пределами указанного отрезка аргументов.

Для доказательства теоремы обозначим сумму всех значений функции, принадлежащих диапазону $\frac{n}{2} \pm k$ в (7) через S. Соответственно сумма — реализаций выборки, принадлежащих вероятностям за пределами указанного интервала, будет равна $m^n - S$.

Для оценки отношения ${}^{C}/_{S}$ используем формулу среднеквадратического отклонения вида:

$$\sigma_K = \sqrt{np_0\,(x_\omega)\,q_0\,(x_\omega)} = \sqrt{n}\sqrt{\frac{1}{m}\left(1-\frac{1}{m}\right)}$$

При этом с учетом соотношения (5) имеем:

$$\sqrt{\frac{2\,(m-1)}{m^2} \ln \frac{2^l m}{\sqrt{2\pi n\,(m-1)}}} \ge 3\sqrt{\frac{1}{m}\left(1-\frac{1}{m}\right)}$$

Отсюда следует:

$$2\,\ln \frac{2^l m}{\sqrt{2\pi n\,(m-1)}} \ge 9,$$

При $n = \frac{2^l}{r}$, $l = \log_2 rn$, имеем результат, приведенный в теореме.

В частном случае, если $m = 2$, из (7) имеем $n > 0,5\pi e^9 \approx 12\,728$. Иными словами, при данной длине выборки отклонение от k_0 в (6) оказывается большим интервала $3\sigma_K$ со всеми вытекающими отсюда последствиями.

Таким образом, уровень отношения ${}^{C}/_{S}$ зависит от длины выборки и ${}^{C}/_{S} < 1$, при указанном в теореме значении n.

Что касается больших значений разрядности процесса r, то для всех $r \ge 5$ прямая вероятности пропуска ошибки (2) проходит выше уровня функции (1), определяемого границами среднеквадратического отклонения. Соответственно и преимущества сигнатурного интегрирования при этом сводятся к минимуму.

Итак, полученный результат показывает, что вероятность пропуска ошибки, соответствующая методу наблюдения заданного вектора в достаточно большой, но не асимптотической выборке при $r \le 4$ определяет преимущество линейного сверточного кодирования перед статистической методологией. Однако для $r \ge 5$, а также с точки зрения поддиапазона аргументов $2k_0 \le k \le n$ преимущество метода СВС в области больших значений статистики возрастает весьма существенно.

1. Кобяк И.П. // Электрон. моделирование. 1996, т.18. №3. С.58-62.
2. Яблонский С.В. Введение в дискретную математику. Москва, 1986.