ЛЕКЦИЯ 5. СИСТЕМЫ ПЕРЕДАЧИ ДАННЫХ С ПРИМЕНЕНИЕМ ПОМЕХОУСТОЙЧИВЫХ КОДОВ, ОБНАРУЖИВАЮЩИХ ОШИБКИ

3. МЕТОДЫ ОБНАРУЖЕНИЯ ОШИБОК ПОМЕХОУСТОЙЧИВЫМИ КОДАМИ CRC (Cyclic Redundancy Check).

3.1. Алгоритмы формирования циклической проверочной суммы

CRC с нулевыми начальными состояниями ячеек регистров деления.

3.1.1. Алгоритм с "простой" CRC.

Этот алгоритм, названный "простым", относится к классическим систематическим циклическим (n,k)-кодам с числом избыточных элементов в кодовой комбинации, равным (n–k) = m, где m – степень образующего примитивного многочлена $P(x)$.

Рассмотрим, прежде всего, *варианты кодов с "простой" CRC с нулевыми начальными состояниями ячеек регистров деления* на многочлена $P(x)$.

Общий *алгоритм кодирования* классическим систематическим (*n*,*k*)-кодом :

1. умножение многочлена исходной информации $\varphi(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{k-1} x^{k-1}$ на одночлен x^m ;
2. деление произведения $x^m \varphi(x)$ на образующий многочлен $P(x)$ степени m и определение остатка от деления
$$r(x) = r_0 + r_1 x + r_2 x^2 + \ldots + r_{m-1} x^{m-1};$$
3. формирование разрешенной кодовой комбинации $f(x) = r(x) + x^m \varphi(x).$
4. Очевидно, что $f(x) \equiv 0 \mod P(x),$ т.е. разрешенная кодовая комбинация делится на $P(x)$ без остатка.

Алгоритм декодирования.

алгоритм декодирования сводится к вычислению CRC по принятым информационным элементам и сравнению с принятыми проверочными элементами. Если циклические проверочные элементы CRC совпадают, то считается, что ошибки в комбинации отсутствуют.

В литературе циклическую проверку, порождённую многочленом $P(x)$ степени m, часто обозначают *CRC-m*.

Различают три варианта длин n комбинаций: *оптимальная* полная длина $n = 2^m - 1$; *укороченный* код с длиной комбинации $n < 2^m - 1$ и *удлинённый* код с $n > 2^m - 1$.

Оптимальный полный код CRC.

В оптимальном варианте при $n = 2^m - 1$ наиболее полно проявляются свойства циклического кода:

- деление разрешённых комбинаций $f(x)$ на образующий многочлен $P(x)$ без остатка;
- поэлементная сумма по mod2 двух или более разрешенных комбинаций порождает новую разрешенную комбинацию;
- циклический сдвиг разрешённой комбинации также порождает другую разрешённую комбинацию;
- наиболее оптимальным образом согласуются обнаруживающая способность кода и скорость кода k/n.

Из равенства $n = 2^m - 1$ следует равенство Хэмминга $(n-k)=m=\log_2(n+1)$, доказывающее, что это действительно оптимальный (n,k)-код, имеющий плотную упаковку и минимальное кодовое расстояние Хэмминга dmin=3

$$P_{\text{но}} = \sum_{w_i = d_{\min}}^{n} A(w_i)\, p^{w_i} (1-p)^{n-w_i},$$

где w_i – вес разрешённой комбинации кода;

$A(w_i)$ – количество разрешённых комбинаций с весом w_i, эту характеристику называют ещё *весовым спектром* кода;

p – вероятность битовой ошибки в канале ДСК.

Для полных двоичных кодов, для которых $n = 2^m - 1$ и $d_{min}=3$ весовой спектр находится довольно просто как коэффициенты при z^{w_i} в разложении по степеням z следующей функции

$$\nu(z) = \frac{1}{n+1}\left[(1+z)^n + n(1+z)^{\frac{n-1}{2}} \cdot (1-z)^{\frac{n+1}{2}}\right].$$

Например, для циклического кода $(n,k)=(15,11)$ с CRC-4 весовой спектр представлен в табл. 1.4:

Весовой спектр циклического кода (15,11) Таблица 1.4

w_i	0	3	4	5	6	7	8	9	10	11	12	15
$A(w_i)$	1	35	105	168	280	435	435	280	168	105	35	1

$P_{оо}=1-P_{пп}-P_{но}$, где $P_{пп}$ – вероятность правильного приёма комбинации, равная $P_{пп}=(1-p)^n$.

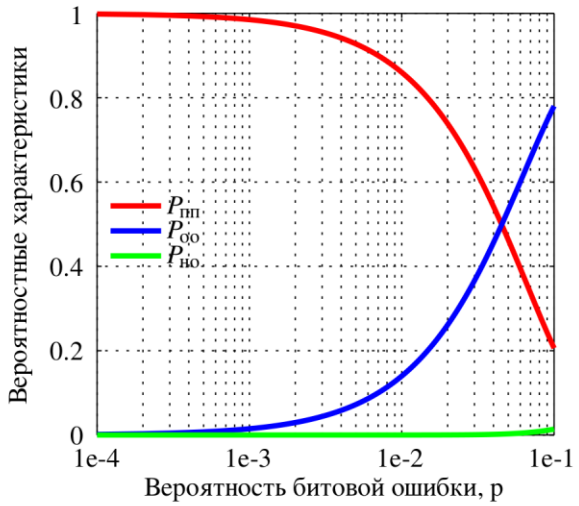
Пример. Рассчитать вероятностные характеристики для кода $(n,k)=(15,11)$ с CRC-4 в канале ДСК с битовой вероятностью ошибки $p = 10^{-3}$.

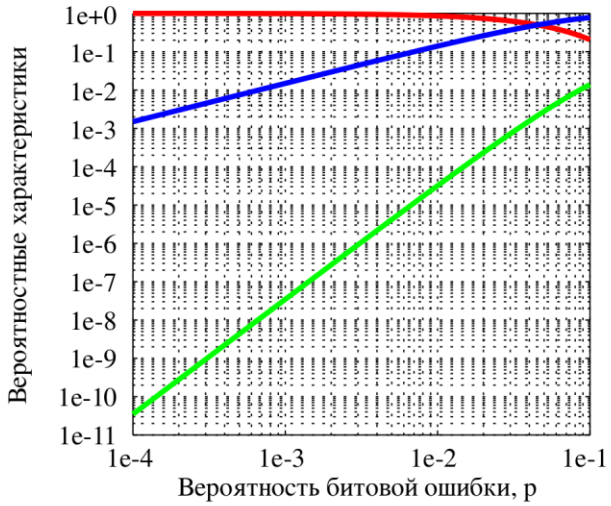
Вероятность правильного приема $P_{пп}=(1-p)^{15}=(0{,}999)^{15}=0{,}9851$.

Расчетные вероятности необнаруживаемых ошибок кратности w_i и суммарная вероятность $P_{но}$ представлены в табл. 1.5.

Таблица 1.5

w_i	3	4	5	6	7	$P_{но\ сумм}\cdot\cong$ $3{,}46\cdot10^{-8}$
$P_{но}(w_i)$	$3{,}46\cdot10^{-8}$	$1{,}04\cdot10^{-10}$	$1{,}66\cdot10^{-13}$	$2{,}77\cdot10^{-16}$	$4{,}31\cdot10^{-19}$	





Укороченные CRC-m коды.

При $n < 2^m - 1$ получим укороченный код с проверочной контрольной суммой CRC-*m* и числом информационных элементов $k=n-m$. Такой код имеет меньшую относительную кодовую скорость k/n и, соответственно, большую относительную избыточность.

Например, для полного кода $(n,k)=(15,11)$ скорость равна $k/n = 0{,}733$ и избыточность $(n-k)/n=0{,}266$. А укороченный (9,5)-код имеет скорость 0,555, а избыточность – 0,444. При этом минимальное кодовое расстояние остаётся равным $d_{min}=3$, чем и определяется обнаруживающая способность укороченного кода.

Вместе с тем, при оценке вероятностных характеристик необходимо учесть, что для укороченного кода поменяется весовой спектр, который чаще всего определяется путем моделирования.

В укороченном (n,k)-коде свойства циклических кодов сохраняются не полностью. Так, сохраняются свойство делимости разрешённой комбинации в виде многочлена $f(x)$ на образующий многочлен $P(x)$ без остатка и свойство линейности, но не сохраняется свойство циклических сдвигов, так как примитивный многочлен $P(x)$ степени m не будет делителем двучлена (x^n+1), если $n < 2^m - 1$.

***Удлинённые CRC-m коды*.**

В ряде систем применяется код с CRC-*m* при длине комбинации $n > 2^m - 1$ при том же числе проверочных элементов $m=n-k$. Такой код, по сравнению с оптимальным полным кодом, имеет большую относительную скорость k/n и, соответственно, меньшую относительную избыточность. Например, если взять код (20,16) вместо (15,11), то получим скорость $k/n=0{,}8$ и избыточность 0,2.

Особенностью такого кода является то, что у него, по сравнению с полным кодом, увеличится доля необнаруживаемых ошибок.

Известно, что образующий многочлен $P(x)$ степени m должен быть делителем двучлена $(x^{2^m-1}+1)$. А так как в удлинённом коде $n > 2^m - 1$, то в комбинации могут возникнуть такие двукратные ошибки с многочленом ,

$$e(x) = x^i (x^{2^m-1}+1)$$

что $i+(2^m-1) \le n-1$, а многочлен $e(x)$ будет делиться на двучлен $(x^{2^m-1}+1)$ и, следовательно, на образующий многочлен $P(x)$ без остатка. Т.е. такие двукратные ошибки не будут обнаруживаться кодом с CRC-*m* при длине комбинации $n > 2^m - 1$.

Способность циклического кода с "простой" CRC-m обнаруживать ошибки.
Так как образующий многочлен $P(x)$ такого кода является примитивным, то он должен иметь нечетное число слагаемых, включая 1 и x^m, т.е. 3,5,7 и т.д.
Исходя из этого, можно дать оценку обнаруживающей способности такого кода, учитывая, что многочлен обнаруживаемых ошибок $e(x)$ не должен делиться на $P(x)$ без остатка. Однократная ошибка: $e(x) = x^i$, где $0 \le i \le n-1$.
Такой многочлен однократной ошибки не может делиться без остатка на другой многочлен, имеющий более одного члена. Значит, однократные ошибки будут гарантированно обнаруживаться.
Двукратную ошибку с многочленом $e(x) = x^i + x^j$, $0 \le i < j \le n-1$, можно представить произведением двух сомножителей $e(x) = x^i (1+ x^{j-i}) = x^i (1+ x^v)$.
Такой многочлен двукратных ошибок $e(x)$ не будет делиться на $P(x)$ без остатка в том случае, если ни один из сомножителей не будет делиться на $P(x)$. Одночлен x^i не делится на $P(x)$ без остатка. Второй сомножитель $(1+ x^v)$ также не будет делиться без остатка на $P(x)$ степени m, так как $v \le n-1=2^m - 2$, а примитивный многочлен степени m является делителем двучлена $(1+ x^i)$ с наименьшей степенью $i = (2^m -1)$. Таким образом, двукратные ошибки код с CRC-*m* также гарантированно обнаруживает.
Кроме того, будут обнаруживаться и те ошибки большей кратности, многочлены которых $e(x)$ не делятся без остатка на $P(x)$.
Наоборот, необнаруживаемыми будут те ошибки, многочлены которых $e(x)$ будут кратны многочлену $P(x)$. Общее количество и вес этих комбинаций ошибок определяется весовым спектром кода $A(w_i)$.

Варианты кодов с «простой» CRC-m и с нулевыми начальными состояниями , применяемых в реальных системах (протоколах).

1. ***Цифровые тракты плезиохронной иерархии PDH, ITU – T G.704***

   CRC – 4, $P(x)=1 + x + x^4$

   Субцикл – 8 циклов Е1, содержит k=8×8 × 32=2048 бит; n – k=4; n=2052 бит

   $f(x) = r(x) + x^4\ \varphi(x)$. Проверочные элементы $r(x) \Rightarrow (r_0, r_1, r_2, r_3)$ передаются в следующем субцикле. Код удлинённый с $d_{min} = 3$. Но не все двукратные ошибки обнаруживаются.

   ***Многочлен* $e(x) = x^i + x^j$, $0 \le i < j \le n-1$ можно представить как $e(x) = x^i (1+ x^{j-i}) = x^i (1+ x^v)$, тогда при *v* кратном числу 15 примитивный многочлен $P(x)$ будет делителем двучлена $(1+ x^v)$, т.е. такие ошибки код не обнаружит.**

2. **Другими примерами простых кодов с CRC-*m* являются, рекомендованные для цифровых сетей синхронной иерархии SDH, код CRC-6 с образующим многочленом $P(x) = 1+x+x^6$, рекомендация ITU–T G.704, и код CRC-7 с образующим многочленом $P(x) = 1+x^3+x^7$, рекомендации ITU–T G.704 и G.832.**

3. Но самым простым является код **CRC-3** в схеме управления пропускной способностью канала (LCAS – Link Capaciti Adjustment Scheme), применяемый для передачи кадров в сети SDH в соответствии с рекомендацией **ITU–T G.707**. Информационная последовательность, состоящая из 29 бит, кодируется систематическим $(n,k)=(32,29)$-кодом с образующим многочленом $P(x) = 1+x+x^3$. Очевидно, код является удлинённым, поэтому он гарантированно обнаруживает все однократные и многие двукратные ошибки. В то же время, как было пояснено выше, часть двукратных ошибок он не обнаруживает.

4. На сегодняшний день одним из самых сложных является пример кода с **CRC-32** стандарта IEEE 802.3, применяемого в кадрах Ethernet на MAC уровне. Образующий многочлен имеет вид:

$$P(x) = 1 + x + x^2 + x^4 + x^5 + x^7 + x^8 + x^{10} + x^{11} + x^{12} + x^{16} + x^{22} + x^{23} + x^{26} + x^{32}.$$

3.1.2. Алгоритмы обнаружения ошибок двоичными (n,k)-кодами с расширенной CRC и нулевыми состояниями ячеек регистров.

Название "расширенная CRC" здесь вводится в связи с тем, что образующий многочлен имеет вид: $G(x)=(1+x)\,P(x)$, где $P(x)$ – примитивный многочлен степени m. Варианты таких кодов будут, как и ранее, оптимальные (полные или с плотной упаковкой) при $n = 2^m - 1$, укороченные – при $n < 2^m - 1$ и удлинённые – при $n > 2^m - 1$. В таких (n,k)-кодах с расширенной CRC число проверочных элементов равно $(n-k)=m+1$, а число информационных элементов – $k = n-m-1$.

Рассмотрим общие свойства (n,k)-кода с расширенной CRC на примере оптимального систематического кода (плотная упаковка) с $n = 2^m - 1$. Для большей наглядности будем вести построение такого кода с образующим многочленом $G(x)=(1+x)\,P(x)$ в два этапа. На первом этапе информационная k-элементная комбинация $\varphi(x)=a_0 + a_1 x + \ldots + a_{k-1}x^{k-1}$ умножается на x^m ($m\geq 3$) и делится на многочлен $P(x)$. Остаток от деления

$$r(x) = r_0 + r_1 x + \ldots + r_{m-1}x^{m-1}$$

добавляется к информационным элементам со стороны младшего разряда. При этом будет получена комбинация из (n–1)-го элементов: $(r_0, r_1, \ldots, r_{m-1}, a_0, a_1, a_2, \ldots, a_{k-1})$ (1.15) с минимальным кодовым расстоянием Хэмминга $d_{min}=3$.

На втором этапе комбинация (1.15) проверяется на четность и к ней со стороны младшего разряда добавляется ещё один проверочный элемент *b на четноть*: $(b, r_0, r_1, \ldots, r_{m-1}, a_0, a_1, a_2, \ldots, a_{k-1})$.

Следовательно, разрешенные комбинации кода с расширенной CRC будут иметь только четный вес, начиная с $w_{min}=4$ среди ненулевых комбинаций. Отсюда также следует, что минимальное кодовое расстояние кода с расширенной CRC также будет равно $d_{min}=4$. Исходя из этого, такой полный код способен гарантированно обнаруживать однократные, двукратные и трёхкратные ошибки. Кроме того, код будет обнаруживать также ошибки более высокой нечетной кратности и более высокой четной кратности , если многочлен ошибок $e(x)$ не будет делиться на $G(x)$ без остатка. Следовательно, код не сможет обнаружить только те ошибки, многочлен которых $e(x)$ кратен образующему многочлену $G(x)$.

Например, для кода $(n,k) = (15,10)$ с CRC-5 и примитивным многочленом $P(x) = 1+x+x^4$ получим образующий многочлен $G(x)=(1+x)P(x)=1+x^2+x^4+x^5$. Весовой спектр такого кода с $n = 2^4 - 1 = 15$ будет только четным с числами $A(w_i)$ комбинаций веса w_i , показанным в табл. 1.6

w_i	0	4	6	8	10	12
$A(w_i)$	1	105	280	435	168	35

Особенностью кодов с расширенной CRC и $d_{min} = 4$ является и то, что такие полные (n,k)-коды с $n=2^m - 1$ могут работать как в режиме только обнаружения ошибок (обнаруживать все однократные, двукратные и все другие ошибки нечетной кратности), так и в режиме исправления однократных ошибок и гарантированного обнаружения всех двукратных ошибок.

Примеры систем и рекомендаций по использованию для обнаружения ошибок (n, k)-кодов с расширенной CRC и с нулевыми начальными состояниями ячеек регистра деления.

1. Рекомендация ITU-T G.704, 1998 года, предписывает применение для обнаружения ошибок кода с CRC-5 и образующим многочленом $G(x)=(1+x)\ P(x) = (1+x)(1+x+x^4) = 1 + x^2 + x^4 + x^5$.

Получим полный (оптимальный) $(n, k) = (15,10)$ циклический код с $d_{min} = 4$. Такой код, как было доказано выше, будет обнаруживать все ошибки нечетной кратности и все двукратные ошибки. В табл. 1.6 приведен весовой спектр разрешённых комбинаций этого кода, из которого следует, что код не будет обнаруживать определенные ошибки четной кратности, начиная с 4-ой. Так, вероятность появления 4-х кратной необнаруживаемой ошибки в канале ДСК будет определяться выражением:

$$P(4,n) = A(w=4)\ p^4\ (1-p)^{n-4},$$

где p – вероятность битовой ошибки в канале ДСК; $A(w=4)$ – количество разрешённых комбинаций кода с весом $w=4$, число которых для данного кода равно 105 (табл.1.6).

В реальных системах обычно $n > 2^m - 1$, т.е. применяют удлинённый код с расширенной CRC-5.

Проведем анализ, на сколько ухудшится обнаруживающая способность и общая эффективность кода с расширенной CRC-5 из-за возможных двукратных необнаруживаемых ошибок.

Удлинённым кодом не будут обнаружены двукратные ошибки, многочлен которых будет иметь вид:

$$e(x) = x^{i}(1 + x^{\lambda \cdot n_{\text{o}}}),$$

где i и j целые числа, $i \geq 0$,. $\lambda \geq 1, 2, \ldots, s$, но такие, что $(i + \lambda \cdot n_{\text{o}}) \leq n-1$ для всех λ. Величина s определяется из равенства

$$n = s \cdot n_{\text{o}} + r, \qquad (1.16)$$

где r – остаток от деления n на $n_{\text{o}}=2^m-1$, т.е. $n \equiv r \pmod{n_{\text{o}}}$.
Количество таких ошибок точно определяется выражением:

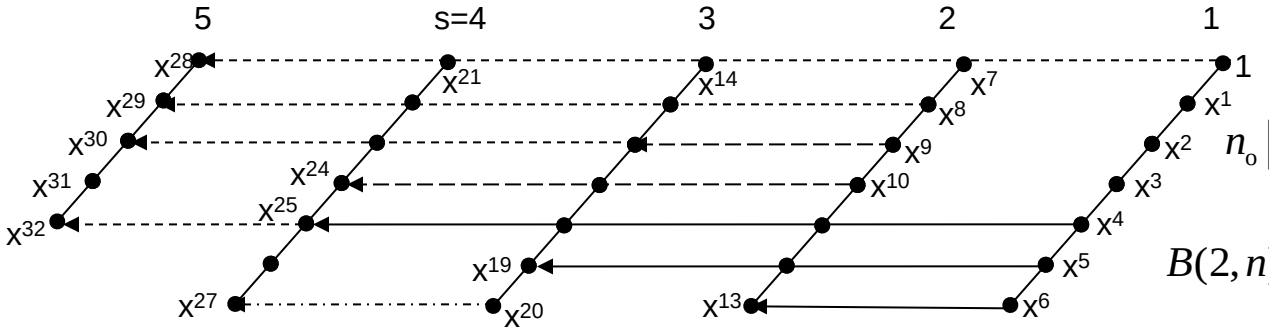
$$B(2, n) = C_s^2 n_{\text{o}} + r \cdot s. \qquad (1.17)$$

Пример 1.3. Рассмотрим в качества примера код (n, k) = (33,29) с расширенной CRC-4 с образующим многочленом

$G(x)=(1+x)\,P(x) = (1+x)(1+x+x^3) = 1 + x^2 + x^3 + x^4$.

Этот код образован на базе кода (n,k)=(32,29) с "простой" CRC-3 и с образующим многочленом $P(x) = 1+x+x^3$, применяемого для передачи кадров в сети SDH [4] в соответствии с рекомендацией **ITU–T G.707**. В комбинации к k=29 информационным элементам добавляются, вместо 3, четыре проверочных элемента, что вызвано введением сомножителя $(1+x)$ в образующий многочлен $G(x)$.

На рис. 1.4 представлена иллюстрация возможных двукратных ошибок, не обнаруживаемых удлинённым кодом с расширенной CRC-4, и поясняющая формулу (1.17). Так как m = 3, то получаем значение $n_o=2^m-1=7$. Тогда из (1.16) находим, что s = 4, а r = 5. На рисунке условно наклонными линиями показаны s = 4 секции по n_o = 7 разрядов n-элементной комбинации. Двукратные не обнаруживаемые кодом ошибки располагаются по горизонтальным линиям. Это будут двукратные ошибки – переходы из точек 1-го наклонного ряда во второй, как, например, переход $x^6 \rightarrow x^{13}$. Таких переходов будет n_o= 7 и каждому из них соответствует двукратная ошибка с многочленом $e(x)=x^i(1+x^7)$, где i принимает значения от 0 до n_o–1 =6. Аналогичные переходы из точек первой наклонной линии в точки 3-й и 4-ой наклонных линий по nо= 7 переходов в каждом случае, как, например, $x^5 \rightarrow x^{19}$ и $x^4 \rightarrow x^{25}$. Таким переходам соответствуют двукратные ошибки с многочленами $e(x)=x^i(1+x^{14})$ и $e(x)=x^i(1+x^{21})$. Всего таких двукратных ошибок – переходов из n_o= 7 точек первого наклонного ряда в точки 2, 3 и 4-го наклонных рядов будет равно $n_o\,(s-1)=n_o\cdot 3=21$.



$$n_o\left[1+2+\ldots+(s-1)\right]=n_o\frac{s(s-1)}{2}=n_oC_s^2.$$

$$B(2,n)=s\cdot r+n_oC_s^2=4\cdot 5+7\cdot\frac{4\cdot 3}{2}=62.$$

Приведем другие рекомендации по применению кодов с расширенными CRC и нулевыми начальными состояниями ячеек регистра деления

2). В сетях с технологией ATM заголовки ATM-ячеек проверяются на наличие в них ошибок, в соответствии с рекомендацией **ITU-T I.432**, систематическим укороченным кодом $(n,k) = (40,32)$ с расширенной CRC-8 и образующим многочленом $G(x)=(1+x)\,P(x)$, где примитивный многочлен $P(x)$ степени 7 имеет вид: $P(x) = 1 + x^2 + x^3 + x^4 + x^5 + x^6 + x^7$.

3). Ещё одним широко применяемым кодом с расширенной CRC-16 с образующим многочленом $G(x)=1+x^2 + x^{15} + x^{16}$, который применяется в протоколе **бинарной синхронной связи BSC фирмы IBM**. Длина кадра канального уровня переменная. Кодирование и декодирование осуществляется по алгоритму систематического циклического (n,k)-кода с нулевыми начальными состояниями ячеек регистра деления.

4). В виртуальных локальных сетях VLAN в соответствии со стандартом **IEEE 802.1Q** применяется (n, k)-код с расширенной CRC-32 с образующим многочленом $G(x) = 1 + x + x^3 + x^5 + x^7 + x^8 + x^{14} + x^{16} + x^{22} + x^{24} + x^{31} + x^{32}$.

5). Наиболее сложным кодом является (n, k)-код с расширенной CRC-64 стандарта **ECMA-182** с образующим многочленом $G(x)$ 64-й степени, имеющим вид в шестнадцатеричной форме записи: $(142F0E1EBA9EA3693)_{16}$, при этом старшая степень слева.