In order to win the prize for most cookies sold, my friend Alice and I are going to merge our Girl Scout Cookies orders and enter as one unit.

Each order is represented by an "order id" (an integer).

We have our lists of orders sorted numerically already, in arrays. Write a function to merge our arrays of orders into one sorted array.

For example:

```c
size_t i;

int myArray[6] = {3, 4, 6, 10, 11, 15};
int alicesArray[6] = {1, 5, 8, 12, 14, 19};

size_t myLength = sizeof(myArray) / sizeof(myArray[0]);
size_t alicesLength = sizeof(alicesArray) / sizeof(alicesArray[0]);

int *mergedArray = mergeArrays(myArray, myLength, alicesArray, alicesLength);
size_t mergedLength = myLength + alicesLength;

assert(mergedArray);

printf("[");
for (i = 0; i < mergedLength; i++) {
    if (i > 0) {
        printf(", ");
    }
    printf("%d", mergedArray[i]);
}
printf("]\n");
// prints [1, 3, 4, 5, 6, 8, 10, 11, 12, 14, 15, 19]
```

Gotchas

We can do this in $O(n)$ time and space.

If you're running a built-in sorting function, your algorithm probably takes $O(n \lg n)$ time for that sort.

Think about edge cases! What happens when we've merged in all of the elements from one of our arrays but we still have elements to merge in from our other array?

Breakdown

We could simply concatenate (join together) the two arrays into one, then sort the result:

```c
int compareInts(const void *x, const void *y)
{
    return *(int *)x - *(int *)y;
}

int * mergeSortedArrays(const int *myArray, size_t myLength,
                        const int *alicesArray, size_t alicesLength)
{
    size_t mergedLength = myLength + alicesLength;
    int *mergedArray = malloc(sizeof(int) * mergedLength);
    assert(mergedArray != NULL);

    memcpy(mergedArray, myArray, sizeof(int) * myLength);
    memcpy(mergedArray + myLength, alicesArray, sizeof(int) * alicesLength);
    qsort(mergedArray, mergedLength, sizeof(int), compareInts);

    return mergedArray;
}
```

What would the time cost be?

$O(n \lg n)$, where n is the total length of our output array (the sum of the lengths of our inputs).

We can do better. With this algorithm, we're not really taking advantage of the fact that the input arrays are themselves *already sorted*. How can we save time by using this fact?

A good general strategy for thinking about an algorithm is to try writing out a sample input and performing the operation by hand. If you're stuck, try that!

Since our arrays are sorted, we know they each have their smallest item in the 0th index. **So the smallest item overall is in the 0th index of one of our input arrays!**

Which 0th element is it? Whichever is smaller!

To start, let's just write a function that chooses the 0*th* element for our sorted array.

```c
int * mergeArrays(const int *myArray, size_t myLength,
                  const int *alicesArray, size_t alicesLength)
{
    // make a array big enough to fit the elements from both arrays
    size_t mergedLength = myLength + alicesLength;
    int *mergedArray = malloc(sizeof(int) * mergedLength);
    assert(mergedArray != NULL);

    int headOfMyArray = myArray[0];
    int headOfAlicesArray = alicesArray[0];

    // case: 0th comes from my array
    if (headOfMyArray < headOfAlicesArray) {
        mergedArray[0] = headOfMyArray;

    // case: 0th comes from Alice's array
    }
    else {
        mergedArray[0] = headOfAlicesArray;
    }

    // eventually we'll want to return the merged array
    return mergedArray;
}
```

Okay, good start! That works for finding the 0th element. Now how do we choose the next element?

Let's look at a sample input:

```c
[3,  4,  6, 10, 11, 15]  // myArray
[1,  5,  8, 12, 14, 19]  // alicesArray
```

To start we took the 0th element from `alicesArray` and put it in the 0th slot in the output array:

```c
[3,  4,  6, 10, 11, 15]  // myArray
[1,  5,  8, 12, 14, 19]  // alicesArray
[1,  x,  x,  x,  x,  x]  // mergedArray
```

We need to make sure we don't try to put that 1 in `mergedArray` again. We should mark it as "already merged" somehow. For now, we can just cross it out:

```
[3,  4,  6, 10, 11, 15]  // myArray
[x,  5,  8, 12, 14, 19]  // alicesArray
[1,  x,  x,  x,  x,  x]  // mergedArray
```

Or we could even imagine it's removed from the array:

```
[3,  4,  6, 10, 11, 15]  // myArray
[5,  8, 12, 14, 19]      // alicesArray
[1,  x,  x,  x,  x,  x]  // mergedArray
```

Now to get our next element we can use the same approach we used to get the 0th element—it's the smallest of the *earliest unmerged elements* in either array! In other words, it's the smaller of the leftmost elements in either array, assuming we've removed the elements we've already merged in.

So in general we could say something like:

1. We'll start at the beginnings of our input arrays, since the smallest elements will be there.
2. As we put items in our final `mergedArray`, we'll keep track of the fact that they're "already merged."
3. At each step, each array has a *first* "not-yet-merged" item.
4. At each step, the next item to put in the `mergedArray` is the smaller of those two "not-yet-merged" items!

Can you implement this in code?

C

```
int * mergeArrays(const int *myArray, size_t myLength,
                  const int *alicesArray, size_t alicesLength)
{
    size_t mergedLength = myLength + alicesLength;
    int *mergedArray = malloc(sizeof(int) * mergedLength);
    assert(mergedArray != NULL);

    size_t currentIndexAlices = 0;
    size_t currentIndexMine   = 0;
    size_t currentIndexMerged = 0;

    while (currentIndexMerged < mergedLength) {
        int firstUnmergedAlices = alicesArray[currentIndexAlices];
        int firstUnmergedMine = myArray[currentIndexMine];

        // case: next comes from my array
        if (firstUnmergedMine < firstUnmergedAlices) {
            mergedArray[currentIndexMerged] = firstUnmergedMine;
            currentIndexMine++;

        // case: next comes from Alice's array
        }
        else {
            mergedArray[currentIndexMerged] = firstUnmergedAlices;
            currentIndexAlices++;
        }

        currentIndexMerged++;
    }

    return mergedArray;
}
```

Okay, this algorithm makes sense. To wrap up, we should think about edge cases and check for bugs. What edge cases should we worry about?

Here are some edge cases:

1. One or both of our input arrays is 0 elements or 1 element
2. One of our input arrays is longer than the other.

3. One of our arrays runs out of elements before we're done merging.

Actually, (3) will *always* happen. In the process of merging our arrays, we'll certainly exhaust one before we exhaust the other.

Does our function handle these cases correctly?

If both arrays are empty, we're fine. But for all other edge cases, we'll get `a segfault`.

How can we fix this?

We can probably solve these cases at the same time. They're not so different—they just have to do with indexing past the end of arrays.

To start, we could treat each of our arrays being out of elements as a separate case to handle, in addition to the 2 cases we already have. So we have 4 cases total. Can you code that up?

Be sure you check the cases in the right order!

C

```
int * mergeArrays(const int *myArray, size_t myLength,
                  const int *alicesArray, size_t alicesLength)
{
    // case: both arrays are zero elements
    if (myLength == 0 && alicesLength == 0) {
        return NULL;
    }

    size_t mergedLength = myLength + alicesLength;
    int *mergedArray = malloc(sizeof(int) * mergedLength);
    assert(mergedArray != NULL);

    size_t currentIndexAlices = 0;
    size_t currentIndexMine   = 0;
    size_t currentIndexMerged = 0;

    while (currentIndexMerged < mergedLength) {
        int firstUnmergedAlices = alicesArray[currentIndexAlices];
        int firstUnmergedMine = myArray[currentIndexMine];

        // case: my array is exhausted
        if (currentIndexMine >= myLength) {
            mergedArray[currentIndexMerged] = alicesArray[currentIndexAlices];
            currentIndexAlices++;

        // case: Alice's array is exhausted
        }
        else if (currentIndexAlices >= alicesLength) {
            mergedArray[currentIndexMerged] = myArray[currentIndexMine];
            currentIndexMine++;

        // case: next comes from my array
        }
        else if (firstUnmergedMine < firstUnmergedAlices) {
            mergedArray[currentIndexMerged] = firstUnmergedMine;
            currentIndexMine++;

        // case: next comes from Alice's array
        }
        else {
```

```
            mergedArray[currentIndexMerged] = firstUnmergedAlices;
            currentIndexAlices++;
        }

        currentIndexMerged++;
    }

    return mergedArray;
}
```

Cool. This'll work, but it's a bit repetitive. We have these two lines twice:

```
mergedArray[currentIndexMerged] = myArray[currentIndexMine];
currentIndexMine++;
```

Same for these two lines:

```
mergedArray[currentIndexMerged] = alicesArray[currentIndexAlices];
currentIndexAlices++;
```

That's not DRY. Maybe we can avoid repeating ourselves by bringing our code back down to just 2 cases.

See if you can do this in just one "if else" by combining the conditionals.

You might try to simply squish the middle cases together:

```
if (isAlicesArrayExhausted
        || (myArray[currentIndexMine] < alicesArray[currentIndexAlices])) {

    mergedArray[currentIndexMerged] = myArray[currentIndexMine];
    currentIndexMine++;
```

But what happens when `myArray` is exhausted?

We'll get a `segfault` when we try to access `myArray[currentIndexMine]`!

How can we fix this?

Solution

First, we allocate our answer array, getting its size by adding the size of `myArray` and `alicesArray`.

We keep track of a current index in `myArray`, a current index in `alicesArray`, and a current index in `mergedArray`. So at each step, there's a "current item" in `alicesArray` and in `myArray`. The smaller of those is the next one we add to the `mergedArray`!

But careful: we also need to account for the case where we exhaust one of our arrays and there are still elements in the other. To handle this, we say that the current item in `myArray` is the next item to add to `mergedArray` only if `myArray` is *not* exhausted AND, either:

1. `alicesArray` is exhausted, or
2. the current item in `myArray` is less than the current item in `alicesArray`

C

```
int * mergeArrays(const int *myArray, size_t myLength,
                  const int *alicesArray, size_t alicesLength)
{
    // case: both arrays are zero elements
    if (myLength == 0 && alicesLength == 0) {
        return NULL;
    }

    size_t mergedLength = myLength + alicesLength;
    int *mergedArray = malloc(sizeof(int) * mergedLength);
    assert(mergedArray != NULL);

    size_t currentIndexAlices = 0;
    size_t currentIndexMine   = 0;
    size_t currentIndexMerged = 0;

    while (currentIndexMerged < mergedLength) {
        int isMyArrayExhausted = currentIndexMine >= myLength;
        int isAlicesArrayExhausted = currentIndexAlices >= alicesLength;

        // case: next comes from my array
        // my array must not be exhausted, and EITHER:
        // 1) Alice's array IS exhausted, or
        // 2) the current element in my array is less
        //    than the current element in Alice's array
        if (!isMyArrayExhausted && (isAlicesArrayExhausted
                || (myArray[currentIndexMine] < alicesArray[currentIndexAlices]))) {

            mergedArray[currentIndexMerged] = myArray[currentIndexMine];
            currentIndexMine++;

        // case: next comes from Alice's array
        }
        else {
            mergedArray[currentIndexMerged] = alicesArray[currentIndexAlices];
            currentIndexAlices++;
        }

        currentIndexMerged++;
    }
```

```
    return mergedArray;
}
```

The if statement is carefully constructed to avoid undefined behavior from indexing past the end of the array. We take advantage of C short circuit evaluation and check *first* if we've reached the end of the arrays.

Complexity

$O(n)$ time and $O(n)$ additional space, where n is the number of items in the merged array.

The added space comes from allocating the `mergedArray`. There's no way to do this " in place" because neither of our input arrays are necessarily big enough to hold the merged array.

But if our inputs were linked lists, we could avoid allocating a new structure and do the merge by simply adjusting the `next` pointers in the list nodes!

In our implementation above, we could avoid tracking `currentIndexMerged` and just compute it on the fly by adding `currentIndexMine` and `currentIndexAlices`. This would only save us one integer of space though, which is hardly anything. It's probably not worth the added code complexity.

Bonus

What if we wanted to merge *several* sorted arrays? Write a function that takes as an input *an array of sorted arrays* and outputs a single sorted array with all the items from each array.

Do we absolutely have to allocate a new array to use for the merged output? Where else could we store our merged array? How would our function need to change?

What We Learned

We spent a lot of time figuring out how to cleanly handle edge cases.

Sometimes it's easy to lose steam at the end of a coding interview when you're debugging. But keep sprinting through to the finish! Think about edge cases. Look for off-by-one errors.

⬅ course home (/table-of-contents)

Next up: Cafe Order Checker ➡ (/question/cafe-order-checker?course=fc1&section=array-and-string-manipulation)

Want more coding interview help?

Check out **interviewcake.com** for more advice, guides, and practice questions.