

String Deduplication During Garbage Collection in Virtual Machines

Konstantin Nasartschuk, Kenneth B. Kent,
Aleksandar Micic, Charlie Gracie

University of New Brunswick, IBM Canada
Faculty of Computer Science

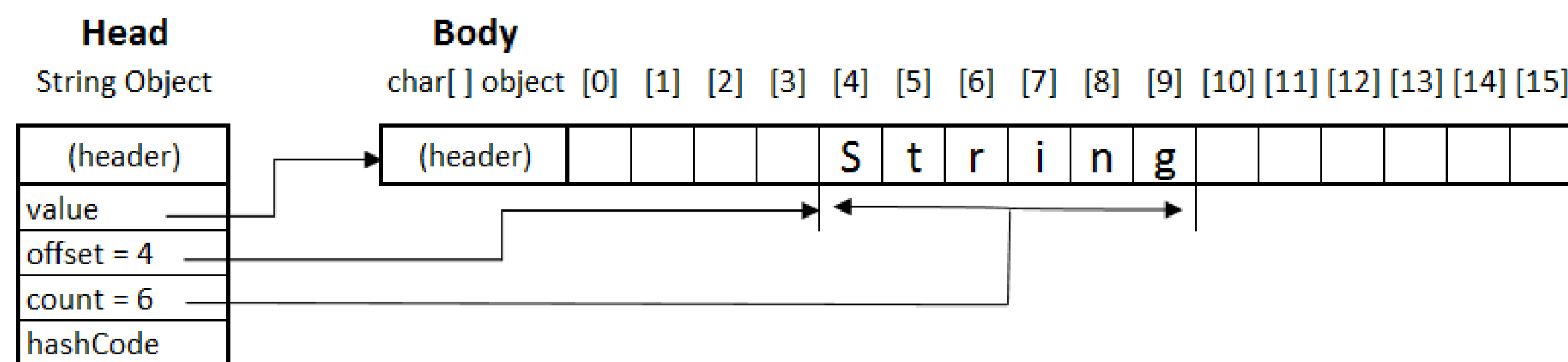
{kons.na, ken}@unb.ca, {aleksandar_micic, charlie_gracie}@ca.ibm.com

ABSTRACT

Duplicate strings stored on the heap of a virtual machine still offer an opportunity for heap use decrease and GC time reduction. The approach presented utilizes the stop-the-world phase of the garbage collector to detect duplicate strings and to deduplicate them before they are evacuated. The goal of this algorithm is to reduce memory duplication and to prevent unnecessary copying of memory.

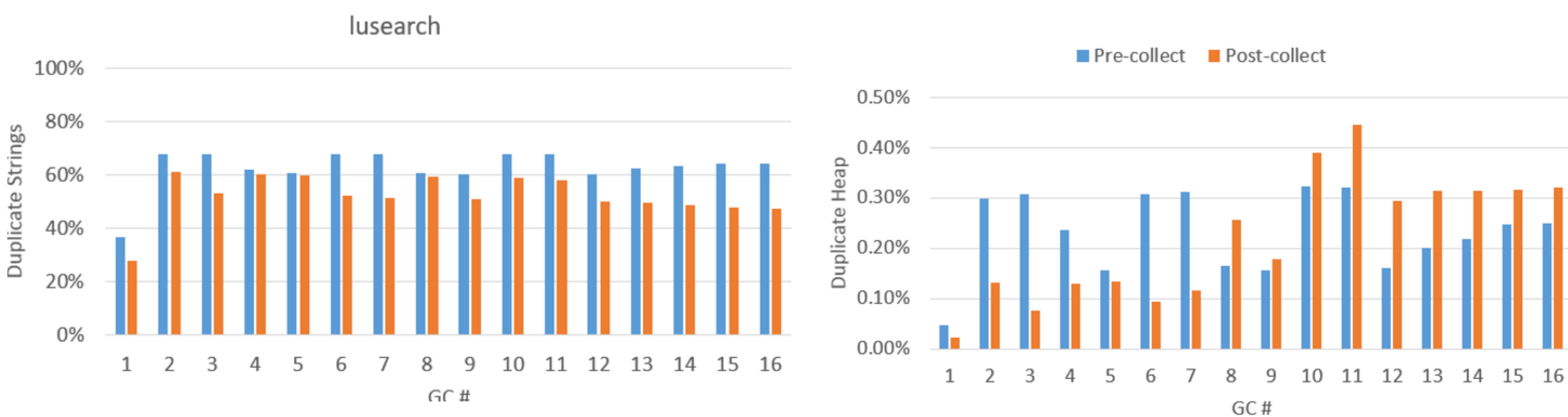
INTRODUCTION

- String objects in Java are immutable
- Strings in Java can occupy up to 40% of the heap
- Deduplication possible by pointing multiple String headers to one Char array



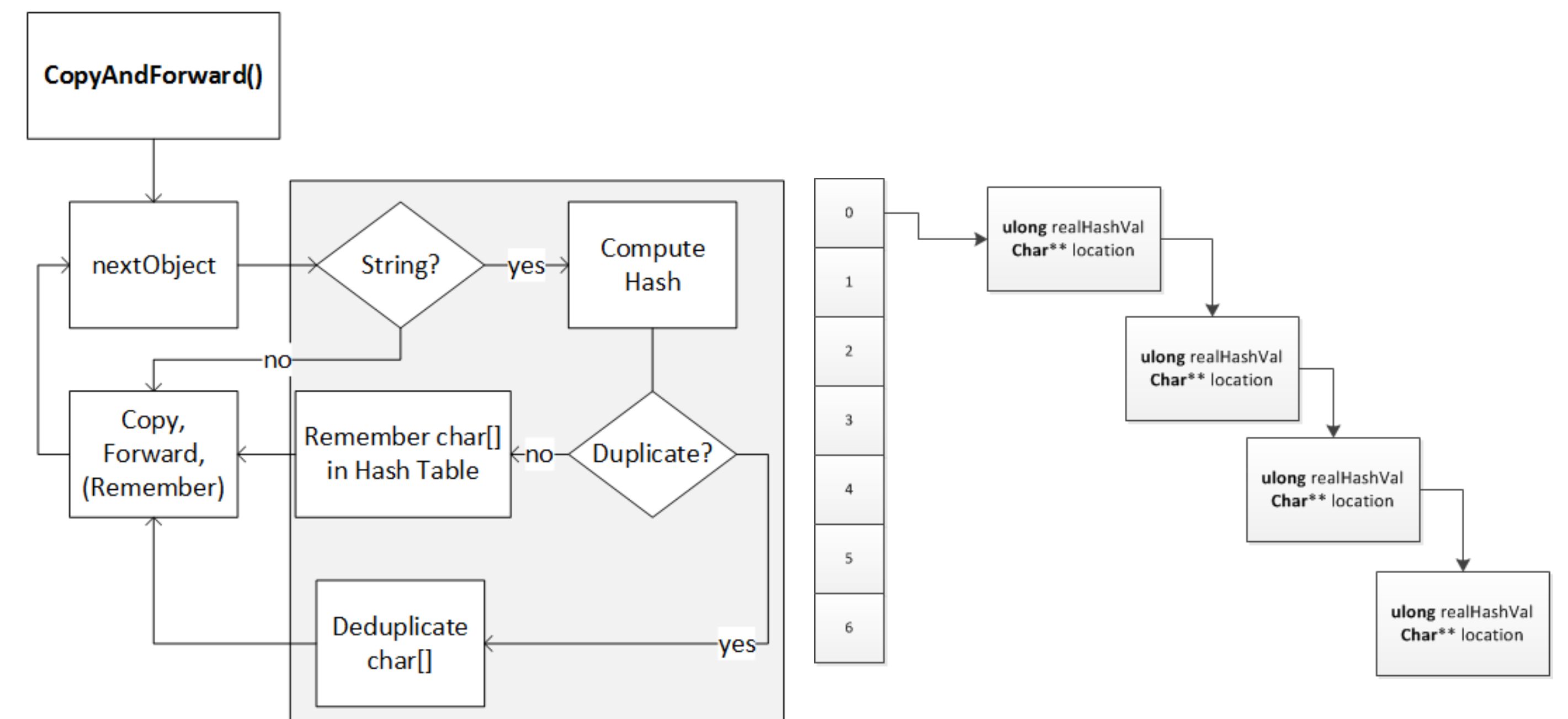
MOTIVATION

- Up to 40% of strings candidates
- Up to 2% of heap can consist of duplicate strings



APPROACH

- Use existing stop-the-world-phase
- Remember strings encountered in hash table
- Deduplicate before objects are copied if possible



RESULTS

- A: No SD, No hash table
- B: No SD, Hash table
- C: SD, Hash table

Benchmark	B	C
avroa	-1.07%	-0.09%
batik	+0.42%	+1.13%
fop	+1.89%	-1.12%
h2	+2.29%	-0.18%
kython	+2.53%	+2.01%
luindex	-0.04%	+0.17%
lusearch	+8.58%	-3.36%
pmd	+2.07%	-0.29%
sunflow	+13.31%	-2.66%
tomcat	+1.97%	+2.71%
xalan	+7.57%	+4.23%

