An Historical Survey of Early Real-Time Computing Developments in the U.S.

PHILLIP A. LAPLANTE laplante@fdu.edu
Department of Mathematics and Computer Science, Fairleigh Dickinson University, Madison, NJ 07940

EILEEN P. ROSE epf@mtgzfs3.mt.att.com
MAREA GRACIA-WATSON mig@mtnms.att.com
AT&T Bell Laboratories, Middletown, NJ 07746

Abstract. In this paper the development of real-time computing terms, systems, hardware, and software from the 1940s through the 1960s in the United States is examined. In addition, the contributions of significant individuals, corporations, and government through the projects they fostered during these early years are surveyed. And in particular it is contended that Project Whirlwind, a post-World War II flight simulator, was the first American real-time system.

1. Introduction

A real-time system is one whose logical correctness depends on the correctness of its outputs as well as their timeliness. Inherent in this definition is the notion of control – the system under real-time control must be both deterministic and stable. The first definitions of “real-time” emerged in the 1950s. For example, Dinneen stated “a digital computer is said to operate in ‘real-time’ when it is an integral part of a physical control system. One of the requirements for real-time operation of a digital computer is rapid computation consistent with changes in the physical inputs and the output data rates required by the system” (Dinneen, Lebow, and Reed, 1958). During the 1960s, the term “real-time” became part of everyday computing parlance even as more precise definitions of “real-time” emerged. All known definitions shared common ideas, but not one mentioned “simulating real-response times”, a concept that was so important to earlier systems such as Whirlwind and SAGE. Robert Head defined “real-time” as, “paralleling data processing with a physical process in such a fashion that the results of the data processing are immediately useful to the physical operation” (Head, 1964). This definition singles out the feedback characteristic of a real-time system as the predominant one by requiring that system results be available in sufficient time to affect the external process. This incorporates an aspect often seen in real-time systems, that is direct control of the environment via immediately useful results.

Desmonde defined the notion of “real-time” as usually implying that “a central processor responded to inputs within a very short time, perhaps a few seconds. If the duration of time between the creation of information and the computer response was longer than a matter of hours, most people did not regard the system as ‘real-time’” (Desmonde, 1964). Using Desmonde’s definition, it is difficult to establish a time crite-
ration to determine whether or not the system was "real-time", since the range of acceptable
response times was so large. The deciding factor was usually how soon the computer
took action in the situation it was controlling. Nevertheless, the notion of a "bounded
response time" in a controlling situation seems to have been introduced here.

Evidently, the definitions were slowly evolving into the definition of real-time used to-
day. The most often-cited early definition for real-time was given by Martin (Martin, 1965)
who stated, "a real-time computer system may be defined as one which controls an envi-
ronment by receiving data, processing them and returning the results sufficiently quickly
to affect the functioning of the environment at that time." With this definition of "real-
time", in order to "control", an environment a short response time for the computer is
necessary. And the response time can differ from machine to machine depending on the
application. For example, while a radar scanning response time must be in milliseconds,
an airline reservation system response time may be several seconds (Martin, 1965). Al-
though this definition again includes the notion of "quick response and controlling the
environment", it goes further to imply that all computers are in a sense "real-time" a
notion which many still contend today. It suggests that response times must be bounded,
and the range is a function of the application the computer is serving. Martin's definition
seems to come closest to our present day definition of "real-time".

2. Developments in the 1940s

Let us begin our historical tour with the Second World War, which profoundly influenced
technological progress. Most of the U.S. was devoted to the war effort, and the concept of
"business as usual" was suspended. The war necessitated improvement of computational
systems needed for: the preparation of ballistics tables, aircraft systems, atomic weapons
design, fire control, and logistics, and the U.S. government felt compelled to allocate
substantial resources to these problems (Holbrook, 1982). Major corporations such as
Bell Telephone Laboratories, were also devoting large amounts of time and money in
support of the war effort. In addition, many of the computing needs identified during
the war provided the research problems of the late 1940s and 1950s, which led to
crucial developments in real-time systems. Finally, the developments of this decade
were influenced by the contributions of many, notably John von Neumann, J. Presper
Eckert, Jr., John William Mauchly, Jay W. Forrester, and other scientists at Bell Telephone
Laboratories, and the Massachusetts Institute of Technology (MIT).

While systems of this era exhibited certain real-time characteristics, such as expected
response time, they were still not classified as "real-time". However, it is our belief
that Project Whirlwind, which was started in the late 1940s, was the first "real-time"
computer system. We also believe that the term, "real-time", was coined during this
project's development.