COMPRESSING JAVA BINARIES

DENIS N. ANTONIOLI

antonioli@ifi.unizh.ch
<http://www.ifi.unizh.ch/~antonio>

February 1st, 2001

DEPARTMENT OF INFORMATION TECHNOLOGY,
UNIVERSITY OF ZURICH

ifi

MOTIVATION

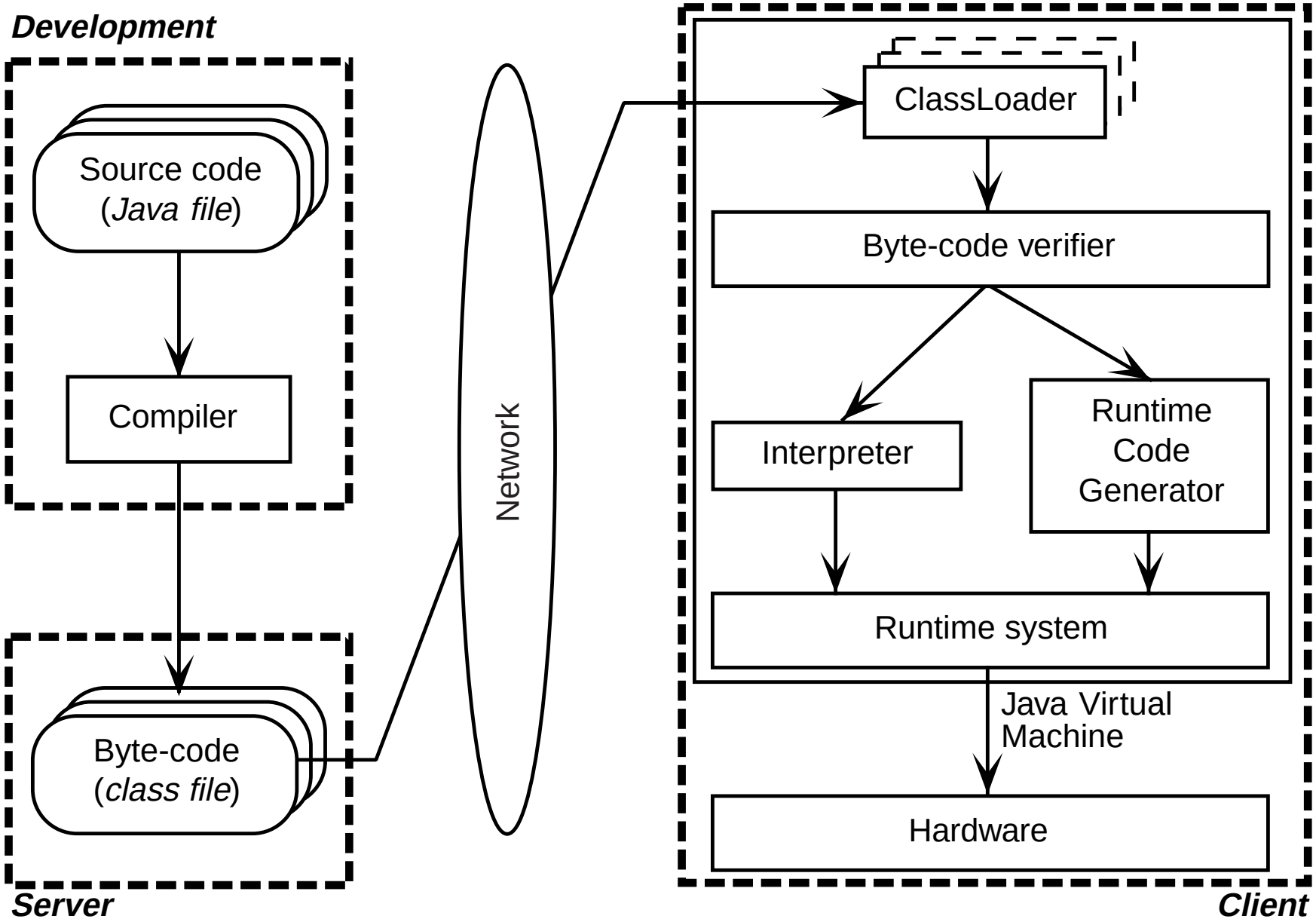
The combination of html and Java made the web a much nicer place to wait in.

Jim Waldo, Senior Staff Engineer, JavaSoft

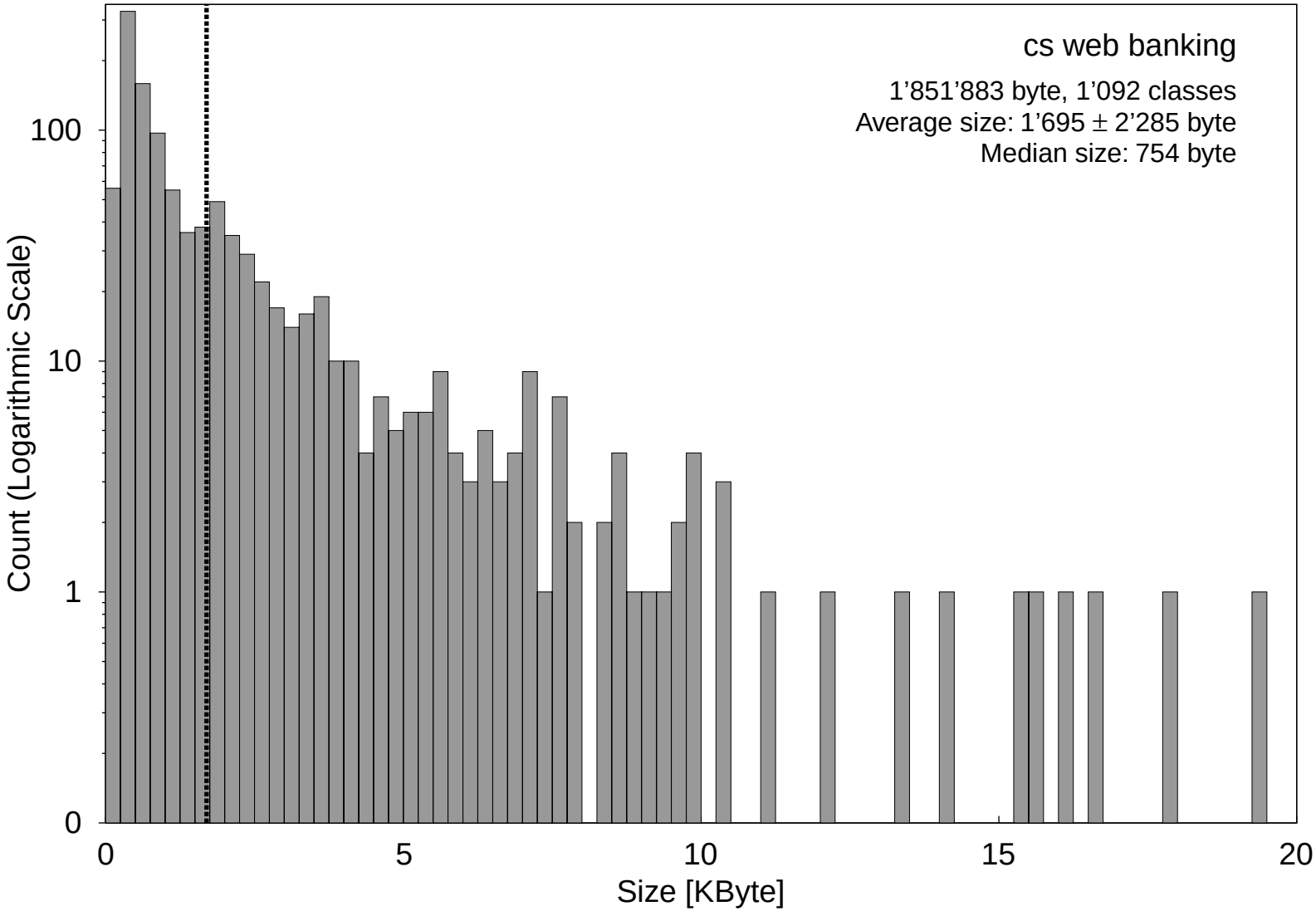
Reasons for the delay:

- **Execution**
  - **Processor speed**
  - **Quality of the encoding of the class files**
- **Transmission**
  - **Network speed**
  - **Size of the class files**
  - **Number of class files**

THE JAVA EXECUTION PLATFORM



JAVA CLASS FILES TODAY



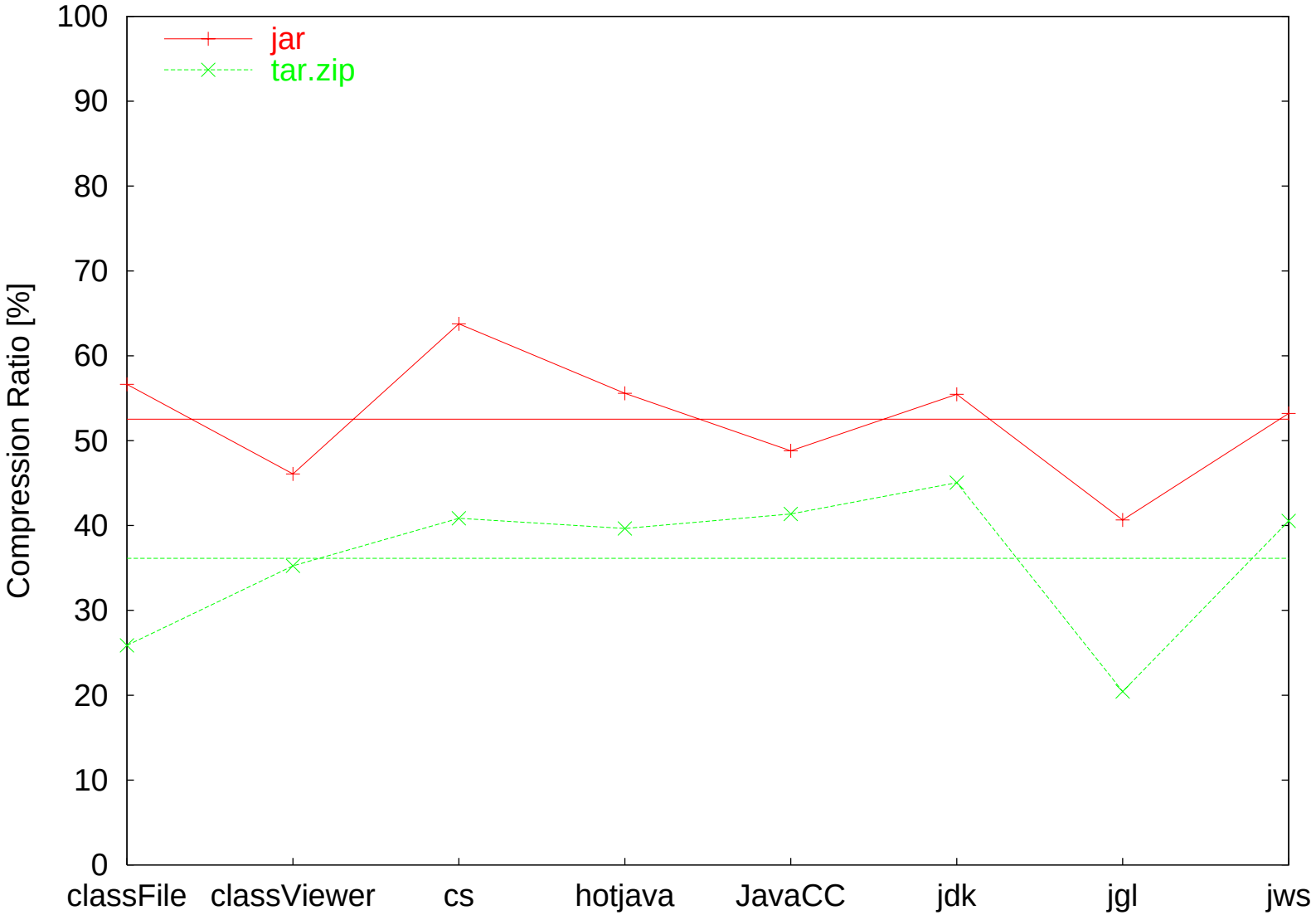
ifi

JAR: JAVA ARCHIVE

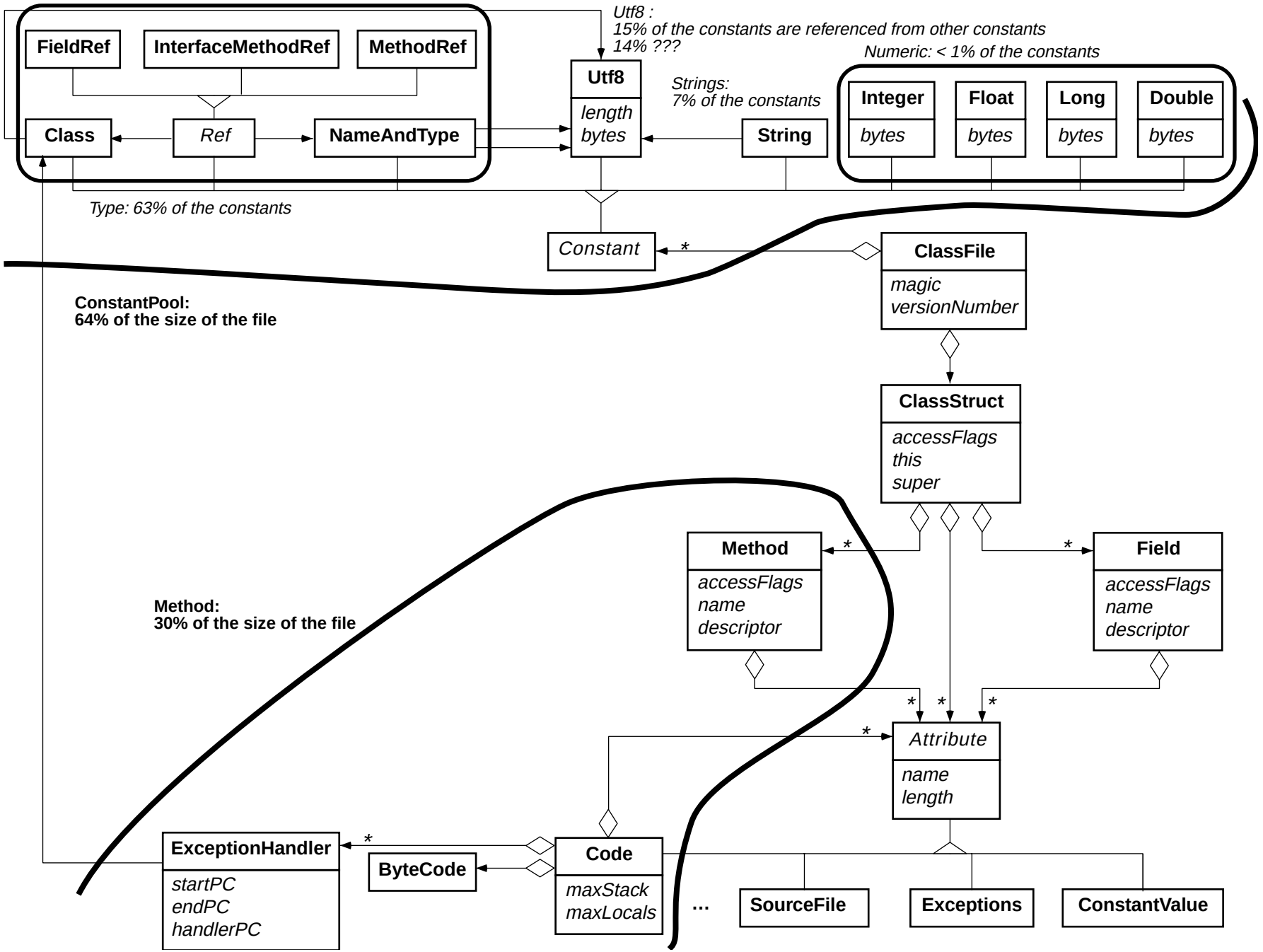
- **stores complete Java application**
  - **classes**
  - **resources (images,...)**
  - **meta-informations**
- **provides random access to the archive's members**
- **is a zip file:**

```
zipFile = directory members.
directory = { Name Size Attributes ... }.
members = { compressed Data }.
```

COMPRESSION WITH JAR

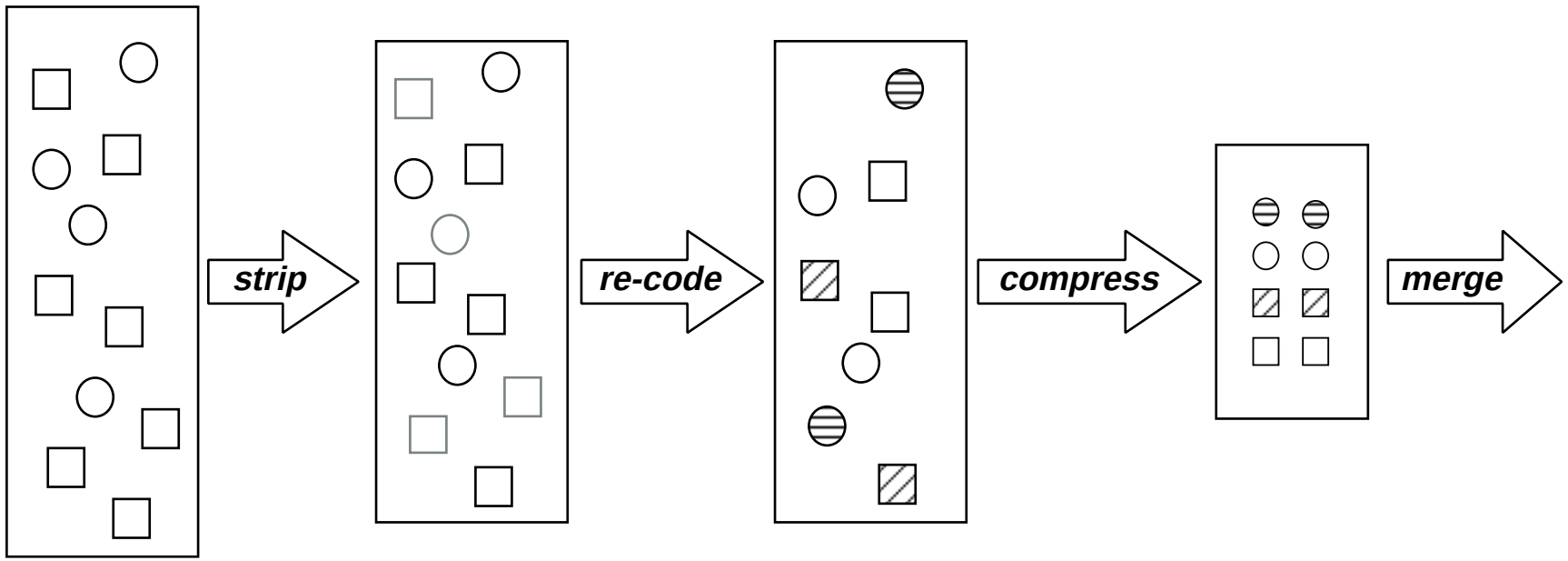


FORMAT OF THE CLASS FILE



4-TIER STRATEGY TO COMPRESSION

- **Stripping**
- **Re-coding**
- **Compressing**
- **Merging**



ifi

STRIPPING

In a given context, some parts of the class file are superfluous:

- **Symbolic information for debuggers**
- **Comments for other tools**
- **Unused methods, fields, classes?**

Possible contexts:

- **Development**
- **Final deployment**
- **Deployment on embedded system**
- **...**

ifi

RE-CODING

The format of the class file reproduces the data layout in the JVM

- **simple to decode**
- **optimized for faster execution**

In a file, alternate representations are possible, which are

- **smaller**
- **amenable to further compressions**

ifi

COMPRESSING

Text, byte-code,... can not be (easily) re-coded.

java.util.zip is

- **a good general purpose, statistical compressor**
- **byte oriented**
- **part of the standard distribution**

java.util.zip best compresses large, homogeneous data sets

1. **group and prepare the data in a single block**
2. **compress the resulting block**

ifi

MERGING CLASSES

Java binaries are groups of classes that belong together; they

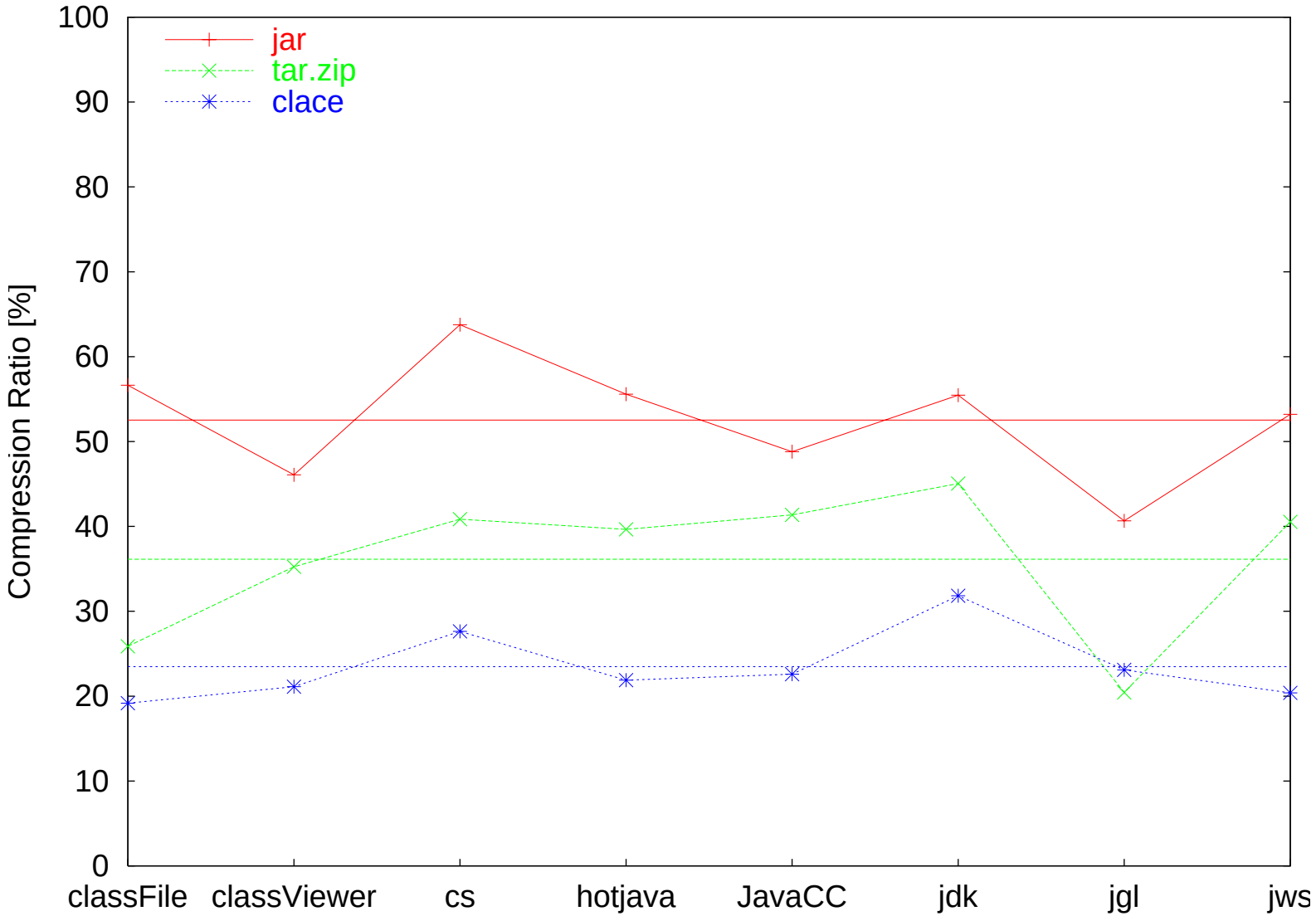
- **have privileged access to one another**
- **use the same external libraries**

The classes have common references that can be shared! Agglomerating all the classes reduces the overall size, but still results in a large archive.

How to select the classes to merge?

- ***all* the classes in a** `package`**?**
- ***cluster* according to some *similarity measure*?**

COMPRESSION WITH CLACE



ifi

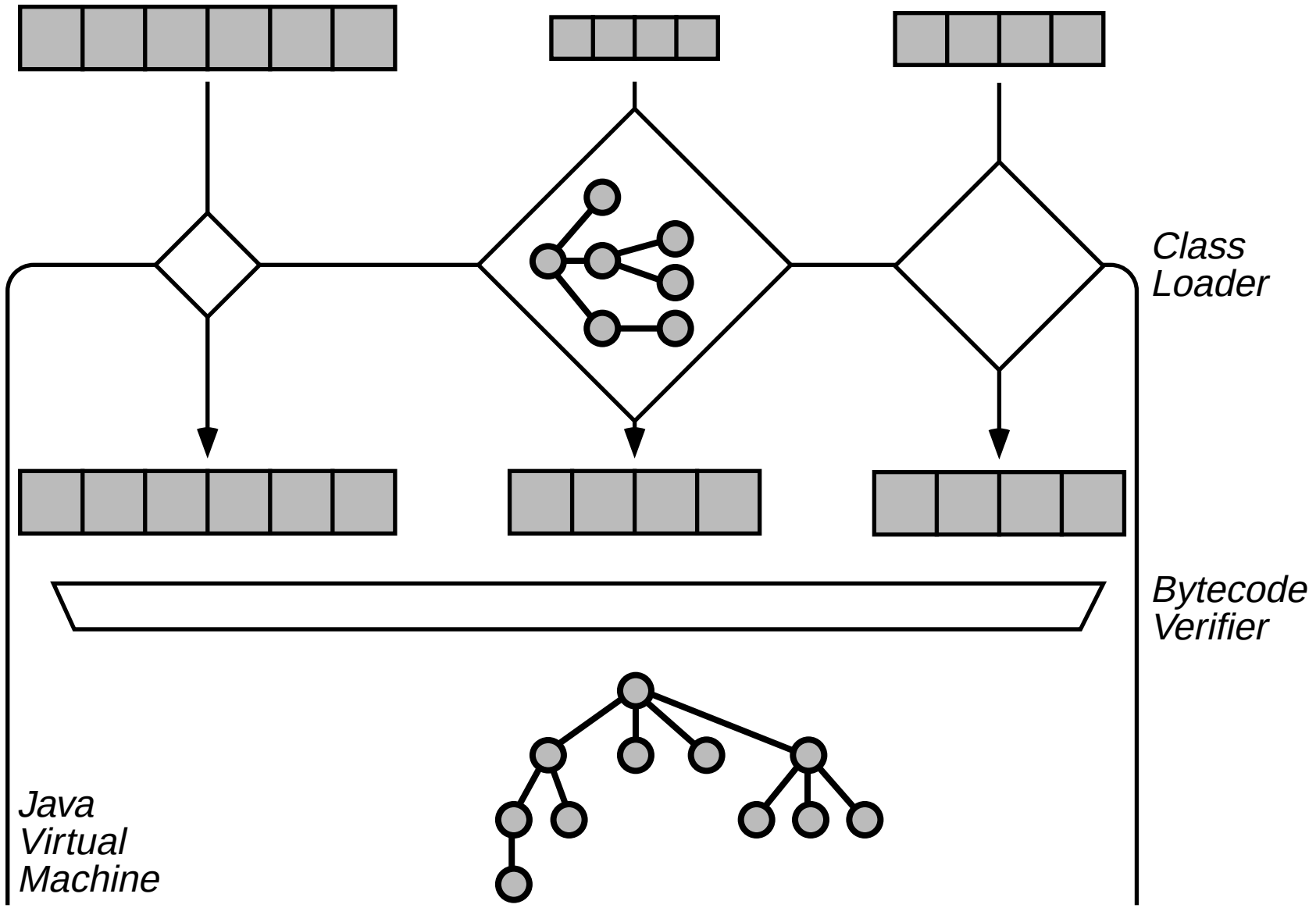
CAR: CLASS ARCHIVE

This new format is as small as possible *and* as simple as possible.

- **applies minimal transformations:**
  - **strip the class file**
  - **re-order the constant pool**
- **stores essential meta-information only:**
  - **name of the class**
  - **size of the class file**
- **compresses the resulting archive as a block.**

```
carFile = compressed { member }.
member = Name Size Data.
```

CLASS LOADERS



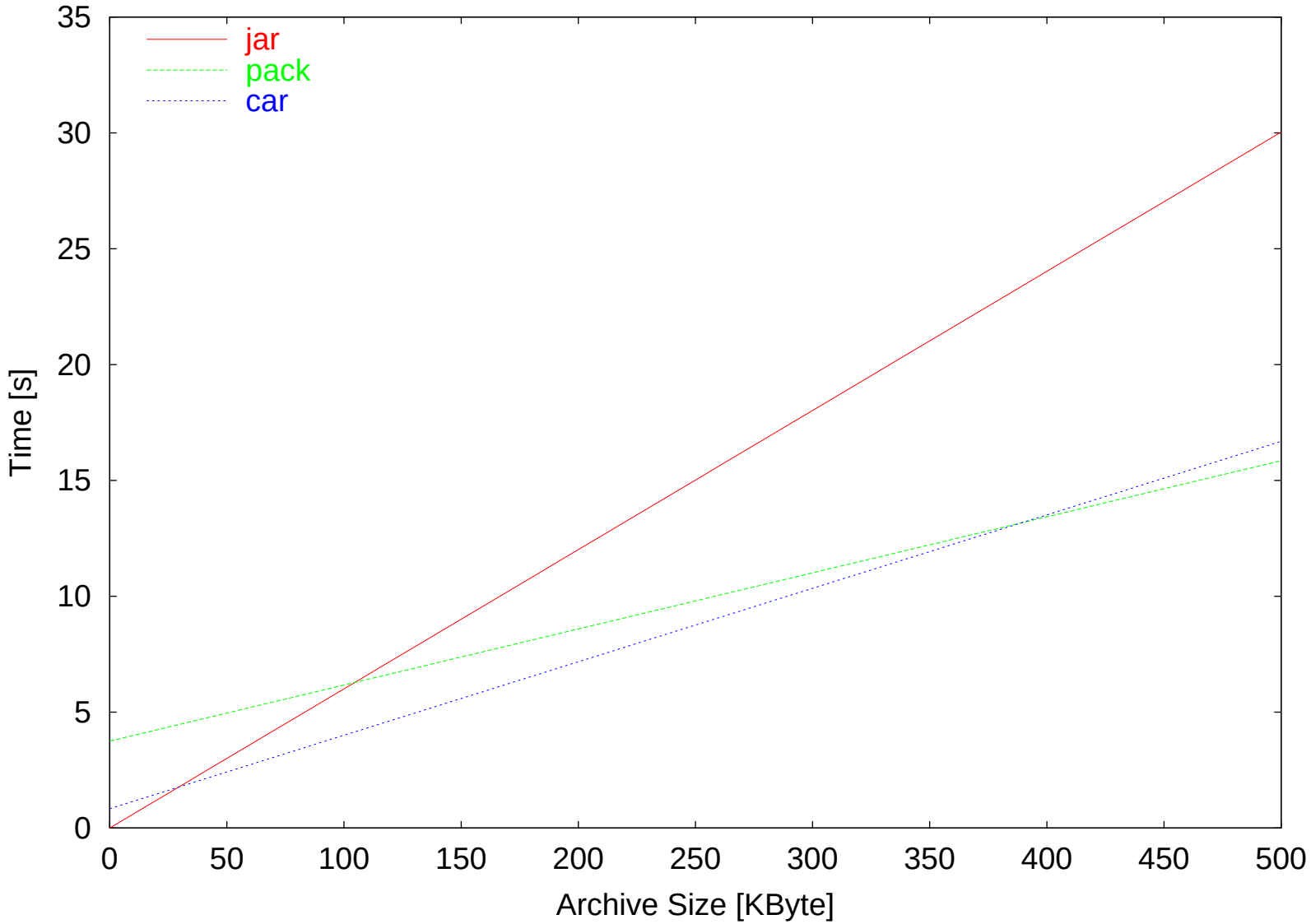
A Comparison Model

Relevant is the *total time* t, which is the sum of the *transmission time* t_{trm} and of the *decoding time* t_{dec}.

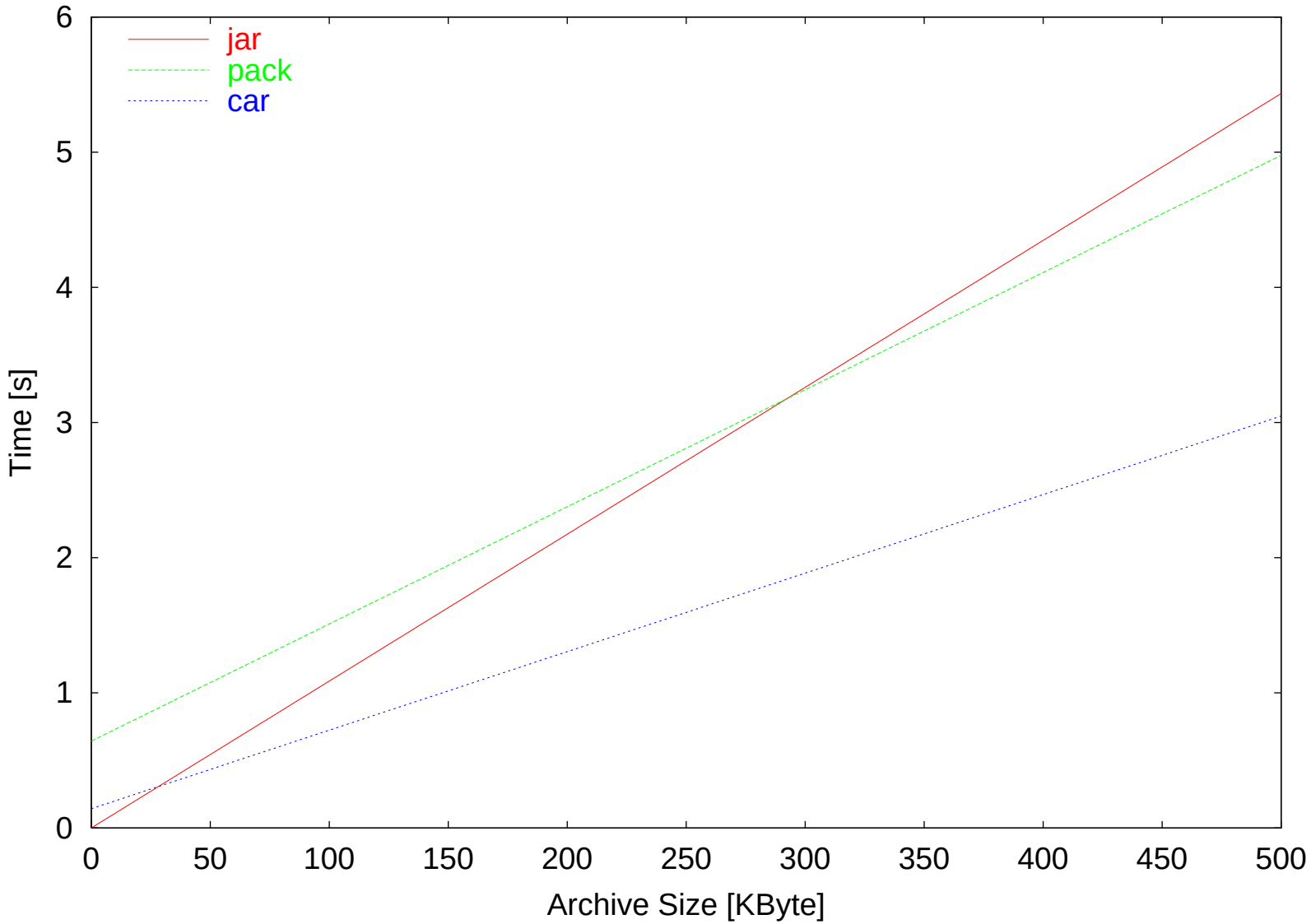
$$t = \frac{\text{sizeof}(dec) + k\,\text{sizeof}(members)}{v_{trm}} + \frac{k\,\text{sizeof}(members)}{v_{dec}}$$

		jar	pack	car
Median compression ratio [% of raw]	k	57	18	30
Size of decompressor [KByte]	sizeof(*dec*)	0	36	8
Median decompression speed [KByte/s]	v_{dec}	826	33	663

At 9'600 Bit/s



At 56'000 Bit/s



ifi

CONCLUSIONS

- class
  - many files
  - formatted for smooth interpretation
  - contain superfluous material
- jar
  - one smaller file: 57% of original size
- pack
  - very compact: 18% of original size
  - requires a customized JVM for speed
- car
  - compact: 30% of original size
  - fast with common JVM

ifi

REFERENCES

- **Denis N. Antonioli and Markus Pilz, *Analysis of the Java Class File Format*, Technical report ifi-98.04, Departement of Information Technology, University of Zürich, Switzerland, April 1998, <ftp://ftp.ifi.unizh.ch/pub/techreports/TR-98/ifi-98.04.pdf>**
- **Denis N. Antonioli, *Car: The Class Archive Format*, Technical report ifi-2001.01, Departement of Information Technology, University of Zürich, Switzerland, January 2001, <ftp://ftp.ifi.unizh.ch/pub/techreports/TR-2001/ifi-2001.01.pdf>**