

# Re: More of an Algorithms question

---

*Source:* <http://sci.tech-archive.net/Archive/sci.math/2005-11/msg02793.html>

---

- *From:* quasi <[quasi@xxxxxxxx](mailto:quasi@xxxxxxxx)>
  - *Date:* Sat, 19 Nov 2005 22:02:47 -0500
- 

On Sun, 20 Nov 2005 02:04:07 GMT, LC Killingbeck  
<[NOTlynn.killingbeck@xxxxxxxxxxxxxxxxxxxx](mailto:NOTlynn.killingbeck@xxxxxxxxxxxxxxxxxxxx)> wrote:

>quasi <[quasi@xxxxxxxx](mailto:quasi@xxxxxxxx)> wrote in [news:d19qn1pbiauikij1tmslcr6p4e1e2hda5b5@4ax.com](mailto:news:d19qn1pbiauikij1tmslcr6p4e1e2hda5b5@4ax.com):

>

>> On Fri, 18 Nov 2005 00:11:37 GMT, LC Killingbeck  
>> <[NOTlynn.killingbeck@xxxxxxxxxxxxxxxxxxxx](mailto:NOTlynn.killingbeck@xxxxxxxxxxxxxxxxxxxx)> wrote:

>>

>>>quasi <[quasi@xxxxxxxx](mailto:quasi@xxxxxxxx)> wrote in  
>>>[news:t7ipn118h117f4p3sbtetcnsqv7jj250f3@xxxxxxxx](mailto:news:t7ipn118h117f4p3sbtetcnsqv7jj250f3@xxxxxxxx):

>>>

>>>> On Thu, 17 Nov 2005 13:13:06 -0500, quasi <[quasi@xxxxxxxx](mailto:quasi@xxxxxxxx)> wrote:

>>>>

>>>>>On Thu, 17 Nov 2005 17:59:28 +0000 (UTC), Dave Seaman  
>>>>><[dseaman@xxxxxxxxxxxxxxxx](mailto:dseaman@xxxxxxxxxxxxxxxx)> wrote:

>>>>>

>>>>>>On Thu, 17 Nov 2005 12:51:17 -0500, quasi wrote:

>>>>>>> On Thu, 17 Nov 2005 17:21:45 +0000 (UTC), Dave Seaman  
>>>>>>><[dseaman@xxxxxxxxxxxxxxxx](mailto:dseaman@xxxxxxxxxxxxxxxx)> wrote:

>>>>>>>

>>>>>>>>On Thu, 17 Nov 2005 12:15:12 -0500, quasi wrote:

>>>>>>>>> On Thu, 17 Nov 2005 15:07:58 GMT, [cri@xxxxxxxx](mailto:cri@xxxxxxxx) (Richard Harter)  
>>>>>>>>> wrote:

>>>>>>>>>

>>>>>>>>>>>On 15 Nov 2005 14:55:51 -0800, "Filter"  
>>>>>>>>>>><[filtermedialtd@xxxxxxxx](mailto:filtermedialtd@xxxxxxxx)> wrote:

>>>>>>>>>>>

>>>>>>>>>>>>Hi,

>>>>>>>>>>>>

>>>>>>>>>>>>>How can I prove that no Algorithm can compress every file of  
>>>>>>>>>>>>>length  $10^6$ ?

>>>>>>>>>>>>

>>>>>>>>>>>>>>However there is an algorithm that will losslessly compress  
>every

>>>>>>>>>>>>>>file of length  $10^6$  except one.

>>>>>>>>>>>>>>

>>>>>>>>>>>>>>

Re: More of an Algorithms question

>>>>>>>> I don't believe that.  
>>>>>>>>  
>>>>>>>> The pigeonhole principle applies easily to invalidate such a  
>>>>>>>> claim ...  
>>>>>>>>  
>>>>>>>> Let  $n=10^6$ .  
>>>>>>>>  
>>>>>>>> There are  $2^n$  possible files with  $n$  bits, but only  $2^{(n-1)}$  files  
>>>>>>>> with  $n-1$  bits. Hence, at most half the files can be compressed.  
>>>>>>>>  
>>>>>>>> There are  $2^{n-1}$  files with fewer than  $n$  bits. Hence, the claim  
>>>>>>>> stands.  
>>>>>>>>  
>>>>>>>> Not by my count.  
>>>>>>>>  
>>>>>>>> For example, 16 possible files with 4 bits but only 8 possible  
>>>>>>>> files with 3 bits, no?  
>>>>>>>>  
>>>>>>>> Are there any other files with fewer than 4 bits that you have  
>>>>>>>> overlooked in your count?  
>>>>>>>>  
>>>>>>>> Ok, I see my confusion.  
>>>>>>>>  
>>>>>>>> I failed to take into account files with less than  $(n-1)$  bits.  
>>>>>>>>  
>>>>>>>> I get it now -- thanks.  
>>>>>>>>  
>>>>>>>> quasi  
>>>>>>>>  
>>>>>>>> Ok, but then along the lines of what I was thinking, the following  
>>>>>>>> claim is valid:  
>>>>>>>>  
>>>>>>>> In any lossless compression scheme, less than half of the files with  
>>>>>>>>  $n$   
>>>>>>>> or fewer bits can be compressed.  
>>>>>>>>  
>>>>>>>> quasi  
>>>>>>>>  
>>>>>>>> Isn't "less than half" even stronger as "none"? A 1-bit file cannot be  
>>>>>>>> compressed, since the only smaller file (0 bits) is needed to  
>>>>>>>> represent itself. Then a 2-bit file cannot be compressed, since all of  
>>>>>>>> the 0- and 1-bit files are needed to represent themselves. Induction.  
>>>>>>>>  
>>>>>>>> Yes, I saw that the two 1-bit files were a problem as far compression,  
>>>>>>>> but I was assuming that a compression scheme is allowed to compress  
>>>>>>>> some files while potentially expanding others. Thus, a compression  
>>>>>>>> scheme could map some files of  $\leq n$  bits to more than  $n$  bits, and, as  
>>>>>>>> you point out, something like that is necessary. If anything gets  
>>>>>>>> smaller, something has to get bigger -- I guess it's a kind of  
>>>>>>>> conservation law for bits.  
>>>>>>>>

## Re: More of an Algorithms question

>> quasi

>>

>

>Can one conserve the total size, over all files? For example, suppose we  
>have all 5-bit and smaller files, and some 5-bit file compresses to a  
>2-bit file. The original 2-bit file has to become larger – and there is a  
>natural empty slot where the uncompressed 5-bit file came from. The net  
>effect is that the 2-bit and the 5-bit file are simply interchanged, with  
>one compressing and the other growing, while the total size over all  
>files remains constant.

>

>If the above holds, and if there is a theorem that any random  
>permutation of the integers 1,2,3,...,N can be re-ordered (or, even,  
>placed into any other order) by pair-wise interchanges, then the  
>result should be a theorem that the total size can be retained (or,  
>of course made bigger). The truth of the pair-wise swap idea seems  
>obvious: works trivially for N=2; and for N>2 do a swap that moves  
>any number into its final position, reducing the remaining numbers  
>to the same problem with only (N-1) to re-order.

>

>Lynn Killingbeck

Yes, as you suggest, any permutation on the set of n-or-less bit files can be used to define a compression scheme.

Note that if such a compression scheme is used on the full set of n-or-less bit files, less than half of the files will get smaller.

Of course, as I pointed out earlier in this thread, the pigeonhole principle implies that in any compression scheme, less than half of the full set of n-or-less bit files will get smaller.

Also, for the permutation scheme, at least one set must also end up with exactly the same number of bits, so might as well let the permutation fix that one.

quasi

.

---

### • References:

- ◆ **Re: More of an Algorithms question**  
◇ From: Richard Harter
- ◆ **Re: More of an Algorithms question**  
◇ From: quasi
- ◆ **Re: More of an Algorithms question**  
◇ From: Dave Seaman
- ◆ **Re: More of an Algorithms question**  
◇ From: quasi
- ◆ **Re: More of an Algorithms question**

Re: More of an Algorithms question

◇ *From:* Dave Seaman

◆ ***Re: More of an Algorithms question***

◇ *From:* quasi

◆ ***Re: More of an Algorithms question***

◇ *From:* quasi

◆ ***Re: More of an Algorithms question***

◇ *From:* LC Killingbeck

◆ ***Re: More of an Algorithms question***

◇ *From:* quasi

◆ ***Re: More of an Algorithms question***

◇ *From:* LC Killingbeck

- Prev by Date: ***Re: Defining the Fibonacci numbers with only one starting value***
- Next by Date: ***Re: divergence of improper integral implies divergence of series***
- Previous by thread: ***Re: More of an Algorithms question***
- Next by thread: ***Re: More of an Algorithms question***
- Index(es):
  - ◆ ***Date***
  - ◆ ***Thread***