Федеральное государственное бюджетное образовательное учреждение высшего образования

«Сибирский государственный университет телекоммуникаций и информатики»

(СибГУТИ)

Кафедра вычислительных систем

**ОТЧЕТ**

по практической работе 3

по дисциплине «**Программирование**»

|  |  |  |
| --- | --- | --- |
| Выполнил:  студент гр. ИС-241  «25» апреля 2023 г. | \_\_\_\_\_\_\_\_\_\_\_\_ | /Олейников Д.О./ |
|  |  |  |
| Проверил:  Старший преподаватель  «25» апреля 2018 г. | \_\_\_\_\_\_\_\_\_\_\_\_ | /Фульман Д.О./ |

Оценка «\_\_\_\_\_\_\_\_\_\_\_\_\_»

Новосибирск 2023

**ОГЛАВЛЕНИЕ**

[ЗАДАНИЕ 3](#_Toc35593781)

[ВЫПОЛНЕНИЕ РАБОТЫ 4](#_Toc35593782)

[ПРИЛОЖЕНИЕ 5](#_Toc35593783)

# **ЗАДАНИЕ**

# 

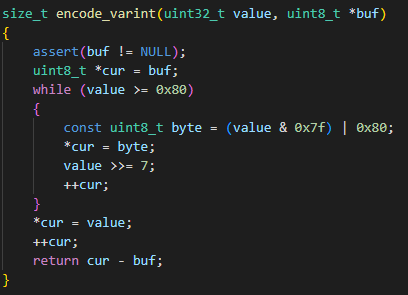
# 

# 

**ВЫПОЛНЕНИЕ РАБОТЫ**

**1 Задание.**

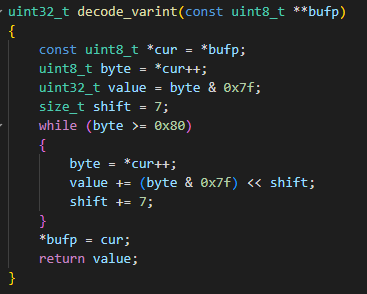
В данном задание нам требуется 3 функции: для кодирования чисел, для декодирование, и для генерации случайного числа.



Функция **encode\_varint** кодирует целочисленное значение value в формате Base 128 Varint и сохраняет результат в буфер buf.

В самом начале идёт проверка, с помощью макроса assert, равен ли buf NULL.

Переменная cur указывает на текущую позицию в buf. В цикле while проверяется, является ли значение value больше или равным 0x80. Если это так, значит, должен записаться ещё один байт в буфер, т.к. каждый байт в формате Base 128 Varint содержит 7 бит данных и 1 бит продолжения. Создаётся byte, который содержит младшие 7 битов значения value и бит продолжения (0x80). Затем записывается этот в текущую позицию cur в буфере buf, сдвигается значение value на 7 бит вправо и увеличивается значение cur на 1. Когда значение value становится меньше, чем 0x80, записывается оставшееся значение value в текущую позицию в буфере buf и увеличивается значение cur на 1.

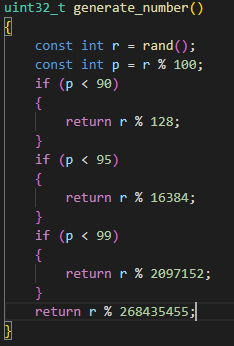


Функция **decode\_varint** декодирует значение Base 128 Varint из буфера bufp и возвращает раскодированное значение.

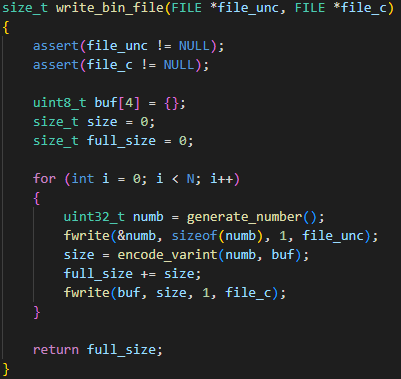
Переменная cur указывает на текущую позицию в буфере bufp. Первый байт в буфере считывается в переменную byte, а младшие 7 битов этого байта сохраняются в переменной value.

Затем цикл while, который продолжается до тех пор, пока значение байта byte имеет установленный старший бит (0x80). Внутри цикла считывается следующий байт из буфера bufp в переменную byte и добавляются младшие 7 битов этого байта к переменной value, сдвигая их на shift битов влево. Затем увеличивается значение shift на 7, чтобы готовиться к чтению следующего байта.

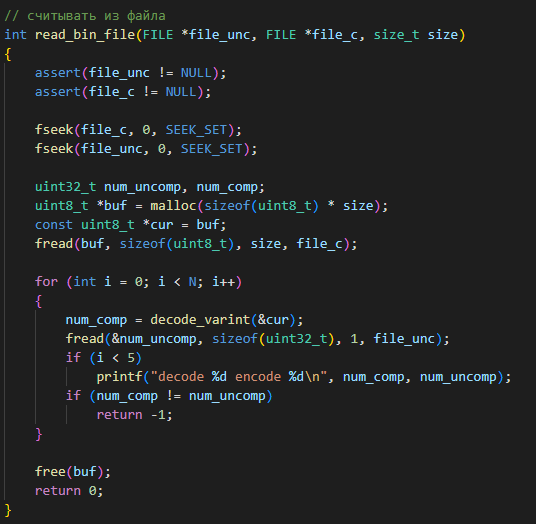
После того, как прочитались все байты, необходимые для кодирования значения value, устанавливается указатель bufp на следующую пустую позицию в буфере, где закончилось чтение. Возвращается раскодированное значение value.



Функция generate\_number генерирует случайное число.



Функция **write\_bin\_file** записывает сгенерированные числа в два файла: в несжатом формате, в формате varint. Макрос assert проверяет не равны ли файлы NULL. Создаётся массив буфер, для передачи данных в функцию кодирования числа. Создаётся цикл для генерации случайных чисел и запись в файлы сначала без кодирования, а затем с кодированием. Возвращается полный размер закодированных чисел.

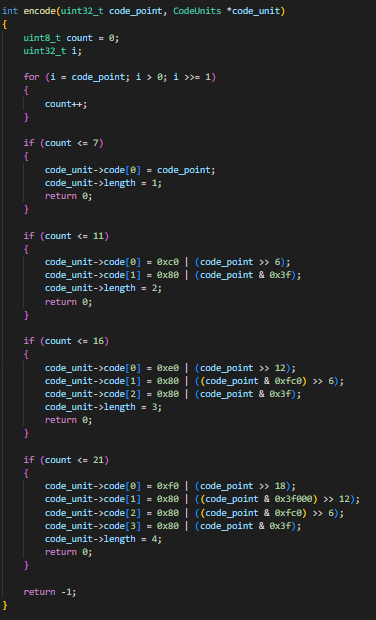


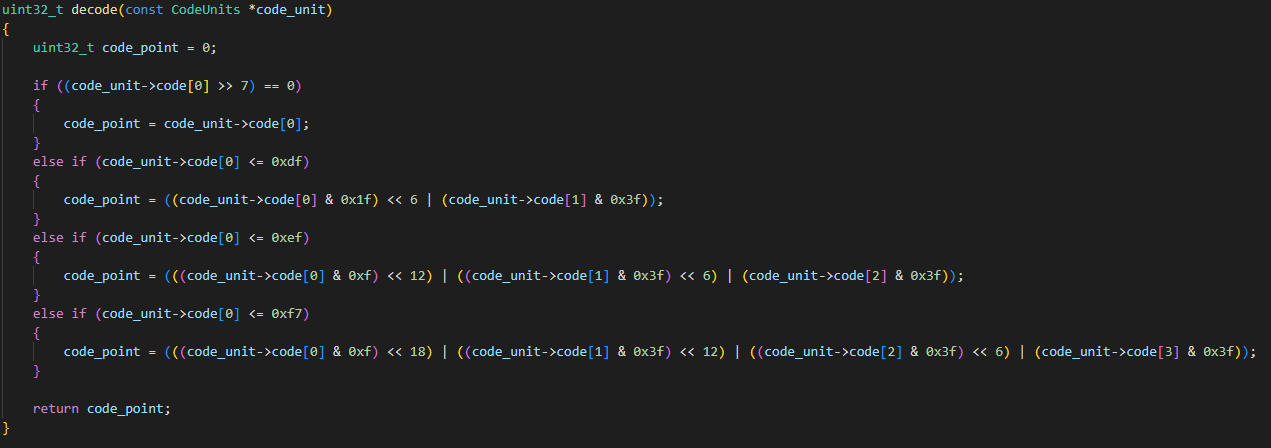
Функция, которая считывает цифры из файлов. Для начала, проверка с помощью макроса assert равны ли файлы NULL. С помощью fseek указатель перемещается на начало файла. Далее выделяется память для раскодирования чисел. Затем в buf считываются данные из кодированного файла. Числа декодируются и выводится 5 сообщений с кодированным значением и декодируемым. В результате buf очищается.



В начале функции main устанавливается таймер, с помощью функции srand и внутри неё функции time, для того, чтобы когда функция выполнялась время было разное и генерировалось разное число. Далее создан указатель на файл, в котором будет храниться указатель на файл, в котором хранятся закодированные цифры. Если файл равен NULL, то выводится сообщение и программа завершается с оишбкой.

**2 Задание**

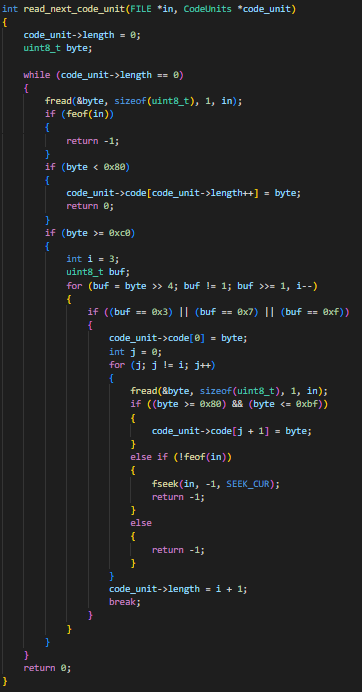
****Функция принимает два параметра: число, которое необходимо закодировать, параметр, который будет являться результатом кодирования. С помощью цикла for считывается количество значащих бит кодируемого числа. После этого функция проверяет, сколько байтов потребуется для кодирования символа, и заполняет массив code в code\_units соответствующими байтами. Если символ может быть закодирован в один байт, то функция сохраняет этот байт в code\_units->code[0] и устанавливает длину code\_units->length в 1. Если байт больше чем 1, то с помощью масок и логических операций кодируются остальные байты, заполняется массив и устанавливается соответствующая длина.



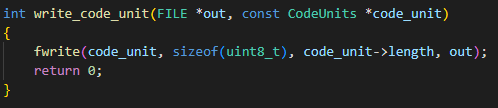
Функция принимает в качестве параметра закодированное представление числа. Есть 4 условия:

1. Если старший байт, сдвинутый вправо на 7 битов, равен нулю, то это число соответствует шаблону в первой строке таблицы, где количество значащих бит кодируемого числа равно 7. В данном случае переменной code\_point присваивается значение старшего байта.
2. Если старший байт меньше, либо равен 0xdf (223), то закодированное число соответствует шаблону во второй строке таблицы, т.к. 0xdf – это 1101 1111 (т.е. максимальное число, которое может соответствовать второму шаблону). В данном случае, сначала с помощью побитового И и сдвига влево на 6 бит избавление от шаблонных битов в старшем байте, затем используется побитовое И для второго байта с числом 0x3f (63), чтобы снова избавиться от шаблонных битов. После этого используется операция побитовое ИЛИ, чтобы «склеить» байты.
3. Если старший байт меньше, либо равен 0xef (239), то закодированное число соответствует шаблону в третьей строке таблицы, т.к. 0xef – это 1110 1111 (т.е. максимальное число, которое соответствовать третьему шаблону). В данном случае сначала с помощью побитового И и сдвигов вправо избавление от шаблонных битов во всех байтах числа. После этого используется операция побитовое ИЛИ, чтобы «склеить» байты.
4. Если старший байт меньше, либо равен 0xf7 (247), то закодированное число соответствует шаблону в четвёртой строке таблицы, т.к. 0xf7 – это 11110111 (т.е. максимальное число, которое может соответствовать четвёртому шаблону).

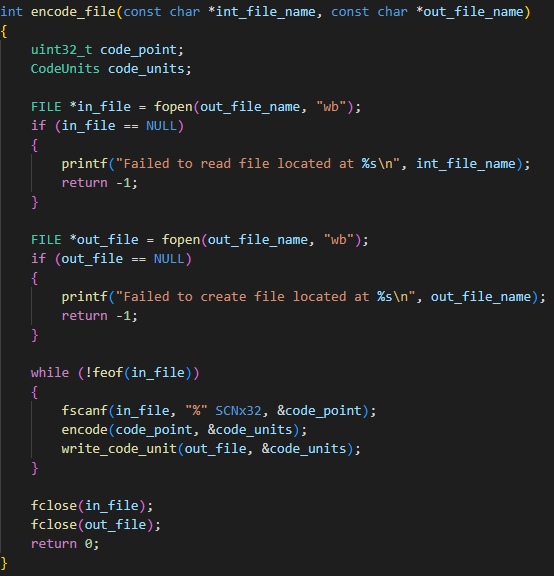
Функция возвращает результат декодирования в виде переменной code\_point.



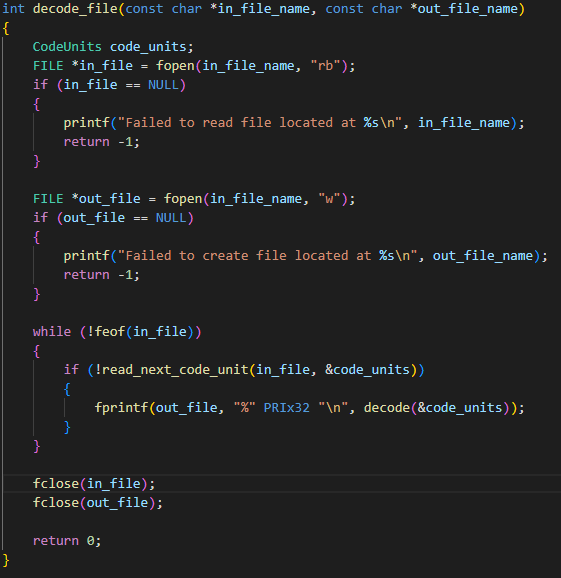
Функция read\_next\_code\_unit принимает несколько параметров: поток файла in и указатель на структуру code\_unit. Сначала функция fread читает следующий байт из файла и сохраняет его в переменную byte, если достигнут конец файла, функция возвращает -1. if(byte < 0x80) – проверяет, является ли byte однобайтовой кодовой единицей. Если является, то он записывается в последовательность кодовых единиц «code\_units», длина последовательности увеличивается на 1 и функция возвращает значение 0. If(byte => 0xc0) – проверяет, является ли byte началом многобайтовой кодовой единицы. Если да, то начинается цикл. Затем начинается цикл, который продолжается до тех пор, пока buf не станет равным единице. Внутри цикла переменная buf сдвигается на один бит вправо, а переменная i уменьшается на единицу.if((buf == 0x3) || (buf == 0x7) || (buf == 0xf)) – проверяет, является ли ‘byte’ началом двухбайтовой, трёхбайтовой или четырёхбайтовой кодовой единицей. Далее записывается первый байт и с помощью ещё одного цикла начинают записываться остальные байты, в 98 строчке проверяется, достигнут ли конец файла. Если нет, то происходит возврат указателя на предыдущий байт с помощью функции fseek, и функция завершается с ошибкой. Если нет, то выполняются следующие условия. В завершение в 105 строчке устанавливается длина массива code на i+1 и выход из внутреннего цикла.



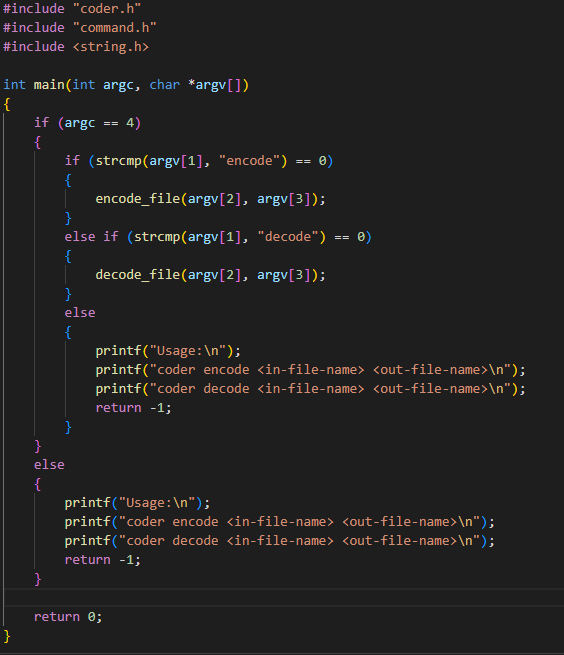
Функция write\_code\_unit записывает последовательность кодовых единиц из структуры CodeUnits в поток out.



Функция encode\_file принимает два параметра: путь к входному текстовому файлу и путь к выходному бинарному файлу. Сначала происходит попытка открыть для чтения входной текстовый файл, в случае неудачи, возвращается -1. Если всё успешно, то происходит попытка создать бинарный файл, в случае неудачи, возвращается -1. Если всё прошло успешно, то далее идёт цикл, который будет работать, пока не достигнет концы файла. В цикле из файла считываются числа и записываются в переменную code\_point. Далее вызывается функция encode, чтобы закодировать считанные числа. После этого вызывается функция write\_code\_unit, чтобы записать в выходной бинарный файл закодированные числа. После завершения цикла, закрывается как текстовый фал, так и бинарный файл. В случае успеха, возвращается 0.



Функция decode\_file принимает два параметра: путь к входному бинарному и путь к выходному текстовому файлу. Сначала попытка открыть для чтения входной бинарный файл, в случае неудачи, возвращается -1. Если всё успешно, то далее попытка создать или открыть уже имеющийся текстовый файл, в случае неудачи возвращается -1. Если всё прошло успешно, то далее идёт цикл, который будет работать, пока не достигнет конца файла. В цикле наблюдается условие, если возвращаемое значение функции read\_next\_code\_unit равно нулю, то записывается в текстовый файл декодированное число. После завершения цикла, бинарный и текстовый файл закрываются. В случае успеха возвращается 0.



В функции main проверяется количество введённых аргументов командной строки, затем с помощью функции strcmp проверяется, какое действие хочется совершить: закодировать или декодировать число. В случае неправильного ввода аргументов, main вернёт -1. В случае успеха 0.

# **ПРИЛОЖЕНИЕ**

**1 Задание**

**main.c**

|  |  |
| --- | --- |
| 1  2 | #include <assert.h>  #include <stddef.h>  #include <stdint.h>  #include <stdio.h>  #include <stdlib.h>  #include <string.h>  #include <time.h>  #define N 1000000  **uint32\_t** **generate\_number**()  {  **const** **int** r = rand();  **const** **int** p = r % **100**;  **if** (p < **90**)  {  **return** r % **128**;  }  **if** (p < **95**)  {  **return** r % **16384**;  }  **if** (p < **99**)  {  **return** r % **2097152**;  }  **return** r % **268435455**;  }  **size\_t** **encode\_varint**(**uint32\_t** value, **uint8\_t** \*buf)  {  assert(buf != NULL);  **uint8\_t** \*cur = buf;  **while** (value >= **0x80**)  {  **const** **uint8\_t** byte = (value & **0x7f**) | **0x80**;  \*cur = byte;  value >>= **7**;  ++cur;  }  \*cur = value;  ++cur;  **return** cur - buf;  }  **uint32\_t** **decode\_varint**(**const** **uint8\_t** \*\*bufp)  {  **const** **uint8\_t** \*cur = \*bufp;  **uint8\_t** byte = \*cur++;  **uint32\_t** value = byte & **0x7f**;  **size\_t** shift = **7**;  **while** (byte >= **0x80**)  {  byte = \*cur++;  value += (byte & **0x7f**) << shift;  shift += **7**;  }  \*bufp = cur;  **return** value;  }  // записывать в файл  **size\_t** **write\_bin\_file**(**FILE** \*file\_unc, **FILE** \*file\_c)  {  assert(file\_unc != NULL);  assert(file\_c != NULL);  **uint8\_t** buf[**4**] = {};  **size\_t** size = **0**;  **size\_t** full\_size = **0**;  **for** (**int** i = **0**; i < N; i++)  {  **uint32\_t** numb = generate\_number();  fwrite(&numb, **sizeof**(numb), **1**, file\_unc);  size = encode\_varint(numb, buf);  full\_size += size;  fwrite(buf, size, **1**, file\_c);  }  **return** full\_size;  }  // считывать из файла  **int** **read\_bin\_file**(**FILE** \*file\_unc, **FILE** \*file\_c, **size\_t** size)  {  assert(file\_unc != NULL);  assert(file\_c != NULL);  fseek(file\_c, **0**, SEEK\_SET);  fseek(file\_unc, **0**, SEEK\_SET);  **uint32\_t** num\_uncomp, num\_comp;  **uint8\_t** \*buf = malloc(**sizeof**(**uint8\_t**) \* size);  **const** **uint8\_t** \*cur = buf;  fread(buf, **sizeof**(**uint8\_t**), size, file\_c);  **for** (**int** i = **0**; i < N; i++)  {  num\_comp = decode\_varint(&cur);  fread(&num\_uncomp, **sizeof**(**uint32\_t**), **1**, file\_unc);  **if** (i < **5**)  printf("decode %d encode %d**\n**", num\_comp, num\_uncomp);  **if** (num\_comp != num\_uncomp)  **return** -**1**;  }  free(buf);  **return** **0**;  }  **int** **main**()  {  srand(time(NULL));  **FILE** \*file\_c;  **if** ((file\_c = fopen("compressed.dat", "wb+")) == NULL)  {  fprintf(stderr, "Can't open compresed.dat**\n**");  exit(EXIT\_FAILURE);  }  **FILE** \*file\_unc;  **if** ((file\_unc = fopen("uncompressed.dat", "wb+")) == NULL)  {  fprintf(stderr, "Can't open uncompressed.dat**\n**");  exit(EXIT\_FAILURE);  }  **size\_t** full\_size = write\_bin\_file(file\_unc, file\_c);  **size\_t** size = ftell(file\_c);  printf("Size with no encode: %d**\n**Size with encode: %ld**\n**",  N \* **4**,  full\_size);  printf("Коэффициент сжатия: %.1f**\n\n**", (**double**)(N \* **4**) / size);  **int** test = read\_bin\_file(file\_unc, file\_c, size);  **if** (test == -**1**)  printf("Последовательность чисел неверна**\n**");  **else**  printf("Последовательность в файлах верна**\n**");  fclose(file\_c);  fclose(file\_unc);  **return** **0**;  } |