О ВОЗМОЖНОСТИ СИГНАТУРНЫХ АНАЛИЗАТОРОВ СЖИМАТЬ ДАННЫЕ БЕЗ ПОТЕРИ ИНФОРМАЦИИ

Барашко А.С.
Кафедра ПМИ ДонНТУ

Abstract

Barashko A.S. On possibility of signature analyzers to compress the data lossless information. *The criterion of signature analyzer possibility to compress the sequence lossless information is found. The question about the number of given length sequences which may be lossless information to compress with high efficiency is investigated.*

В [1] предложено несколько методов сжатия сигнатурным анализатором (СА) анализируемой последовательности с вероятностями необнаружения любых ее искажений равными нулю. В качестве СА рассматриваются такие регистры сдвига с линейными обратными связями (РСЛОС), которые описываются полиномами и выполняют деление полинома, задающего анализируемую (входную) последовательность, на полином, задающий РСЛОС. На выходе последнего разряда РСЛОС появляется выходная последовательность, соответствующая частному от деления упомянутых полиномов, а содержимое РСЛОС после подачи последнего символа входной последовательности определяет остаток от деления (сигнатуру). Предложенные в [1] методы основаны на анализе не только сигнатуры, но и частного от деления. При этом РСЛОС выбирается таким образом, чтобы в случае отсутствия искажения входной последовательности его выходная последовательность была достаточно простой, скажем, состоящей из повторений одного и того же символа или повторений некоторой короткой последовательности.

Для заданной длины между последовательностями символов некоторого конечного поля и полиномами над этим полем, степень которых не превышает длину без единицы, существует взаимно однозначное соответствие. Поэтому будем отождествлять эти понятия. РСЛОС также можно отождествить с задающим его полиномом, степень которого равна числу разрядов регистра. В [1] определена также эффективность сжатия регистром входной последовательности $\eta = 1 - k/(N+1)$, где N - степень входного полинома, а k - длина сигнатуры (степень полинома, задающего РСЛОС). Ограничившись бинарными последовательностями, будем считать, что входная последовательность сжимается РСЛОС без потери информации, если его выходная последовательность состоит из повторений единиц. В [1] показано, что любой полином можно сжать без потери информации с эффективностью сжатия, близкой к 0,5. Такая эффективность сжатия вполне приемлема для последовательностей небольшой длины, однако, чтобы сжать без потери информации последовательность длины, скажем, 100 потребуется РСЛОС из 50 разрядов, что неприемлемо. С другой стороны, существуют последовательности, которые можно сжимать без потери информации с высокой (близкой к 1) эффективностью. Возникает вопрос, какова доля таких "хороших" последовательностей среди всех последовательностей фиксированной длины.

В данной работе найден критерий возможности сжатия данной последовательности сигнатурным анализатором без потери информации и показано, что, к сожалению, "хороших" последовательностей очень мало. Проведенные исследования базируются на двух утверждениях и их следствиях, в которых используются следующие обозначения.

Пусть $GF(q)$- произвольное конечное поле, $g(x)$, $h(x)$ - полиномы над этим полем и $\deg g(x)$, $\deg h(x)$ - их степени [2]. Положим $Q_{h(x)}[g(x)]$ - частное, а $R_{h(x)}[g(x)]$ - остаток от деления $g(x)$ на $h(x)$, т.е. $g(x) = h(x)Q_{h(x)}[g(x)] + R_{h(x)}[g(x)]$, где $\deg R_{h(x)}[g(x)] < \deg h(x)$.

Утверждение 1. *Пусть $g(x)$ и $h(x)$ - такие полиномы, что* $\deg h(x) < \deg g(x) = N$. *Если* $\Phi(x) = Q_{h(x)}[g(x)]$ *и* $\deg R_{h(x)}[g(x)] < N - \deg h(x)$, *то* $R_{h(x)}[g(x)] = R_{\Phi(x)}[g(x)]$ *и* $h(x) = Q_{\Phi(x)}[g(x)]$.

Доказательство. Пусть $h'(x) = Q_{\Phi(x)}[g(x)]$. Тогда $g(x) = \Phi(x)h(x) + R_{h(x)}[g(x)] = \Phi(x)h'(x) + R_{\Phi(x)}[g(x)]$. И, значит, $\Phi(x)(h(x) - h'(x)) = R_{\Phi(x)}[g(x)] - R_{h(x)}[g(x)] = \varphi(x)$. Поскольку $\deg R_{\Phi(x)}[g(x)]$, $\deg R_{h(x)}[x)] < \deg \Phi(x) = N - \deg h(x)$, то $\deg \varphi(x) < \deg \Phi(x)$. Последнее неравенство может иметь место только тогда, когда $h(x) = h'(x)$. Отсюда следует, что $\varphi(x) = 0$ и $R_{h(x)}[g(x)] = R_{\Phi(x)}[g(x)]$. Утверждение доказано.

Следствие 1. *Пусть $g(x)$и $h(x)$-такие полиномы, что* $\deg h(x) \le \deg g(x)/2$ *и* $\Phi(x) = Q_{h(x)}[g(x)]$. *Тогда* $R_{h(x)}[g(x)] = R_{\Phi(x)}[g(x)]$ *и* $h(x) = Q_{\Phi(x)}[g(x)]$.

Доказательство. Поскольку $\deg \Phi(x) + \deg h(x) = \deg g(x)$, то $\deg \Phi(x) \ge \deg g(x)/2$. Поэтому $\deg R_{h(x)}[g(x)] < \deg h(x) \le \deg g(x)/2 \le \deg \Phi(x)$ и условие утверждения 1 выполняется. Значит, имеют место соответствующие равенства следствия.

Следствие 1 позволяет сделать вывод, что для любой входной последовательности и любой последовательности (в том числе состоящей из повторений одного символа), длина которой не превышает половину длины входной, можно построить такой СА (определяемый полиномом $\Phi(x)$), что вторая последовательность в случае неискажения входной появится на выходе последнего разряда СА. Если вторая последовательность состоит из повторений одного и того же символа, то для ее контроля требуется простейшее оборудование и сжатие будет определяться только сигнатурой. При выполнении этих условий можно установить, произошло ли искажение (любое) входной последовательности.

Хотя представленные ниже результаты можно было бы получить для любого конечного поля, для упрощения изложения ограничимся полем $GF(2)$. Для произвольного натурального n положим $f_n(x) = x^n + x^{n-1} + \ldots + x + 1$ (этому полиному соответствует последовательность единиц длины $n + 1$).

Определение. Полином $g(x)$ степени N сжимаем k-разрядным ($1 \le k < N$) СА без потери информации, если существует такой полином $\Phi(x)$ степени k , что $Q_{\Phi(x)}[g(x)] = f_{N-k}(x)$. При этом эффективность сжатия равна $\eta = 1 - k/(N+1)$.

Пусть z - действительное неотрицательное число. Через $[z]$ обозначим целую часть числа z , а через $\lceil z \rceil$ - наименьшее целое, которое не меньше z. Для натурального числа $N \ge 2$ зафиксируем соотношения

$$\lceil N/2 \rceil + [N/2] = N \,,\ \lceil N/2 \rceil - [N/2] \le 1 \,. \qquad (1)$$

Найдем критерий возможности сжатия без потери информации полинома степени $N \ge 2$.

Утверждение 2. *Полином $g(x)$ степени $N \geq 2$ сжимаем без потери информации некоторым полиномом степени $1 \leq k < N$ тогда и только тогда, когда* $\deg R_{f_{N-k}(x)}[g(x)] < k$.

Доказательство. Необходимость. Согласно определению существует такой полином $\Phi(x)$ степени k, что $Q_{\Phi(x)}[g(x)] = f_{N-k}(x)$. Положим $\Phi'(x) = Q_{f_{N-k}(x)}[g(x)]$ и рассмотрим два случая.

Случай $\lceil N/2 \rceil \leq k < N$. Используя (1), находим $N - k \leq N - \lceil N/2 \rceil = [N/2] \leq N/2$. Из следствия 1 вытекает равенство $R_{f_{N-k}(x)}[g(x)] = R_{\Phi'(x)}[g(x)]$, а так как $\deg \Phi'(x) = k$, то $\deg R_{f_{N-k}(x)}[g(x)] < k$.

Случай $1 \leq k < \lceil N/2 \rceil$. Используя (1), находим $N - k > k$. Поэтому $\deg R_{\Phi(x)}[g(x)] < N - k$. Считая в утверждении 1 $h(x) = \Phi(x)$, получаем

$$R_{\Phi(x)}[g(x)] = R_{f_{N-k}(x)}[g(x)].$$

Откуда $\deg R_{f_{N-k}(x)}[g(x)] < k$.

Достаточность. Пусть $\deg R_{f_{N-k}(x)}[g(x)] < k$. Полагая $\Phi(x) = Q_{f_{N-k}(x)}[g(x)]$, из утверждения 1 находим $Q_{\Phi(x)}[g(x)] = f_{N-k}(x)$, т.е. согласно определению полином $g(x)$ сжимаем без потери информации полиномом степени k. Утверждение доказано.

Непосредственно из утверждения 2 получаем следующий результат.

Следствие 2. *Любой полином степени $N \geq 2$ сжимаем без потери информации некоторым сигнатурным анализатором, соответствующим полиному степени $\lceil N/2 \rceil$.*

Таким образом, для сжатия без потери информации полинома степени N всегда найдется полином степени $\lceil N/2 \rceil$. Однако могут существовать полиномы более низкой степени, которые способны сжимать без потери информации исходный полином.

Для данного полинома $g(x)$ степени $N \geq 2$ можно определить $M_{g(x)}$ - наименьшую степень полинома, способного сжать $g(x)$ без потери информации. Положим $M_{g(x)} = \min\{N > k \geq 1 \mid \deg R_{f_{N-k}(x)}[g(x)] < k\}$. Следствие 2 позволяет указать верхнюю границу $M_{g(x)} \leq \lceil \deg g(x)/2 \rceil$. Возникает вопрос о достижимости верхней границы, т.е. для любого ли $N \geq 2$ существует полином $g(x)$ степени N, удовлетворяющий равенству $M_{g(x)} = \lceil N/2 \rceil$. Покажем, что ответ на этот вопрос утвердительный.

Для произвольного $N \geq 2$ положим $g(x) = x^N + x^{N-1} + \ldots + x^{\lceil N/2 \rceil}$. Тогда, используя (1), находим $g(x) = x^{\lceil N/2 \rceil} f_{[N/2]}(x)$, т. е. $g(x)$ сжимаем без потери информации сигнатурным анализатором, соответствующим полиному $x^{\lceil N/2 \rceil}$. При этом сигнатура равна нулевому полиному. Покажем, что $g(x)$ не сжимаем без потери информации никаким полиномом степени k, где $\lceil N/2 \rceil > k \geq 1$. В самом деле, поскольку $N - k > N - \lceil N/2 \rceil = [N/2] \geq \lceil N/2 \rceil - 1$, то $g(x) = x^k f_{N-k}(x) + x^{\lceil N/2 \rceil - 1} + \ldots + x^k$ и $R_{f_{N-k}(x)}[g(x)] = x^{\lceil N/2 \rceil - 1} + \ldots + x^k$. Так как $\deg R_{f_{N-k}(x)}[g(x)] = \lceil N/2 \rceil - 1 \geq k$, то утверждение 2 позволяет сделать вывод, что не существует полинома степени k, сжимающего $g(x)$ без потери информации, т. е. $M_{g(x)} = \lceil N/2 \rceil$. Итак, достижимость верхней границы для величины $M_{g(x)}$ доказана.

Любой полином $g(x)$ степени $N \geq 2$ характеризуется наибольшей эффективностью сжатия без потери информации $\eta_{g(x)} = 1 - M_{g(x)}/(N+1)$. При этом нижняя граница эффективности сжатия определяется неравенством $\eta_{g(x)} \geq ([N/2] + 1)/(N + 1)$. Хотя нижняя граница эффективности сжатия зависит от четности степени

сжимаемого полинома, в любом случае имеет место неравенство $\eta_{g(x)} \geq 1/2$ и, таким образом, наибольшая эффективность сжатия любого полинома не меньше $1/2$.

Теперь для произвольного натурального $1 \leq k < N$ определим количество полиномов степени $N \geq 2$, для которых $M_{g(x)} \leq k$. Положим $P(N, k) = |\{g(x) \mid \deg(x) = N \ \& \ M_{g(x)} \leq k\}| = |\{ g(x) \mid \deg g(x) = N \ \& \ \eta_{g(x)} = 1 - k/(N+1)\}|$, т.е. $P(N, k)$ - количество полиномов степени N, для которых эффективность сжатия без потери информации не меньше $1 - k/(N+1)$. Из следствия 2 вытекает, что при $\lceil N/2 \rceil \leq k < N$ $P(N, k) = 2^N$, так как именно столько существует полиномов степени N.

Пусть далее, $1 \leq k < \lceil N/2 \rceil$. Полином $g(x)$ степени N, сжимаемый без потери информации полиномом $\Phi(x)$ степени k, можно представить в виде $g(x) = \Phi(x) f_{N-k}(x) + R_{\Phi(x)}[g(x)]$. Число таких полиномов не превышает суммы числа полиномов степени k и числа полиномов, степень которых меньше k, т. е. 2^{2k}. Покажем, что упомянутая сумма равна 2^{2k}. Для этого достаточно доказать, что если два полинома $g(x) = \Phi(x) f_{N-k}(x) + R_{\Phi(x)}[g(x)]$ и $g'(x) = \Phi'(x) f_{N-k}(x) + R_{\Phi'(x)}[g'(x)]$ равны и $\deg \Phi(x) = \deg \Phi'(x) = k$, то $\Phi(x) = \Phi'(x)$ и $R_{\Phi(x)}[g(x)] = R_{\Phi'(x)}[g'(x)]$. Если $g(x) = g'(x)$, то

$$f_{N-k}(x)(\Phi(x) - \Phi'(x)) = R_{\Phi'(x)}[g(x)] - R_{\Phi(x)}[g(x)]. \qquad (2)$$

Полиномы в левой и правой части равенства (2) обозначим соответственно L и R. Поскольку $\deg \Phi(x) = \deg \Phi'(x) = k$, то $\deg R \leq k - 1$. Если $\Phi(x) \neq \Phi'(x)$, то $\deg L \geq N - k$. Учитывая рассматриваемый диапазон для k и соотношения (1), находим $N - k > N - \lceil N/2 \rceil = [N/2] \geq \lceil N/2 \rceil - 1 \geq k$ и $\deg L > k$. Полученные неравенства противоречат равенству $\deg L = \deg R$. Поэтому предположение $\Phi(x) \neq \Phi'(x)$ неверно и, значит, $\Phi(x) = \Phi'(x)$. Из (2) находим $R_{\Phi'(x)}[g(x)] = R_{\Phi(x)}[g(x)]$.

Поскольку $\eta = 1 - k/(N+1)$ - эффективность сжатия полинома степени N полиномом степени k, то $k = (N+1)(1 - \eta)$. Пусть $V(N, \eta) = \dfrac{P(N, (N+1)(1-\eta))}{2^N}$ - доля полиномов степени N, сжимаемых без потери информации с эффективностью не менее η. Ясно, что при $0 \leq \eta \leq ([N/2] + 1)/(N+1)$ $V(N, \eta) = 1$, а при $([N/2] + 1)/(N+1) < \eta \leq N/(N+1)$

$$V(N, \eta) = 2^{N + 2(1 - \eta(N+1))}. \qquad (3)$$

Если $\eta \to N/(N+1)$, то $V(N, \eta) \to 2^{2-N}$. Из (3) видно, что доля полиномов, сжимаемых без потери информации с эффективностью $\eta > 0{,}5$, быстро убывает с ростом η, причем тем быстрее, чем больше N.

Таким образом, существует небольшой процент полиномов, которые можно сжать без потери информации с высокой эффективностью сжатия. Этот вывод контрастирует с утверждением, полеченным в [1] с использованием кодов Юстесена, о том, что вероятность необнаружения искажения можно сделать произвольно малой при произвольно высокой эффективности сжатия для достаточно длинных последовательностей. Однако, это утверждение имеет небольшую практическую ценность, так как относится к очень длинным анализируемым последовательностям.

Литература

1. Pradhan D.K., Gupta S.K. A new framework for designing and analyzing BIST techniques and zero aliasing compression // IEEE Trans. Comput. - 1991. - 40, № 6. - P. 743 - 763.
2. Блейхут Р. Теория и практика кодов, контролирующих ошибки. - М.: Мир, 1986. - 576 с.

Поступила в редакційну колегію 1. 11.2001 р.