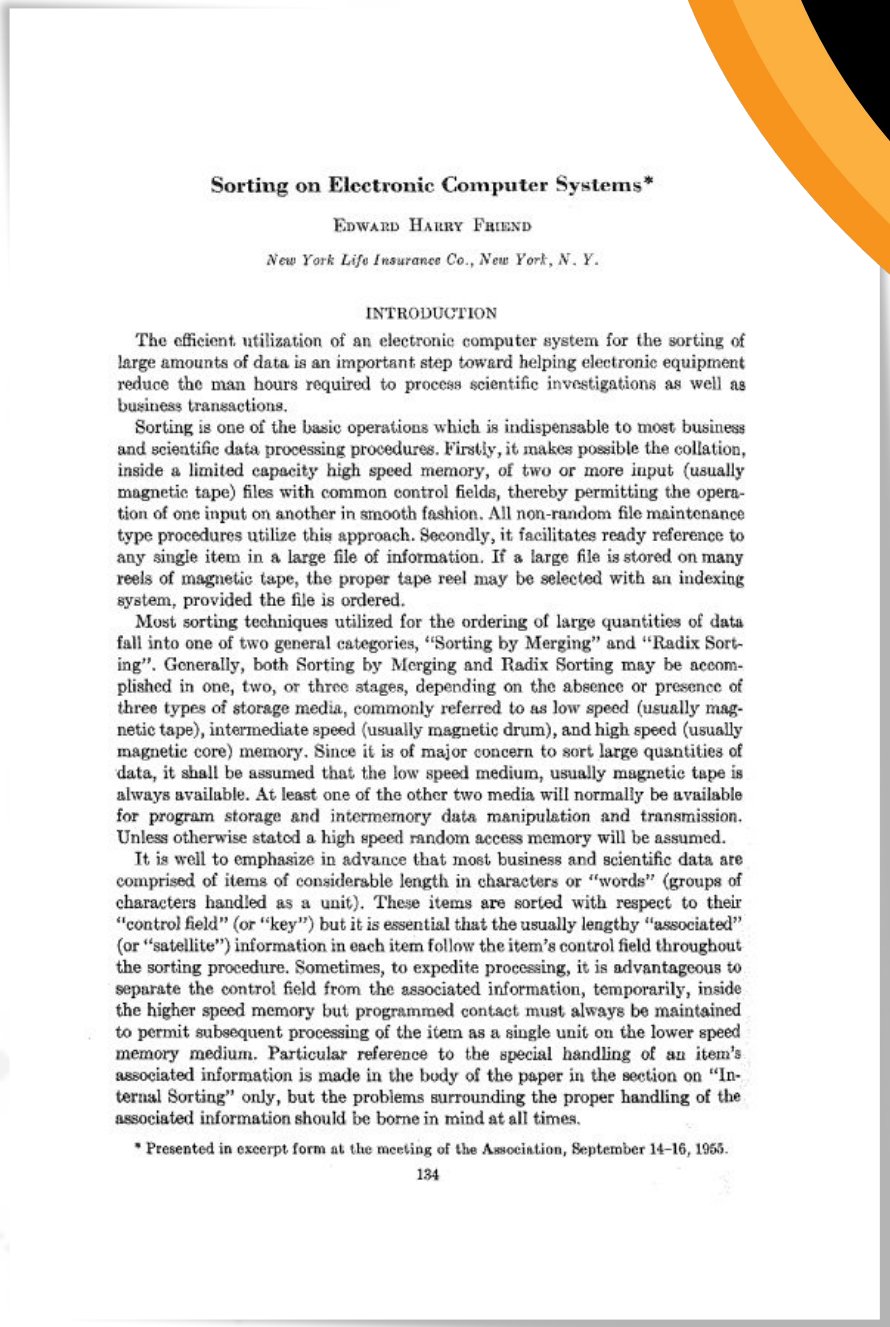
Sorting Algorithm Complexity in Response to Varying Sizes of Data Sets

Harry Haisty

The History of Bubble Sort

- *Sorting on Electronic Computer Systems* – 1956
- Association for Computing Machinery, founded 1947
- First documented proposal was by Edward Harry Friend, listed as "Exchange Sort."
- Complexity n^2



Sorting on Electronic Computer Systems*

EDWARD HARRY FRIEND

New York Life Insurance Co., New York, N. Y.

INTRODUCTION

The efficient utilization of an electronic computer system for the sorting of large amounts of data is an important step toward helping electronic equipment reduce the man hours required to process scientific investigations as well as business transactions.

Sorting is one of the basic operations which is indispensable to most business and scientific data processing procedures. Firstly, it makes possible the collation, inside a limited capacity high speed memory, of two or more input (usually magnetic tape) files with common control fields, thereby permitting the operation of one input on another in smooth fashion. All non-random file maintenance type procedures utilize this approach. Secondly, it facilitates ready reference to any single item in a large file of information. If a large file is stored on many reels of magnetic tape, the proper tape reel may be selected with an indexing system, provided the file is ordered.

Most sorting techniques utilized for the ordering of large quantities of data fall into one of two general categories, "Sorting by Merging" and "Radix Sorting". Generally, both Sorting by Merging and Radix Sorting may be accomplished in one, two, or three stages, depending on the absence or presence of three types of storage media, commonly referred to as low speed (usually magnetic tape), intermediate speed (usually magnetic drum), and high speed (usually magnetic core) memory. Since it is of major concern to sort large quantities of data, it shall be assumed that the low speed medium, usually magnetic tape is always available. At least one of the other two media will normally be available for program storage and intermemory data manipulation and transmission. Unless otherwise stated a high speed random access memory will be assumed.

It is well to emphasize in advance that most business and scientific data are comprised of items of considerable length in characters or "words" (groups of characters handled as a unit). These items are sorted with respect to their "control field" (or "key") but it is essential that the usually lengthy "associated" (or "satellite") information in each item follow the item's control field throughout the sorting procedure. Sometimes, to expedite processing, it is advantageous to separate the control field from the associated information, temporarily, inside the higher speed memory but programmed contact must always be maintained to permit subsequent processing of the item as a single unit on the lower speed memory medium. Particular reference to the special handling of an item's associated information is made in the body of the paper in the section on "Internal Sorting" only, but the problems surrounding the proper handling of the associated information should be borne in mind at all times.

* Presented in excerpt form at the meeting of the Association, September 14–16, 1955.

Visualization of Bubble Sort

Usage of Bubble Sort Today

- Most people don't use Bubble Sort

6/1/2015

The worst sorting algorithm by far, incredibly slow, doesn't work for what you need. I will be taking my data sets elsewhere.

6/1/2015

Bubble sort killed my family.

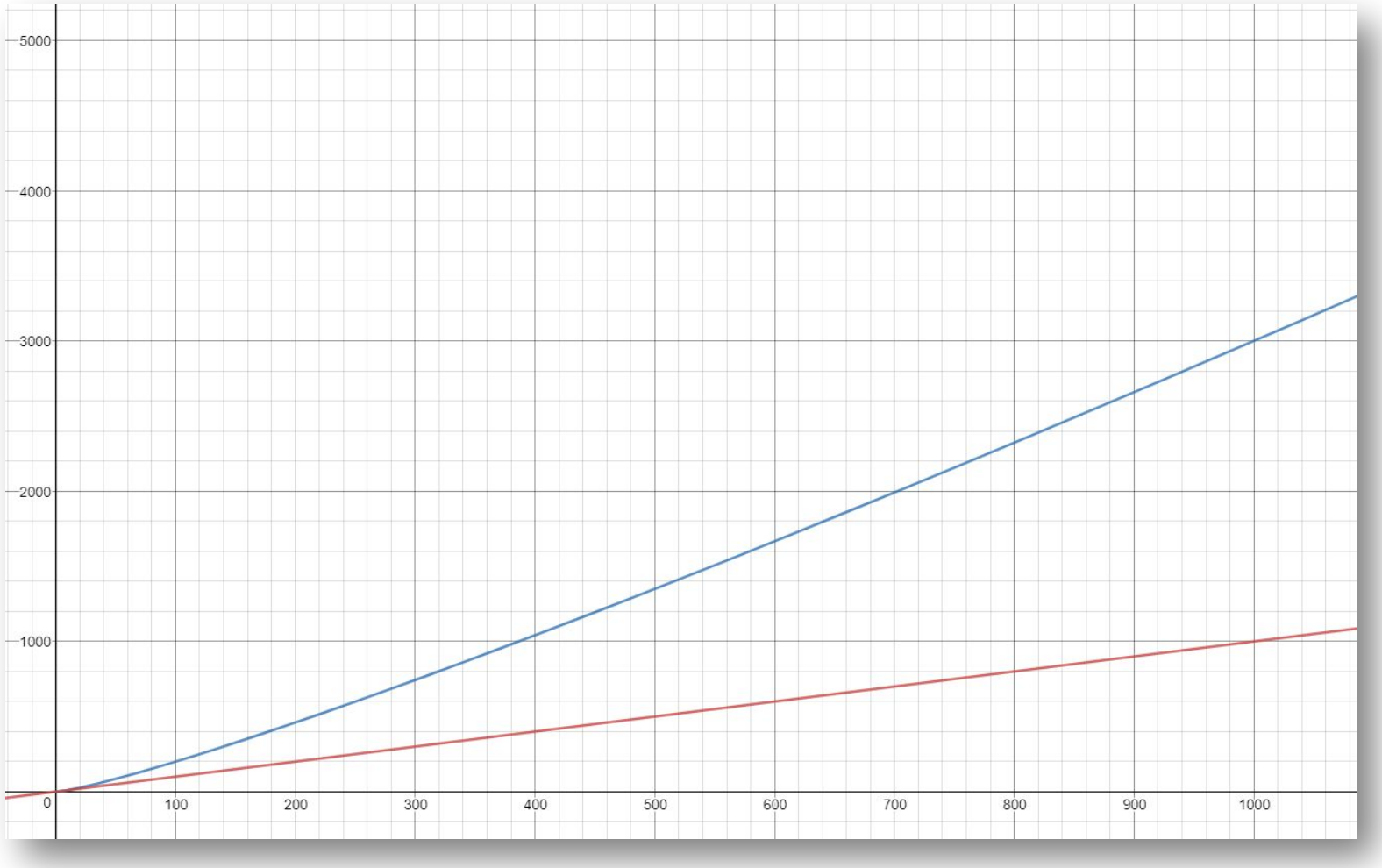
The History of Merge Sort

- Attributed to John von Neumann
- Breaks problem into smaller, more manageable problems (aka “divide and conquer”)
- One of the earliest sorting algorithms proposed for a computer, 1947

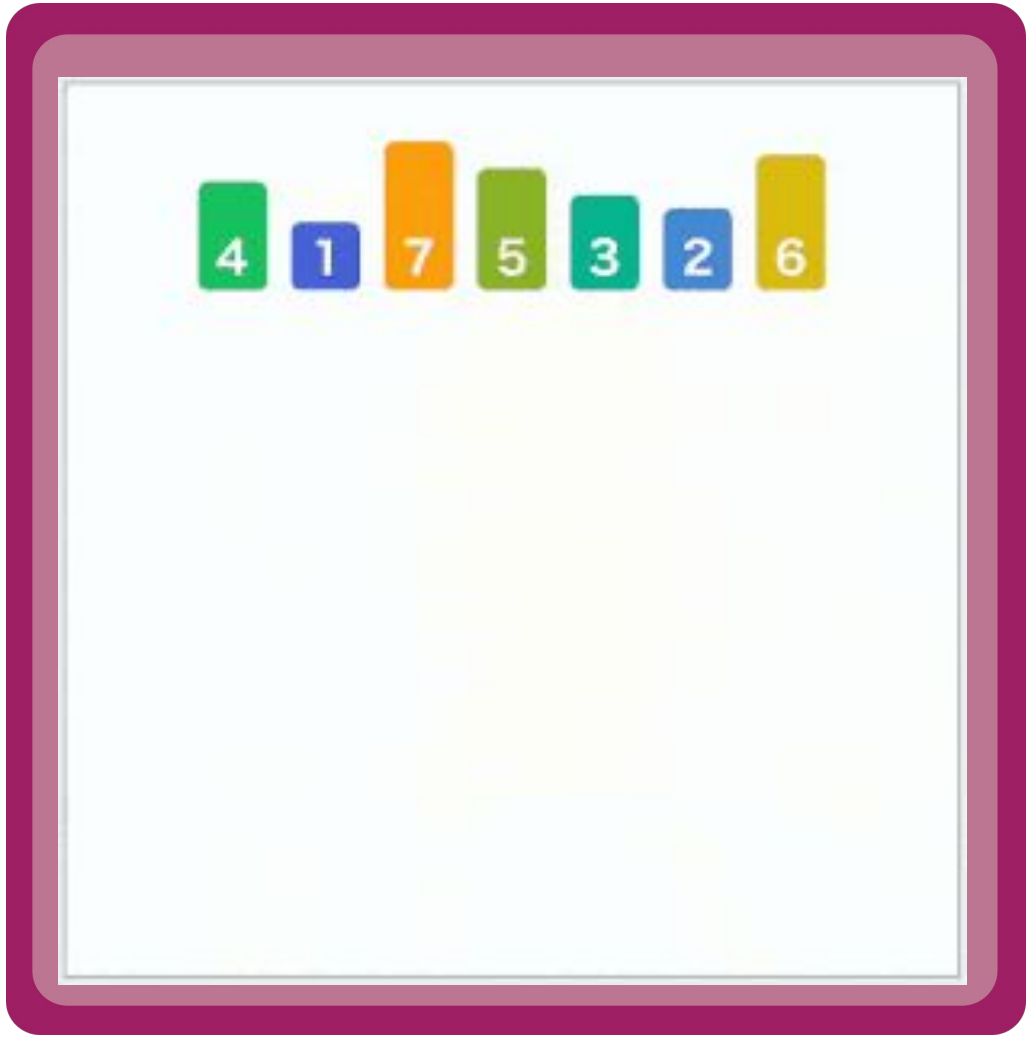
John von Neumann looking fierce

Time Complexity

- The best case scenario for Merge Sort's time complexity can be measured as O(n)
- The worst case scenario for Merge Sort's time complexity can be measured as O($n$log$n$)



Visualization of Merge Sort



Now it's your turn!

- I need 8 volunteers from the class.
- Please draw a number from the receptacle.