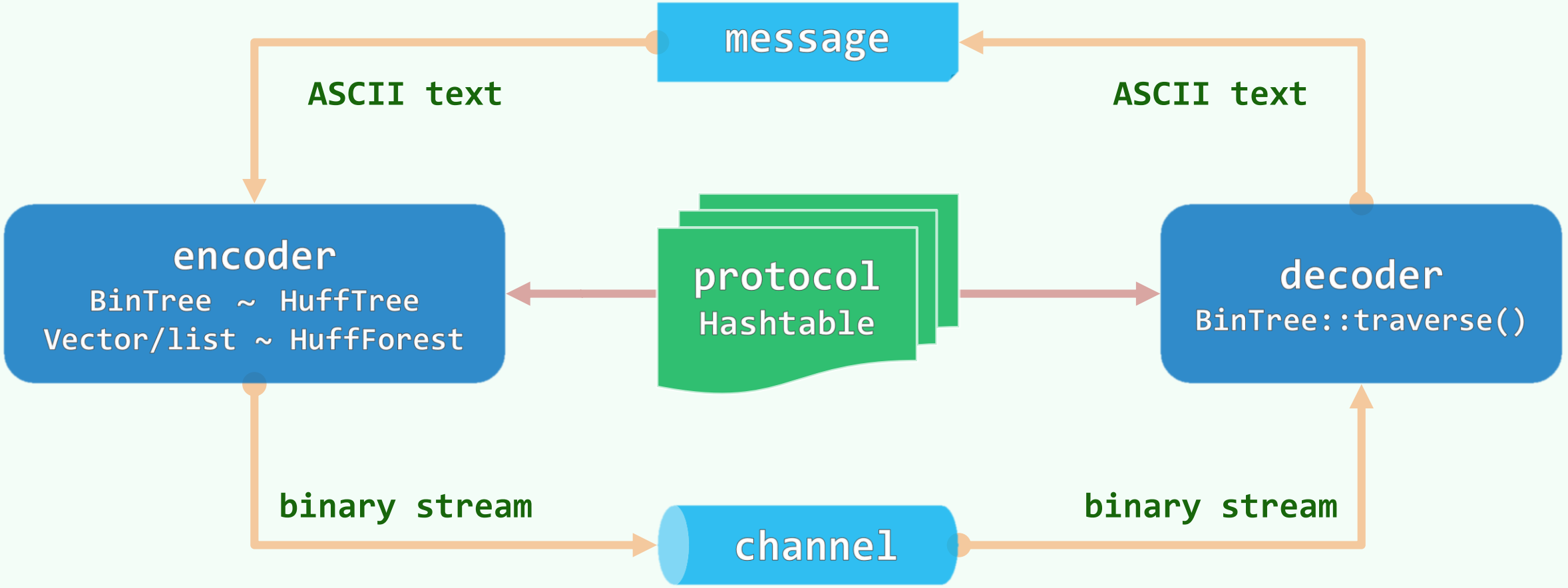
二叉树

Huffman编码树：算法实现

05-13

树形建筑也出现了，看上去规模与地球上的差
不多，只是挂在树上的建筑叶子更为密集

邓俊辉
deng@tsinghua.edu.cn


message
ASCII text
ASCII text
encoder
BinTree ~ HuffTree
Vector/list ~ HuffForest
protocol
Hashtable
decoder
BinTree::traverse()
binary stream
binary stream
channel

Huffman（超）字符

```
#define  N_CHAR  (0x80 - 0x20) //仅以可打印字符为例

struct HuffChar { //Huffman（超）字符
   char ch; unsigned int weight; //字符、频率
   HuffChar ( char c = '^', unsigned int w = 0 ) : ch ( c ), weight ( w ) {};
   bool operator<  ( HuffChar const& hc ) { return weight >  hc.weight; } //比较器
   bool operator== ( HuffChar const& hc ) { return weight == hc.weight; } //判等器
};
```

Huffman树与森林

❖ Huffman（子）树

```
using HuffTree = BinTree< HuffChar >;
```

❖ Huffman森林

```
using HuffForest = List< HuffTree >;
```

❖ 待日后掌握了更多数据结构之后，可改用更为高效的方式，比如：

```
using HuffForest = PQ_List< HuffTree >; //基于列表的优先级队列
using HuffForest = PQ_ComplHeap< HuffTree >; //完全二叉堆
using HuffForest = PQ_LeftHeap< HuffTree >; //左式堆
```

❖ 得益于已定义的统一接口，支撑Huffman算法的这些底层数据结构可直接彼此替换

构造编码树：反复合并二叉树

```
HuffTree* generateTree( HuffForest * forest ) { //Huffman编码算法
   while ( 1 < forest->size() ) { //每迭代一步，森林中都会减少一棵树
      HuffTree T1 = delMax( forest ), T2 = delMax( forest ); //取出权重最小的两棵树
      HuffTree S; //将其合并成一棵新树
      S.insert( HuffChar('^', T1.root()->data.weight + T2.root()->data.weight) );
      S.attach( T2, S.root() ); S.attach( S.root(), T1 ); //T2权重不小于T1
      forest->insertLast( S ); //再插回至森林
   } //森林中最终唯一所剩的那棵树，即Huffman编码树（且其层次遍历序列必然单调非增）
   return forest->first()->data; //故返回之
}
```

遍历森林（List），取出优先级最高（权重最小）的树

```
HuffTree delMax( HuffForest* forest ) {
   ListNodePosi<HuffTree> m = forest->first(); //从首节点出发，遍历所有节点
   for ( ListNodePosi<HuffTree*> p = m->succ; forest->valid( p ); p = p->succ )
      if ( m->data < p->data ) //不断更新（因已定义比较器，故能简捷）
         m = p; //优先级更高（权重更小）者
   return forest->remove( m ); //取出最高者并返回
} //O(n)，改用优先级队列后可做到O(logn)
```

构造编码表：遍历二叉树

```
#include "Hashtable.h" //用HashTable（第09章）实现
using HuffTable = Hashtable< char, char* >; //Huffman编码表
static void generateCT //通过遍历获取各字符的编码
   ( Bitmap* code, int length, HuffTable* table, BinNodePosi<HuffChar> v ) {
   if ( IsLeaf( v ) ) //若是叶节点（还有多种方法可以判断）
      { table->put( v->data.ch, code->bits2string( length ) ); return; }
   if ( HasLChild( v ) ) //Left = 0，深入遍历
      { code->clear( length ); generateCT( code, length + 1, table, v->lc ); }
   if ( HasRChild( v ) ) //Right = 1
      { code->set( length );   generateCT( code, length + 1, table, v->rc ); }
} //总体O(n)
```