Are Your Incoming Aliases Really Necessary? Counting the Cost of Object Ownership

Alex Potanin, Monique Damitio, James Noble
Victoria University of Wellington, New Zealand
{alex|monique|kjx}@ecs.vuw.ac.nz

***Abstract*—Object ownership enforces encapsulation within object-oriented programs by forbidding incoming aliases into objects' representations. Many common data structures, such as collections with iterators, require incoming aliases, so there has been much work on relaxing ownership's encapsulation to permit multiple incoming aliases. This research asks the opposite question: Are your aliases really necessary?**

In this paper, we count the cost of programming with strong object encapsulation. We refactored the JDK 5.0 collection classes so that they did not use incoming aliases, following either the owner-as-dominator or the owner-as-accessor encapsulation discipline. We measured the performance time overhead the refactored collections impose on a set of microbenchmarks and on the DaCapo, SPECjbb and SPECjvm benchmark suites. While the microbenchmarks show that individual operations and iterations can be significantly slower on encapsulated collection (especially for owner-as-dominator), we found less than 3% slowdown for owner-as-accessor across the large scale benchmarks.

As a result, we propose that well-known design patterns such as Iterator commonly used by software engineers around the world need to be adjusted to take ownership into account. As most design patterns are used as a building block in constructing larger pieces of software, a small adjustment to respect ownership will not have any impact on the productivity of programmers but will have a huge impact on the quality of the resulting code with respect to aliasing.

I. Introduction

Encapsulation is a crucial attribute of object-oriented programming and design. Object ownership [1] enforces encapsulation by explicitly identifying the internal representation objects: an object *owns* its representation, and owned objects are protected behind the object's interface. An object should act as a "single entry point" for its representation — considering the heap as a graph, an owner dominates the objects that it owns [2] — in other words, there can be no incoming pointers to an owned object that bypass the object's owner. Figure 1 illustrates this graphically: the nodes making up a linked list are owned by the list — while the element data in the list can exist outside the list. The most restrictive form of object ownership, enforced by the ownership types, is both *strong* and *deep*: strong (or prescriptive) because external references to owned objects cannot be used; and deep, because ownership is transitive — an object that owns the list owns the list's links as well.

Strong, deep, ownership offers a number of benefits for program design, generally because of the single entry point property. For example, owned objects may be deleted as soon as their owner becomes inaccessible, supporting real-time memory management [3]. Security checks carried out on the owning object will always govern access to the owned objects [4]. Class invariants can be affected only by the methods defined in that class and its ancestor classes rather than by method calls on unrelated objects [5].

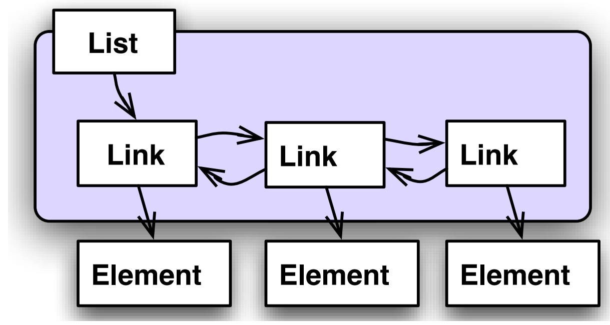

Fig. 1. Encapsulated Linked List

Ownership is emerging as an important technique for designing parallel and concurrent object-oriented systems to take advantage of multicore processors. Many OO actor systems, including Kilim [6], Thorn [7], and Scala Actors [8], impose an ownership discipline to ensure that actors communicate only via message passing and to prevent one actor accessing another actor's internal representation. The high-throughput pipes-and-filters processing system StreamFlex uses ownership to ensure isolation between filters executed in parallel [9]. Parallel languages such as DPJ [10] or AJ [11] use ownership techniques to describe the data they access and detect interfering computations.

The problem this paper addresses is that adopting an ownership or encapsulation discipline may impose runtime costs on programs. As with many other kinds of types, ownership types can be checked statically or dynamically. Static checks, carried out at compile time, do not impose any *direct* costs on program execution, while dynamic checks will impose some overhead directly. Ownership disciplines can also impose *indirect* costs, by precluding designs that bypass objects' interfaces and refer directly to objects' supposedly hidden internal representations. Figure 2 shows how an iterator can rely on an incoming alias into the nodes of a list — such an iterator can move directly and efficiently from one link to the next, and can do so without any reference to the list object within which the links should

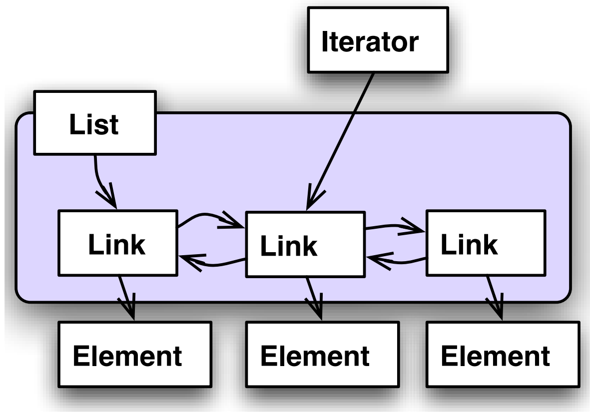

Fig. 2. Linked List with Iterator

be encapsulated. Designs that respect encapsulation, and rely on objects' interfaces, may be less efficient than designs that breach encapsulation with incoming aliases.

In this paper we address the question "*Are your incoming aliases really necessary?*". We study two varieties of strong, deep ownership — owners-as-dominators and owners-as-accessors — and for owners-as-accessors we consider both static and dynamic checking. We then attempt to determine the cost of following each discipline on programs' design and performance. The contributions of this paper are answers to three questions about programming with ownership:

1) How must designs change to respect encapsulation?
2) What performance cost do these changes impose?
3) How does this impact programs' performance?

The remainder of this paper is structured as follows. First, Section II briefly reviews object ownership, and describes the ownership disciplines we investigated for this study. Section III presents a case study of the design of the core collection classes and the refactorings required to adapt them to the object ownership discipline. Section IV describes our benchmarking methodology, and Section V presents the results of both the micro- and macro- benchmarks. Finally Section VI discusses the implications of the results in the context of related work, and Section VII concludes the paper.

II. Ownership

The idea of ownership is to partition the objects accessible from any point in the program according to the object to which they belong [5], [12], [13]. We consider that objects relate to each other in one of three modes: *owned*, *peer*, and *external* objects. An ownership discipline requires programs to keep these modes separate: type casts, assignments, or subsumption must not allow an object in one mode to be accessed as if it were in another mode.

Owned objects are fully private and encapsulated within their owning objects. In Figure 3, object A owns objects B and C, and object C owns D. This is deep ownership because C and D are both owned transitively by A.

Peer objects are siblings with respect to their owning objects, that is, all peers have the same owner. In Figure 3, B and C are peers, and can refer to each other without incoming aliases that would breach encapsulation.

External objects are — as the name implies — outside the current object, but are still accessible without any incoming aliases. In Figure 3, E is external to all the other objects, and B is external to D. External objects are typically passed in and out as parameters to objects, such as the elements of a collection.

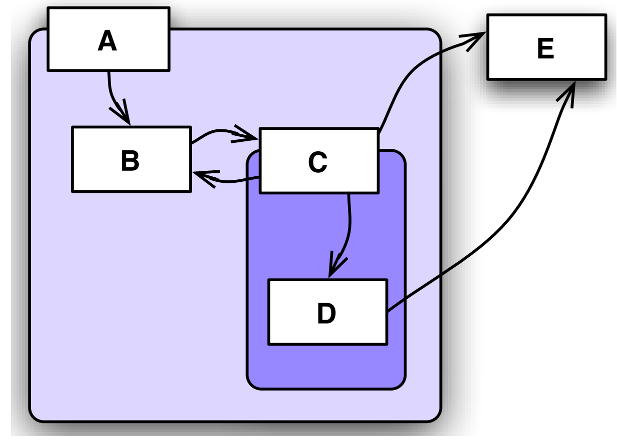

Fig. 3. Relationships between objects with ownership

In this study we investigate two realisations of the basic ownership scheme: owner-as-dominator, and owner-as-accessor.

Owner-as-dominator interprets the "no incoming aliases" constraint in the same way as Clarke et al.'s original ownership types proposal [1]: neither static nor dynamic references are permitted to cross an encapsulation boundary. In Figure 4, only references shown by the solid, black lines are permitted: from external objects to their peers; from an owner to the objects it owns; and between internal peers with the same owner. Aliases that cross encapsulation boundaries (e.g. the dashed red line in Figure 4) are forbidden. Considering the heap as a rooted graph, an owner is a graph-theoretic dominator of all the objects it owns.

Owner-as-accessor takes a different interpretation of the "no incoming aliases" rule, based on Müller et al.'s Universe Types [14], and an experimental dynamically-checked ownership scheme [15]. Owner-as-accessor permits incoming heap references — such as the dashed red line in Figure 4 — provided they are not used directly: and method requests that cross an encapsulation boundary must do so via the boundary's owner object. The external object in Figure 4 cannot use the dashed red incoming link directly, but it could call a method on the Owner (dotted green line "1"). Being inside the encapsulation boundary, just like with owners-as-dominator, that method can then modify the internal object (dotted green line "2" in Figure 4. What makes owner-as-accessor different, is that the Owner object can both accept and return a direct reference to the Internal object (dashed red line). However, that reference can only be stored and passed around but cannot be used to modify the Internal object unless the referrer is inside the appropriate ownership boundary.

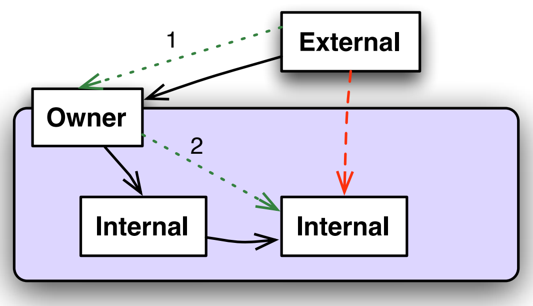

Fig. 4. References and Method calls

For example, Figure 5 shows a simple example following the one in Figure 4. The Internal object can have references to it stored outside its Owner as shown in line 8, which would be illegal under owners-as-dominator. However, the modification is only allowed via an appropriate owner (line 9) and not via external references (line 10).

```
1  class Internal { String s = "abc"; }
2  class Owner { Internal i = new Internal();
3    void modifyI() { i.s = "def"; }
4  }
5  class External {
6    Internal i; Owner o = new Owner();
7    void doIt() {
8      this.i = o.i; // OK. Stores external ref.
9      o.modifyI(); // OK. Modifies Internal.
10     // this.i.s = "ghi"; // WRONG. Not owner!
11 } }
```

Fig. 5. Owner-as-Accessor Code Example

Owner-as-accessor does not constrain the heap topology, but it does constrain the control-flow graph: a method invocation upon an owner must dominate all method invocations upon all objects owned by that owner. This control-flow constraint is implicit in owner-as-dominator encapsulation: any design conforming to owner-as-dominator also satisfies owner-as-accessor.

In this paper we are concerned with design changes, rather than the annotations required to express the properties of a design within an ownership type system, or the properties of the type system itself. Using the JavaCop program constraint system [16], we have implemented an ownership checker based on the theory of Tribal Ownership [17]. Each type used within a class is assigned to a single ownership mode. All instances of inner classes are owned by their enclosing objects, and additional classes can also be annotated as owned. Method invocations (for both owner-as-dominator and owner-as-accessor) and assignments (for owner-as-dominator) are checked to ensure they maintain the ownership invariants.

III. Collections and Ownership Case Study

The Java Collections Framework [18], [19] has formed an important part of the Java platform ever since its first release in JDK 1.2. Designed by Joshua Bloch, there are over 50 classes and interfaces in the framework as a whole. The core of the framework, however, are a relatively small number of interfaces to collection objects (Set, List, and Map), and a similarly small set of implementations of those interfaces [20]. Our case study is based on the Java 5.0 version of Collections, as this is the version required by the DaCapo benchmarks, see section IV-B.

The eight classes: ArrayList, LinkedList, HashMap, LinkedHashMap, TreeMap, HashSet, LinkedHashSet, TreeSet, plus the legacy classes Hashtable and Vector, form the backbone of the mainstream collections usage: our case study analyses these ten implementation classes. We consider the implementations of the two main abstractions — lists and maps — in turn, explaining their design; describing how (and how much) they encapsulate their representations; and if they do not, outlining refactorings to restore encapsulation. Our discussion focuses on lists and maps because HashSet and TreeSet are wrappers that implement sets using HashMap and TreeMap respectively. All our implementations are publicly available[1].

A. *Lists*

Although the collection objects' interfaces are quite rich — ArrayList, for example, defines around thirty methods — for the purpose of this paper we need to consider only a few. Lists (and Maps) define get(index) and set(index,element) methods to read and write collection elements. List (and Sets) also define add(element) and remove(element) methods to add and remove elements. All collections support an iterator () method that returns a dependent Iterator object. An iterator supports at least next() and hasNext() methods that traverse through the collection an element at a time. We say the iterators are "dependent" on their underlying collection because they access (and, in some versions, can update) the actual elements stored in the collection.

1) ***ArrayList:*** An ArrayList is one of the simplest of the collections. An ArrayList<E> stores elements of type E in a primitive array, and copies and replaces that array as necessary as the collection grows and shrinks.

Figure 6 shows the internal structure of an ArrayList and its dependent Iterator. Elements in an ArrayList are accessed by simply accessing the corresponding elements of the array, after checking they are within the range of valid elements:

Crucial to the correct operation of an ArrayList is that each underlying array is owned by the ArrayList whose elements it holds, and so must never be accessible from the outside. This is because if the list grows (or shrinks) the array will

[1] http://homepages.ecs.vuw.ac.nz/~alex/software/files/ownershipcollections20120817.tgz

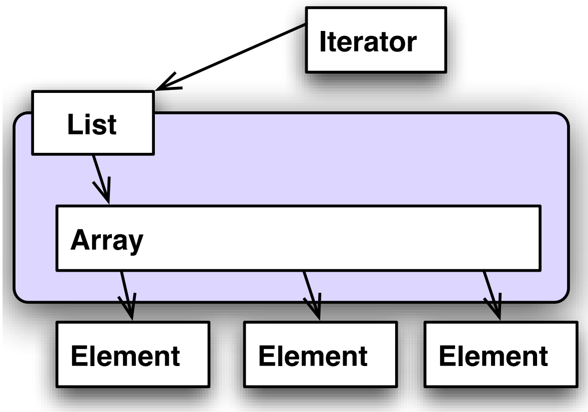

Fig. 6. Encapsulated ArrayList and its Iterator

be replaced with a larger (or smaller) array, and the ArrayList elements copied from the old array into its new replacement.

This encapsulation is respected even by ArrayList's iterators. Implemented as an inner class[2], the iterator refers to its ArrayList via Java's implicit link between every inner class instance and their enclosing "outer" class instance. An ArrayList iterator maintains an integer cursor field that indexes the next element to be returned. The iterator's next() method simply returns the element at the cursor position, and then increments the cursor.

ArrayLists preserve encapsulation because their methods and iterators access their list only via that lists's public methods size() and get().

Even though it is implemented as an inner class, conceptually an ArrayList iterator is outside the ArrayList's encapsulation boundary — the iterator has no privileged access to its underlying ArrayList instance, and in particular does not access any private representation owned by the ArrayList (e.g. the underlying primitive array). An ArrayList object *always* acts as the single entry point for the list abstraction that it represents.

Because ArrayList encapsulates its representation, the implementation does not need to change to satisfy owner-as-dominator encapsulation, and so also satisfies owner-as-accessor encapsulation. In this case at least, ownership does not impose any additional performance cost on the design.

*2) **Vector**:* The design of the legacy Vector class is broadly similar to ArrayList. The only issue is that the vector iterator (an instance of the legacy Enumeration interface) directly accesses the underlying array owned by the vector. We refactored this quite straightforwardly to use the public interface, as in ArrayList.

*3) **LinkedList**:* LinkedList, the other standard List implementation, is a more difficult case than ArrayList or Vector. As we discussed in the introduction, the LinkedList class maintains a doubly-linked list of link nodes (instances of the static inner class Entry<E>). The LinkedList's iterator maintains an incoming pointer (called next) that refers directly to the current link Entry. The iterator's next() method runs this pointer along the list's internal structure, returning the element at the current position. This structure sharing means that a LinkedList object is not a single entry point to its representation — every outstanding iterator on the LinkedList accesses the list entry nodes directly.

*4) **Naïve Owner-as-Dominator LinkedList**:* Our first (naïve) refactoring was simply to adopt the ArrayList iterator described above — the ArrayList iterator requires only a List interface, and LinkedList, like ArrayList, implements List. Unfortunately, we expected that this design would not perform very well and our initial tests confirmed our expectations. The problem is that the iterator calls get(index), and the get(index) method on a linked list must start from the beginning (or end) of the list and count along until it locates the list Entry holding the indexed element: An individual call to get() on a LinkedList will be $O(N)$ and a whole iteration will be $O(N^2)$. So: while adopting an ownership discipline can simplify the design of the LinkedList class — suggesting the use of a more abstract iterator — this naïve design would impose a substantial performance cost [3].

*5) **Single Place Cache**:* The JDK specification makes clear that programmers should expect the performance of LinkedLists will always be $O(N)$ for individual random accesses. The specification also makes clear that programmers should expect $O(N)$ for a full traversal via an iterator. The original iterators with direct pointers into the list delivered this performance. A slightly more complex design, however, can restore $O(N)$ traversals in most cases, while preserving owner-as-dominator encapsulation — the LinkedList object owns its list entries, and remains the single entry point of access to those list entries. We add a cache to the linked list that remembers the last accessed entry and its index. Then, a call to get(index) can look into the cache, and update the cache once it has found the requested element.

In this design, a single (forward or reverse) traversal via an iterator, or even a traversal driven programmatically sometimes forwards and sometimes backwards, should have $O(1)$ performance for a single call to next() and consequently $O(N)$ for a full traversal of the list. The overhead of the cache itself should not be substantial. Random access to a LinkedList will still have $O(N)$ performance for each get(), but this is permitted by the framework specification, since LinkedList is not a random access structure. The specification also invalidates all except one iterator whenever the list is modified. This forbids modifications by more than one iterator, and so we expect a single place cache should suffice for most LinkedList use cases. Multiple simultaneous traversals that purely read data,

[2]Technically, an inner class of the AbstractList superclass of ArrayList, although that does not affect the encapsulation of the ArrayList.

[3]While it may seem unreasonable to even attempt to replace an $O(1)$ operation with an $O(N)$ one, we have examined a number of iterators per each instance of a linked list in all of DaCapo benchmarks and it was very low — 4 or 5 per list in the avrora and pmd cases and less than one per list for most of the others [21].

however, would still revert back to $O(N^2)$ performance, but we expect such uses of LinkedLists to be rare, as ArrayLists are quicker to traverse than LinkedLists, and have around a quarter of the storage overhead.

*6) **Owner-as-accessor via external proxy iterator**:* As Figure 2 showed, the problem with the standard Java Iterator is that it is *outside* the List but must directly manipulate Links. So long as the standard Iterator is only ever used *inside* the List, there are no encpasulaton breaches. Figure 7 shows an alternative to this design that maintains owner-as-accessor encapsulation, but not owner-as-dominator.

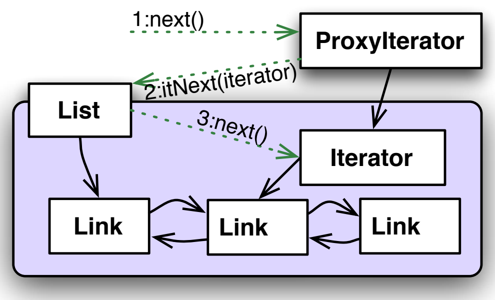

Fig. 7. List Iteration by Proxy

A **proxy iterator** is a peer of the List, and is thus accesible outside. ProxyIterator maintains a reference to a standard Iterator inside the List, but does not use that reference directly. In response to a next() method invocation on the external proxy iterator, the proxy invokes itNext(iterator) on the List, passing the actual internal iterator as a parameter. At this point, control flow passes into the List object (thus maintaining owner-as-accessor) which then invokes next() on the encapsulated Iterator. The cost of this refactoring is creating the external Proxy, plus redirecting its calls via the owning List. An important constraint of this design is that we must be careful to pass the proxy iterator only into the list to which it belongs. This can be checked statically by advanced ownership type systems [22], or it can be checked dynamically [15].

*7) **Owner-as-accessor via indirection iterator**:* The proxy iterator refactoring avoids changing the standard List Iterator, but requires creating the external proxy. Figure 8 shows our final List refactoring, using a single "indirection iterator" object that plays both the roles of the external proxy and internal iterator.

The **indirection iterator** presents the standard Iterator interface (e.g. next()) and forwards those messages to its List, just like the external Proxy. Whereas the external Proxy passed the internal Iterator as an argument, here the indirection iterator passes *itself* as an argument. The implementation of itNext(itr) in List retrieves the current link (getLink) from the iterator and directly gets the next link (getNext). The external iterator keeps references directly into the List's internal Links (rather than to a Proxy) but never uses those Links

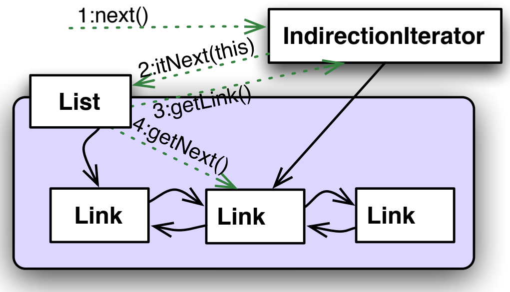

Fig. 8. List Indirection Iterator

directly, thus maintaining owner-as-accessor encapsulation but not owner-as-dominator. We expect this refactoring to be more efficient than a proxy iterator, but it is more expensive to perform, as we cannot reuse the standard iterator object, but rather must incorporate its code into the itrNext()-style methods on the List object. And, as with the proxy design, we must ensure the indirection iterator is only passed into the list to which it belongs, checking either statically or dynamically.

*8) **Other Possible Iterator Refactorings**:* There are more alternatives to consider for ownership-aware list and iterator refactorings. The last author discussed the possibilities in detail in a separate extended paper [23]. For this presentation, we chose the ones we thought would be most representative and easy for software engineers to adopt in practice. More efficient versions include a "magic cookie iterator" [23] that stores an approriate cached position for each existing iterator of the current list communicating by a simple unique id or "magic cookie".

B. The Map Interface

The Map<K,V> interface is the other major interface within the collections framework. Maps provide put(K,V) and V get(K) methods to store and retrieve values V associated with keys K. Maps are not directly iterable — rather they provide methods that return three separate dependent iterable views: Set<K> keySet(); Collection<V> values() and Set<Map.Entry<K,V>> entrySet(), all of which have their own iterators that support modification. The key set (and value collection) contain a set of all the keys (and a bag of all the values) in the underlying map: the entry set is a set of objects that each represent a single key-to-value association and implement the Map.Entry<K,V> interface.

The key encapsulation issue with the core Map implementations is that the Entry objects avaiable via the entry set are the very same entry objects that implement the map. Furthermore, the key and value sets are implemented in terms of the entry set and the entries it contains. This is a tightly coupled design: Figure 9 attempts to show these interrelationships.

*1) **Refactoring HashMap**:* For this case study, we first aimed to find a design for maps with a single point of entry

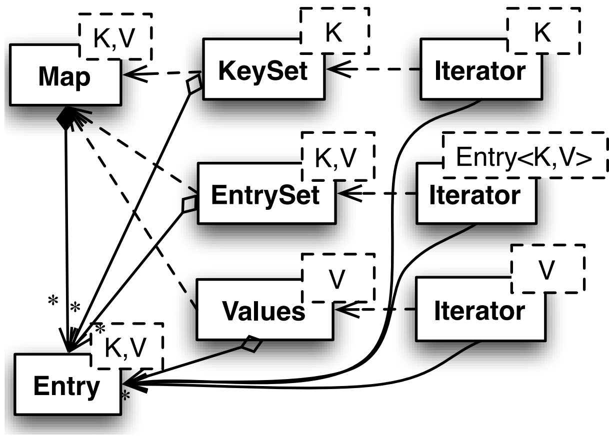

Fig. 9. Map Interfaces. Dashed lines show conceptual dependencies, while solid lines show references in most implementations.

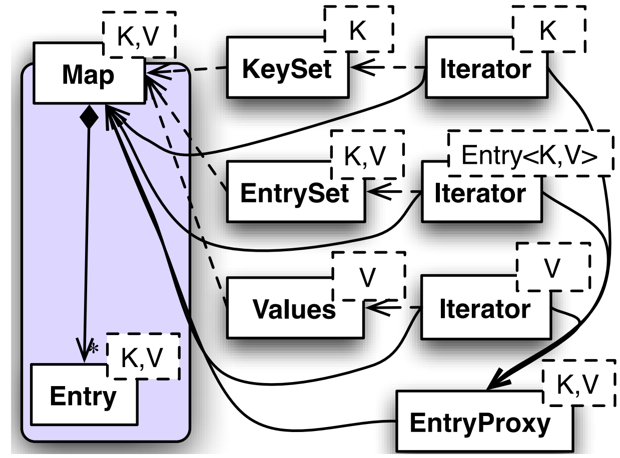

Fig. 10. Refactored Map. The Map Entry instances are owned by the Map object. Dependent views and iterators access the Map only via the Map object's interface, and substitute EntryProxies for Entries.

that could maintain owner-as-dominator encapsulation over as much of the map as possible. In partiuclar, the entry objects storing the key-value mappings need to be protected from outside access. This is straightforward. The problem then is that our refactored maps need to preserve the existing collections' Map interface, so we could exchange implementations to benchmark each design. This means that the dependent set views of a map (the key set, values collection, and entry set); their iterators; and the entry objects must remain available to Map clients.

To support these use cases, we again introduce proxies for these objects as peers of the maps as shown in Figure 10. Writes to the proxies update the underlying map by calling put(K,V) through the main map object's interface, rather than by being part of the map's representation themselves. Within the body of a map implementation (say HashMap) there are three Entry types: Map.Entry is the common public interface used by clients, HashMap.Entry is the owned inner class; and EntryProxy is the (common) peer class. The ownership discipline ensures the peer entry proxies cannot be confused with the owned entry objects and vice versa.

As a result, the Entry<K,V> objects are not accessible outside of the Map<K,V> object that owns them. All three of the map views (key and entry sets and values collection) and their respective iterators *do not have direct references to map's entries*. Instead they work with EntryProxy<K,V> objects that mirror the Entry<K,V> objects encapsulated inside the map and store the same key-value pair as the mirrored entry without allowing direct access to the internal structure of the map. Any modification to such entry proxies will not have any effect on the map's internal representation and any code that relies on modifying the map by manipulating is entries directly as opposed to using map's public interface will no longer work. An important result of our study is that in none of the benchmarks under consideration have we found any instances of such manipulation and thus the only reason for exposing the internal entries behind each map is presumed efficiency improvement.

The dependent view iterators are the most challenging part of this refactoring. All three iterators are subclasses of the abstract inner class HashIterator<E> which traverses the HashMap by accessing the underlying table and Entry objects directly. To restore owner-as-dominator encapsulation, we refactored the design to introduce a completely external iterator, similar in style to the ArrayList iterator, that does not use incoming pointers into the HashMap's implementation. This iterator keeps track of the current and next keys, and uses new K getFirstKey(), K getNextKey(K key), and hasNextKey(K key) methods on the HashMap to manage the iteration. The implementation specific code from the original iterators' nextEntry() method is refactored to call the map's getNextKey() method.

*2) **LinkedHashMap**:* The LinkedHashMap class extends the HashMap class by threading a doubly-linked list between the table Entry objects to support quick and stable traversal. Once we had refactored HashMap, the only change LinkedHashMap required was to ensure the getNextKey() method traversed the list.

*3) **Hashtable**:* The design of the legacy Hashtable class is broadly similar to HashMap, with the same underlying design from Figure 9 and the same epidemic of inner classes and aliases, although the method names and interfaces are not compatible. Like the other legacy class, Vector, all accesses to a Hashtable are synchronized. We were able to encapsulate Hashtable using very similar refactorings to HashMap. Our refactored implementation also reuses the common Map key-based traversal methods and the common EntryProxy class. We had to refactor the existing Hashtable iterators, rather than reuse the shared implementation, because Hashtables must support both the Iterator and the legacy Enumeration interfaces.

*4) **TreeMap**:* We originally planned to refactor the TreeMap class separately, once we had completed the refactorings of the various hash-based maps.

Considering the refactored map design (Figure 10) we realised that we would in fact be able to re-use all the iterator, view, and entry objects from HashMap, because they can only communicate with their underlying map via that map's public interface: encapsulating the representation has also abstracted the representation behind that interface (albeit extended with the various getFirstKey() / getNextKey(K key) methods). In the same way as the residual public interface of a linked list is the List interface, the residual public interface of the various map classes is just this extended Map interface, and so one single, reusable, common external iterator class suffices to iterate over any kind of map.

Again, we expect that the narrower interface between an iterator and its underlying collection will decrease the efficiency of the iteration. Calling getNextKey(currentKey) will require tracing down from the root of the Red-Black tree to the node holding that key, and this is certainly more work than just following a pointer directly. We also tried single place caches (as with LinkedLists, section III-A5), but disabled them for our tests as the DaCapo xalan benchmark did not tolerate the resulting behaviour.

*5) **Owner-as-accessor maps**:* Finally, we refactored all the Maps to maintain owner-as-accessor encapsulation rather than owner-as-dominator. We applied the two refactorings we used with LinkedList to each Map, in spite of the differences between Map implementations. We built external proxies for the various iterators, sets, and entries that simply stored references to the standard iterators etc., which were treated as internal to the Maps, just as in the LinkedList proxy iterator (see Section III-A6 and Figure 7). These proxies were able to be reused across all the various Map implementations. Then we built indirection iterators, sets, and entries following the design of the corresponding indirection objects for the LinkedList (Section!III-A7).

IV. Experimental Methodology

To evaluate our refactorings, we performed a number of benchmark studies comparing the original and refactored collections: we include three small microbenchmarks and the three major bechmark suites: DaCapo, SPECjbb2005 and SPECjvm2008. The refactored collections include: **OasD** that implements owners-as-dominators discipline, **Proxy** that implements owners-as-accessors using a proxy iterator, **Proxy Dynamic** that additionally performs a run-time ownership check, **Indirection** that implements owners-as-accessors using an Indirection Iterator, and **Indirection Dynamic** that additionally performs a run-time ownership check. In this section we describe our methodology, and present the results in the next section.

A. *Microbenchmarks*

We adopted three microbenchmarks to test collections' performance, focusing on the cost of iterating over a whole collection.

1) The IteratorLoops test from the Doug Lea's JSR166 collections microbenchmarks [24]. This runs a large number of traversals over partially filled collections with occasional additions of elements. The result is the time taken for a single `next()` step of an iterator.
2) LinkedList iteration: (a) forwards; (b) backwards; (c) forwards, but with two iterators interleaved, the second iterator indices after the first iterator. We designed the last test to disrupt the caching algorithm in the refactored linked list.
3) The MapMicroBenchmark test from the Doug Lea's JSR166 collections microbenchamrsk [24]. This runs a large number of element-level map operations reflecting a typical usage of maps in the real world and reports an average time an operation takes. This works on on table-based map implementations only (i.e. not TreeMap) and we evaluated different map sizes in increasing order.

To make sure that our numbers were not disturbed by Java's garbage collector or just-in-time compiler, we warmed up the VM before timing the tests. We set the the −XX+PrintGC and −XX+PrintCompilation VM options and carefully checked our traces that compilation and collection did not occur during timed runs. Every microbenchmark was run 25 times and the results analysed.

B. *DaCapo Benchmarks*

The DaCapo benchmark suite [25] is a well-established benchmark suite *representative of typical Java loads*. We included every benchmark in DaCapo in our study: avrora, batik, eclipse, fop, h2, jython, luindex, lusearch, pmd, sunflow, tomcat, tradebeans, tradesoap, and xalan. The DaCapo benchmarks come with data sets of different sizes. We used the *large* size for each benchmark where it was available. The fop and luindex do not include a large size, so for these two benchmarks we used the *default* size.

We carried out 5 runs of 30 iterations of each benchmark, of which the last 5 are used, resulting in 25 data points for each benchmark in each condition [26].

C. *SPECjbb2005 Benchmark*

SPECjbb2005 is a Java Server Benchmark capturing the common types of server side Java applications today. SPECjbb2005 is well-known for making a heavy use of collections and was thus considered essential to be included in our study. Following the other benchmarks in our paper, we ran SPECjbb2005 on a default set of 16 warehouses 25 times and report the averages of these runs for different collections implementations.

D. *SPECjvm2008 Benchmark*

SPECjvm2008 is a Java Virtual Machine Benchmark that measures the performance of a typical Java Runtime Environment using a selection of real life applications focusing on core Java functionality. We included it for completeness as a more traditional macrobenchmark with a caveat that DaCapo was designed to improve on the number of shortcomings of SPEC-style benchmarking [25].

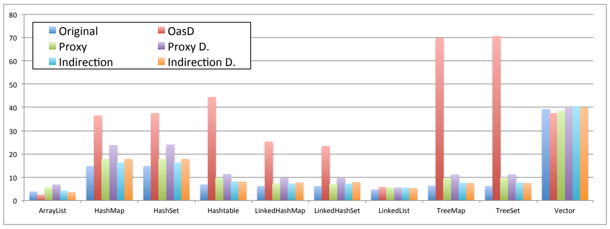

Fig. 11. Microbenchmark 1: IteratorLoops (single iteration time in *nano*seconds)

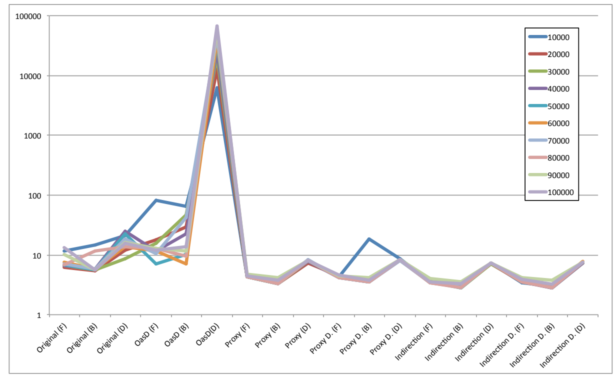

Fig. 12. Microbenchmark 2: Linked List Iteration (single iteration time in *nano*seconds). Note that the time (y-axis) scale is logarithmic.

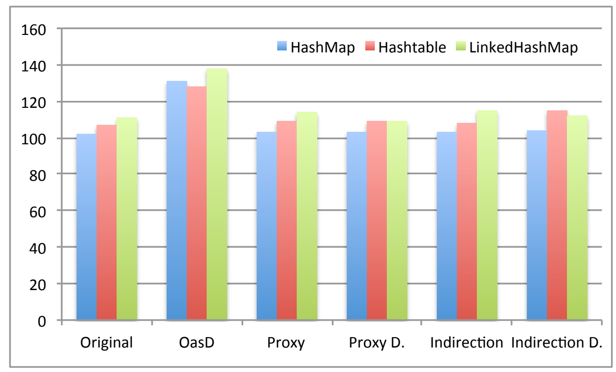

Fig. 13. Microbenchmark 3: MapMicroBenchmark (*nano*seconds per element operation (averaged across get, put etc) across different element types)

E. *Execution Environment*

Our choice of Java Development Kit (JDK) version and thus Java Collections Framework implementation was driven by the latest version of DaCapo Benchmark Suite [25]: v9.12, released in December 2009. DaCapo was built using Java v1.5.0 and this is the version of JDK and Java Collections Framework that we used: in particular the 1.5.0_22 version as the latest Java v1.5.0 version available on the Oracle web site. We executed all the tests on the Java HotSpot(TM) Server VM (build 1.5.0_22-b03, mixed mode). We compiled the modified collection implementations with javac version v1.5.0_22. This choice of JDK version meant that we had to omit one of the SPECjvm2008 sub-benchmark (xml.transform) that requires a later version of Java.

For the purposes of our measurements we placed the modified collections and the same classes that were unmodified in the "ownership" and "original" folders and then utilised the Java –Xbootstrappath/p: option to place our classes in the beginning of the JVM boot class path.

We ran all our tests on a Ubuntu Linux v11.4 machine using v2.6.38-8 SMP kernel with the SMP option selected in the kernel, configured to only use 7 of the 8 cores available in our Dell OptiPlex 790 (Intel Core 2 i7-2600 CPU 3.40 GHz with 4GB of RAM). We used the Linux taskset command to set the CPU affinity of our Java Virtual Machine to the unused core to minimise disturbance from the rest of the operating system.

V. RESULTS

A. *Microbenchmarks*

*1) **IteratorLoops**:* Figure 11 shows the results of the IteratorLoops benchmark from JSR166 benchmarks for all the collections in the *original* and five *ownership refactored* variants. The figure plots the mean time (in nanoseconds) for 1 iteration step (i.e. a call to next).

We can observe that owners-as-dominators refactored implementations were slower per iteration by a factor of three, except for Hashtable and TreeMap (and hence TreeSet), which are slower by factors of seven or eight. In the case of Hashtable the fact that every method is synchronized, including helper methods to get next key or check for modifications, seems to have played a major part. In the case of the tree-based collections, it is indeed much slower to search for the next entry from the root of the tree, rather than following an incoming pointer.

However, observe that all four of the owners-as-accessors refactored implementations were not significantly different from the original collections.

*2) **List Iteration**:* Figure 12 presents a microbenchmark comparing between the original and refactored versions of the linked list: note the log scale on the y-axis. The graph again shows the time for a single next call. The "disruptive" benchmark for owners-as-dominators version shows linear performance for a single step, (thus $O(N^2)$ overall) while the other iterators (including all owners-as-accessors implementations) behave linearly for different collection sizes (i.e. $O(N)$ overall as expected). As collection sizes get larger, the amortised time for the refactored collections approaches that of the original collections.

*3) **Map Iteration**:* Figure 13 presents results of the MapMicroBenchmark benchmark from JSR166 benchmarks for three different kinds of maps: HashMap, LinkedHashMap, and Hashtable. We used the default parameters for the microbenchark and map sizes and report the averaged results for the largest map size. Observe that owners-as-dominators performs only 20% slower than the original while all four of the owners-as-accessors perform with no more than 3-5% slowdown.

Benchmark	Original		OasD		Proxy		Proxy D.		Indirection		Indirection D.		Number	Percent
DaCapo (Time in *ms*; lower is better)														
avrora	23003	300	**22781**	**415**	**22816**	**236**	22867	179	22948	389	22873	323	219049309	78.08%
batik	2516	34	2517	19	2520	25	2519	19	2519	25	**2528**	**30**	26507124	31.37%
eclipse	53793	1031	53480	936	53716	1010	53812	1329	**53554**	**1406**	53608	738	355465429	33.63%
fop	393	27	397	29	394	20	397	23	396	23	399	20	1874892	18.07%
h2	24133	580	24238	593	24141	517	24188	380	23967	320	23934	375	90175446	8.04%
jython	15041	215	**15476**	**100**	**15725**	**107**	**15719**	**53**	**15837**	**110**	**17050**	**145**	159700109	7.49%
luindex	705	18	**687**	**23**	714	22	713	19	710	40	718	51	327466	36.66%
lusearch	7251	184	7334	105	**7181**	**189**	**7120**	**198**	7226	299	7325	78	11979688	5.45%
pmd	3944	47	**3992**	**39**	**4046**	**59**	**4065**	**72**	**4005**	**64**	**4054**	**67**	10712544	36.55%
sunflow	22656	518	22560	365	**23365**	**126**	22970	711	22851	523	**22331**	**207**	171198077	<0.01%
tomcat	7576	108	7641	135	**7687**	**134**	**7736**	**111**	**7733**	**88**	**7661**	**118**	16726923	13.95%
tradebeans	27952	409	**27556**	**494**	**28258**	**506**	28142	275	27998	328	28020	340	1621619	33.00%
tradesoap	64476	1251	65193	1549	65042	1712	65119	1463	64390	1378	65111	1499	1631193	32.82%
xalan	26604	384	26692	318	**26383**	**247**	**26173**	**258**	**26125**	**251**	**26310**	**291**	61153799	13.23%
SPECjbb2005 (Throughput; higher is better)														
SPECjbb2005	29598	405	**14062**	**181**	**28825**	**860**	**28540**	**619**	**28959**	**764**	**28394**	**641**	35542855	4.87%
SPECjvm2008 (Time in *ms*; lower is better)														
compress	46.56	0.81	46.77	0.59	46.71	0.59	46.82	0.42	46.83	0.42	46.84	0.57	199478	20.21%
crypto.aes	18.49	0.18	18.40	0.13	18.46	0.10	18.48	0.18	**18.42**	**0.10**	18.48	0.19	254853	22.00%
crypto.rsa	35.04	0.25	34.89	0.28	34.92	0.31	34.95	0.36	34.94	0.25	34.95	0.27	6535358	11.18%
crypto.signverify	53.06	0.27	52.97	0.31	53.11	0.29	53.16	0.36	53.09	0.38	53.12	0.29	3290783	2.13%
derby	21.55	0.48	21.63	0.40	21.56	0.50	21.73	0.43	21.59	0.38	21.61	0.33	89061937	3.04%
mpegaudio	15.08	0.05	15.10	0.05	15.08	0.06	15.11	0.05	15.10	0.06	15.07	0.06	209089	20.32%
fft.large	23.61	0.27	23.53	0.30	23.49	0.27	**23.43**	**0.28**	23.55	0.24	23.51	0.38	168290	23.78%
fft.small	82.75	4.49	84.24	5.21	84.98	3.94	83.35	4.01	84.46	4.70	83.57	4.26	439646	9.22%
lu.large	6.98	1.44	6.69	1.24	6.72	1.22	7.05	1.39	6.34	0.96	7.02	1.42	166728	23.92%
lu.small	107.98	0.87	107.61	0.92	107.48	0.75	107.58	0.84	107.94	0.70	107.82	0.85	642713	6.39%
monte_carlo	15.33	1.49	15.33	1.48	15.65	0.10	15.67	0.11	15.63	0.03	15.29	1.47	179253	22.58%
sor.large	13.01	0.02	13.01	0.02	13.01	0.01	13.02	0.02	**12.99**	**0.05**	13.01	0.01	167449	23.92%
sor.small	57.47	0.11	57.47	0.10	57.49	0.10	57.47	0.12	57.43	0.07	57.44	0.10	193425	21.16%
sparse.large	11.51	0.23	**11.97**	**1.22**	11.70	0.28	11.82	0.80	11.71	0.34	11.77	0.36	166691	23.98%
sparse.small	43.87	0.11	43.79	0.12	43.80	0.18	43.83	0.11	43.85	0.18	43.80	0.15	182676	22.17%
serial	33.22	0.98	**32.46**	**0.96**	33.27	0.81	33.04	0.92	32.88	1.31	33.31	0.89	51123977	5.68%
sunflow	20.51	0.50	20.48	0.50	20.55	0.32	20.42	0.52	**20.22**	**0.63**	20.51	0.47	42506559	0.12%
xml.validation	63.38	1.09	63.01	0.96	63.36	1.03	63.19	1.42	63.75	1.39	63.23	1.03	10318760	3.33%

TABLE I

FULL RESULTS TABLE (FORMAT: MEAN|SD; STATISTICALLY SIGNIFICANTLY DIFFERENT VALUES SHOWN IN BOLD)

B. DaCapo, SPECjbb2005, and SPECjvm2008

Table I presents the results of the DaCapo, SPECjbb2005, and SPECjvm2008 benchmarks. For each benchmark, for standard and refactored collections, the time taken was recorded for 25 runs. We wanted to test whether the mean time taken differed between the original and five ownership versions, while controlling for the different times needed for different benchmarks. We show the average times across all runs and include a standard deviation for all DaCapo benchmarks. For SPEC benchmarks we report a throughput in either "bops" or "operations per minute" and again include the standard deviation alongside the mean across 25 runs.

Usually one would use Analysis of Variance to test the null hypothesis that the mean time taken did not differ between the standard and refactored collections. However, ANOVA assumes the data are normally distributed and that the variances of time taken are the same in all $2 \times 14 = 28$ groups. In our data, these assumptions did not hold for all benchmarks. Data transformation did not solve this problem. We therefore used non-parametric methods (which do not require normality nor equality of variances) – in particular the Asymptotic p-value of the Mann-Whitney U test – separately for each benchmark, to determine whether we could reject the null hypothesis that the mean time taken for the standard and refactored collections was the same. To perform these tests, we used PASW Statistics 18 Release 18.0.0 (Jul 30 2000), hosted by Microsoft Windows Server 2003 Standard Edition Service Pack 2, running on an Intel Xeon 5130 2.00GHz with 4GB of RAM.

Table I includes the results of the Mann-Whitney test, showing the refactored results with a statistically significant difference from the performance of the original collections in bold. We reject a null hypothesis at significance level $p < 0.05$. We can reject the null hypotheses that the means are the same for the refactored implementations for 40 of the 165 refactored benchmarks: for the others, we were unable to show a statistically significant difference. We refer the interested reader to a technical report accompanying this paper [21] that contains the obtained p-values and clustering results.

Finally, we also measured the number of objects instantiated during each benchmark run and the percentage of them that were collections from the java.util package [27]. Having observed a significant slowdown for the SPECjbb2005 in the case of owners-as-dominators, we have also measured the time this benchmark spent in the methods of java.util collections as the percentage of objects was low (4.87%). We used −Xrunhprof:cpu=times to obtain such timings and found that the original version spent 7.79% of its time in the collections methods and the owners-as-dominators spent 13.94% of its time in the collections methods (almost all of it in TreeMap). We hypothesise that the lack of caching used in our owners-as-dominators map implementations caused this and we can see a number of ways in which this can be improved similar to the way linked lists can be made faster with caching.

VI. Discussion and Related Work

The results we have presented in the previous section have at least two alternative interpretations. First, incoming aliases into collection implementations are absolutely necessary in specific cases, as refactoring to an encapsulated interface means collection operations' runtime performance will be five to ten times slower than otherwise. Second, incoming aliases into collection implementations are clearly unnecessary in general, as the largest significant slowdown in the DaCapo experiment was less than 2%. The truth, no doubt, lies somewhere between these extremes.

The reasons for the microbenchmark results, at least, seem clear: more general (and reusable) interfaces are also by necessity less efficient. Our refactorings imposed additional hash lookups, and additional list and tree traversals, when an incoming pointer could take the program directly to exactly the right place without any such overhead.

The reasons for the macrobenchmark results are less clear. Perhaps there is a small effect, but the variability introduced by a JITting VM, garbage collector, the underlying operating system mean the effect is lost in the noise. Arguably, however, most programs would not be affected by such a small overhead. Perhaps collections do not make up a significant portion of the DaCapo benchmarks execution time? This is certainly the case for sunflow and luindex, although as Table I illustrates, most of the benchmarks create tens of thousands of collection objects, and some benchmarks create millions. The DaCapo benchmarks have been selected to model realistic Java workloads [25]: if they are a reasonably accurate gauge of the use of collections in Java programs, then we would not expect significantly higher runtime impacts upon other Java applications.

a) Evaluating Ownership: There have been a number of implementation studies evaluating ownership types [28], [11] — some quite extensive [29]. Generally these studies were undertaken in the context of validating a particular type system proposal and the performance evaluation did not specifically concentrate on the costs of various ownership-friendly designs. Many of these studies used the collections library as an example, and were able to check the whole of the collections library, albeit with varying amounts of annotation, depending on the system. The AJ collections reimplementation in particular included some performance analysis using a selection of Java applications and SPECjbb benchmark — they found that a tuned version of AJ collections performed only marginally slower than the standard Java version (Figure 22 in [11], which aligns well with our findings.

These studies differ from the approach we have adopted here because they were mostly based on more flexible (i.e. less encapsulating) ownership disciplines — confining objects within regions rather than per-object ownership [30], or by permitting incoming pointers in some circumstances [28]. The problem is that these supposedly "benign" incoming references can drastically reduce encapsulation.

There have been surprisingly few stand-alone case studies evaluating ownership *per se*. The closest research to this work is Stefan Nägeli who studied how ownership affected design patterns and the Swing GUI library [31], and Cele and Stureborg [32] who implemented three medium-sized programs while respecting an strong ownership discipline. Both studies found that ownership could help structure programs, but could also be a cause of refactoring and redesign: this chimes with our experience. Neither study considered the potential performance cost of the encapsulation enforced by ownership.

b) Inversion of Control: Many of the problems we encountered, particularly with the Map iterators, can be understood in terms of inversion of control: we would have had no problem writing an "internal iterator" method for each collection implementation to iterate over all their elements. Because these methods are encapsulated within each collection class, there is no need (or temptation) to breach encapsulation. The reason, of course, is that in an internal iterator, control and data flow are both efferent, flowing from the collection to the iterator . By contrast, an in external iterator, data flow remains efferent, but control flow is afferent: the external iterator calls in to the collection that hands each element back in turn.

CLU-style generators [33] (as popularised in Python, Ruby, and C♯) are an alternative solution to this problem: programs are written as if they used simple internal iterators — with all the benefits for encapsulation that implies — and then the generator construct inverts the control flow.

c) Further Work: The Java collections continue to evolve. We chose to work with the version 5 collections because that was the version that worked with DaCapo. Repeating this study once closures support is integrated into the main Collection APIs (and once benchmarks have evolved to rely upon those APIs) could address some of the hypotheses above regarding inversion of control.

VII. Conclusion

In this paper we present the first experimental evaluation of the cost of ownership types. We examined the use and breaches of encapsulation in the core classes of the Java Collections Framework. We refactored those classes as necessary to fit the owner-as-dominator and owner-as-accessor encapsulation disciplines. We measured the overhead of these refactorings that showed encapsulation reduces iteration performance by factors of 2 to 8. Finally, we compared the performance of the DaCapo, SPECjbb, and SPECjvm benchmark suites, gaining statistically significant results for a number of benchmarks, SPECjbb owners-as-dominators demonstrating the largest slowdown. Owners-as-accessors slowed down no more than 3% for all benchmarks, even with dynamic ownership checking.

We hope these results may encourage object-oriented designers to consider object encapsulation more carefully when designing their programs — especially their use of incoming aliases to circumvent encapsulation — and to ask themselves: are their incoming aliases really necessary?

ACKNOWLEDGMENTS

Thanks to Dalice Sim, VUW Statistical Consultant, for the statistical analysis; to Doug Lea, Tony Hosking, and Andy Georges, for advice about DaCapo and Collections; to Shane Markstrum for help with JavaCOP; and to Christo Fogelberg, and Andrew Walbran for work on the encapsulated collections; and to the anonymous reviewers of previous versions of this paper. This work is partly supported by the Royal Society of NZ Marsden Fund.

REFERENCES

[1] D. Clarke, J. Potter, and J. Noble, "Ownership types for flexible alias protection," in *OOPSLA*. Vancouver, Canada: ACM Press, Oct. 1998, pp. 48–64.

[2] J. Potter, J. Noble, and D. Clarke, "The ins and outs of objects," in *Australian Software Engineering Conference*. Adelaide, Australia: IEEE Press, November 1998.

[3] T. Zhao, J. Baker, J. Hunt, J. Noble, and J. Vitek, "Implicit ownership types for memory management," *Science of Computer Programming*, vol. 71, no. 3, pp. 213–241, 2008.

[4] D. Clarke, M. Richmond, and J. Noble, "Saving the World from Bad Beans: Deployment-Time Confinement Checking," in *OOPSLA*. Anaheim, CA: ACM Press, 2003, pp. 374–387.

[5] W. Dietl and P. Müller, "Universes: Lightweight ownership for JML," *Journal of Object Technology*, vol. 4, no. 8, pp. 5–32, 2005, http://www.jot.fm/issues/issue_2005_10/article1.

[6] S. Srinivasan and A. Mycroft, "Kilim: Isolation-typed actors for Java," in *ECOOP*, 2007.

[7] B. Bloom, J. Field, N. Nystrom, J. stlund, G. Richards, R. Strnisa, J. Vitek, and T. Wrigstad, "Thorn: robust, concurrent, extensible scripting on the JVM," in *OOPSLA*, 2009, pp. 117–136.

[8] P. Haller and M. Odersky, "Capabilities for uniqueness and borrowing," in *ECOOP*, 2010, pp. 354–378.

[9] J. H. Spring, J. Privat, R. Guerraoui, and J. Vitek, "Streamflex: high-throughput stream programming in Java," in *OOPSLA*, 2007.

[10] R. L. Bocchino, Jr., V. S. Adve, D. Dig, S. V. Adve, S. Heumann, R. Komuravelli, J. Overbey, P. Simmons, H. Sung, and M. Vakilian, "A Type and Effect System for Deterministic Parallel Java," in *OOPSLA*, 2009.

[11] J. Dolby, C. Hammer, D. Marino, F. Tip, M. Vaziri, and J. Vitek, "A data-centric approach to synchronization," *ACM Trans. Program. Lang. Syst.*, vol. 34, no. 1, p. 4, 2012.

[12] J. Noble, J. Vitek, and J. Potter, "Flexible Alias Protection," in *ECOOP*, vol. 1445. Springer-Verlag, Jul. 1998, pp. 158–185.

[13] J. Aldrich and C. Chambers, "Ownership domains: Separating aliasing policy from mechanism," in *ECOOP*, vol. 3086. Oslo, Norway: Springer-Verlag, Jun. 2004, pp. 1–25.

[14] P. Müller and A. Poetzsh-Heffter, "Programming Languages and Fundamentals of Programming," Fernuniversität Hagen, Tech. Rep., 2001, poetzsh-Heffter, A. and Meyer, J. (editors).

[15] D. Gordon and J. Noble, "Dynamic ownership in a dynamic language," in *DLS Proceedings*, 2007, pp. 9–16.

[16] S. Markstrum, D. Marino, M. Esquivel, T. Millstein, C. Andreae, and J. Noble, "JavaCOP: Declarative pluggable types for Java," *TOPLAS*, vol. 32, no. 2, 2010.

[17] N. R. Cameron, J. Noble, and T. Wrigstad, "Tribal ownership," in *OOPSLA*, 2010, pp. 618–633.

[18] M. Naftalin and P. Wadler, *Java Generics and Collections*. O'Reilly Media, Inc., 2006.

[19] J. Bloch, *The Java Tutorials — Collections*, Oracle Corporation, 2011, http://download.oracle.com/javase/tutorial/collections/.

[20] Oracle Corporation, "Java collections framework overview," http://download.oracle.com/javase/1.5.0/docs/guide/collections/overview.html, 2011, accessed September 2011.

[21] A. Potanin, M. Damitio, and J. Noble, "Ownership and collections: The statistical analysis," ECS, VUW, http://ecs.victoria.ac.nz/Main/TechnicalReportSeries, Tech. Rep. ECSTR12-22, 2012.

[22] J. Östlund and T. Wrigstad, "Multiple aggregate entry points for ownership types," in *ECOOP*, ser. LNCS, vol. 7313, 2012, pp. 156–180.

[23] J. Noble, "Iterators and encapsulation," in *TOOLS Europe*. IEEE Press, 2000, pp. 431–442.

[24] D. Lea, "Jsr166 loops benchmarks," http://gee.cs.oswego.edu/cgi-bin/viewcvs.cgi/jsr166/src/test/loops/, 2012.

[25] S. M. Blackburn, R. Garner, C. Hoffmann, A. M. Khang, K. S. McKinley, R. Bentzur, A. Diwan, D. Feinberg, D. Frampton, S. Z. Guyer, M. Hirzel, A. Hosking, M. Jump, H. Lee, J. E. B. Moss, A. Phansalkar, D. Stefanovic, T. VanDrunen, D. von Dincklage, and B. Wiedermann, "The DaCapo benchmarks: Java benchmarking development and analysis," in *OOPSLA*, Oct. 2006, pp. 169–190.

[26] A. Georges, D. Buytaert, and L. Eeckhout, "Statistically rigorous Java performance evaluation," in *OOPSLA*. Montreal, Quebec, Canada: ACM Press, Oct. 2007, pp. 57–76.

[27] S. Nelson, D. J. Pearce, and J. Noble, "Profiling object initialization for java," in *Third International Conference on Runtime Verification*, ser. RV 2012. Springer-Verlag, September 2012, pp. 292–307.

[28] Y. Zibin, A. Potanin, S. Artzi, A. Kiezun, and M. D. Ernst, "Object and reference immutability using Java generics," in *Foundations of Software Engineering*, 2007.

[29] J. Aldrich, V. Kostadinov, and C. Chambers, "Alias Annotations for Program Understanding," in *OOPSLA*. Seattle, WA, USA: ACM Press, Nov. 2002, pp. 311–330.

[30] C. Grothoff, J. Palsberg, and J. Vitek, "Encapsulating Objects with Confined Types," *TOPLAS*, vol. 29, no. 6, 2007.

[31] S. Nägeli, "Ownership in design patterns," Master's thesis, Software Component Technology Group, Department of Computer Science, ETH Zurich, 2006.

[32] G. Cele and S. Stureborg, "Ownership types in practice," Master's thesis, Department of Computer and Systems Sciences, Stockholm University, Jan. 2005.

[33] B. Liskov and J. V. Guttag, *Abstraction and Specification in Program Development*. MIT Press/McGraw-Hill, 1986.

Collection	Data	avrora	batik	eclipse	fop	h2	jython	luindex	lusearch	pmd	sunflow	tomcat	tradebeans	tradesoap	xalan
		large	large	large	default	large	large	default	large	large	large	large	large	large	large
ArrayList	instances	69	19581	*379942*	70896	*21786*	2861339	1360	*630387*	*1313504*	225	*385901*	*169892*	*7174701*	*53623*
	max size	26	23	301	821	80660	4901	23	23	7971	23	217	195	195	23
	avrg size	4.88	0.18	*1.07*	31.37	*45.70*	242.64	1.91	2.67	1.22	1.49	*2.54*	*8.64*	*1.93*	1.57
	itr	38	4713	*79509*	5208	*21796*	58658	32	30	478847	58	*286199*	*125632*	*2748582*	84
	listItr	10	0	360	24680	*0*	360393	3	0	33120	1	2	34132	11008	0
	clone	0	0	0	0	0	0	10	0	0	0	3	3	*32157*	0
Vector	instances	36	*2987*	*13306*	2539	58	10003	46	176	2089	98	*402835*	*9009*	*2313696*	*192182*
	max size	0	3	564	564	564	564	564	0	0	5	564	564	564	2
	avrg size	0.00	*1.62*	282.50	1.13	282.50	282.50	189.00	0.00	0.00	5.00	8.29	*13.57*	*0.43*	0.67
	itr	0	11	0	1533	0	0	0	0	0	1	3517	*79*	*192949*	0
	listItr	0	0	0	2942	0	0	1	0	0	0	4	0	0	0
	enum elmt	0	*148*	4	2	4	2	2	0	0	0	10	*201*	*193063*	3
	clone	0	*770*	0	0	0	0	8	0	0	0	0	0	0	0
LinkedList	instances	880665	18813	51	53001	1	6	6	1	*62179*	5	*24137*	*14970*	*143703*	2
	max size	37	11	1	832	1	1	6	1	28	1	1	98	98	1
	avrg size	*1.24*	0.25	1.00	2.53	1.00	0.50	3.39	1.00	2.13	1.00	0.60	1.13	*1.09*	1.00
	itr	*3878507*	30100	1	896	1	2	18	1	374926	1	5	2349	*2438*	1
	listItr	*3878507*	30113	1	16176	1	2	18	1	374926	1	5	2395	*2484*	1
	clone	0	0	0	0	0	0	0	0	0	0	0	0	0	0
PriorityQueue	instances	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	max size	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	avrg size	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
	itr	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	clone	0	0	0	0	0	0	0	0	0	0	0	0	0	0
HashSet	instances	18	2759	*117408*	184	*1698765*	3263025	39	*135*	*335139*	36	*11727*	*1335733*	*1223071*	35
	max size	0	17	7847	7	1	22	6	0	19	22	34	256	256	0
	avrg size	0.00	11.87	*1.41*	3.42	0.23	0.75	2.48	0.00	6.43	2.18	*0.15*	*1.21*	*0.28*	0.00
	itr	0	1719	*167229*	205	56	5441	23	0	7	11	*167285*	*70003*	*695845*	0
	clone	0	0	*249*	0	0	0	0	0	0	0	0	0	0	0
LinkedHashSet	instances	0	0	*924*	0	0	1	0	0	0	1	1	5442	5442	0
	max size	0	0	35	0	0	22	0	0	0	22	22	45	45	0
	avrg size	0.00	0.00	*10.02*	0.00	0.00	22.00	0.00	0.00	0.00	22.00	22.00	*1.65*	*1.30*	0.00
	itr	0	0	*924*	0	0	1	0	0	0	1	1	*28631*	*60753*	0
	clone	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TreeSet	instances	0	0	2	10	0	0	0	0	573	10	1	1830	1830	0
	max size	0	0	63	0	0	0	0	0	402	0	0	242	242	0
	avrg size	0.00	0.00	63.00	0.00	0.00	0.00	0.00	0.00	1.02	0.00	0.00	0.87	0.87	0.00
	itr	0	0	1	0	0	0	0	0	571	0	0	1772	1772	0
	clone	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Collection	Data	avrora	batik	eclipse	fop	h2	jython	luindex	lusearch	pmd	sunflow	tomcat	tradebeans	tradesoap	xalan
		large	large	large	default	large	large	default	large	large	large	large	large	large	large
Hashtable	instances	23	2004	*68257*	114	41	1030	38	26	891	33	*1518*	*3302*	*453584*	*170123*
	max size	0	47	515	76	54	390	0	0	5	0	483	893	893	192
	avrg size	0.00	7.00	*8.68*	11.46	30.43	2.68	0.00	0.00	1.11	0.00	*1.74*	*18.18*	*1.09*	*7.20*
	itr	0	1439	*30504*	1	7	1	0	0	104	0	98	*163*	*162*	1
	itr key	0	1	11	0	6	0	0	0	0	0	0	9	9	0
	itr value	0	2	29176	0	0	0	0	0	0	0	74	0	0	0
	itr entry	0	1436	*1317*	1	1	1	0	0	104	0	24	*154*	153	1
	enum	0	8	*1312*	12	0	839	0	0	1	0	*6787*	*337*	*96753*	*85008*
	enum key	0	8	*1312*	12	0	794	0	0	1	0	*6756*	*327*	*96743*	*85008*
	enum elmt	0	0	0	0	0	45	0	0	0	0	31	10	10	0
	clone	0	752	*5*	2	2	0	0	0	0	0	6	*65*	*65*	*10*
HashMap	instances	290	8309	*196992*	11807	*1827552*	3264459	165	1364	*363559*	969	*35442*	*2956757*	*3552090*	*110216*
	max size	35	67	7847	67	67	519	67	130	580	402	483	952	952	192
	avrg size	17.10	8.32	*1.90*	3.18	3.31	2.72	4.66	2.98	19.97	18.66	*0.21*	*3.53*	*1.26*	*191.36*
	itr	69	10632	*187996*	501	78	8166	47	260	231173	44	*171109*	*372010*	*1160362*	*34120*
	itr key	42	5592	*169948*	342	58	5446	26	129	230027	29	*170593*	*95095*	*658889*	*117*
	itr value	27	76	*11573*	72	18	936	3	129	572	4	*454*	*4188*	*164429*	1
	itr entry	0	4964	*6476*	87	2	1784	18	2	574	11	62	*272726*	*337045*	34002
	clone	0	2088	*1013*	0	0	54	0	0	0	0	5	4	4	34000
LinkedHashMap	instances	6	8	*946*	7	43	863	7	136	13	12	17	*8840*	*8840*	8
	max size	0	67	200	67	67	70	67	130	200	67	84	133	133	67
	avrg size	0.00	32.25	*11.38*	28.75	28.75	4.83	38.33	80.33	9.48	36.50	*17.89*	*1.97*	*1.54*	*6.41*
	itr	2	4	*952*	4	4	2267	3	3	608	8	*17*	*40248*	*72386*	117
	itr key	2	2	*945*	2	2	3	1	1	604	3	*14*	*29007*	*61145*	115
	itr value	0	0	1	0	0	672	0	0	0	1	1	*866*	*866*	0
	itr entry	0	2	6	2	2	1592	2	2	4	4	2	*10375*	*10375*	2
	clone	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TreeMap	instances	0	2	3	12	0	18	0	0	*573*	10	*40*	3486	*3498*	0
	max size	0	33	64	0	0	0	0	0	402	0	20	242	242	0
	avrg size	0.00	22.00	*49.18*	0.00	0.00	0.00	0.00	0.00	1.02	0.00	*6.17*	1.13	*1.13*	0.00
	itr	0	2	*10*	0	0	0	0	0	571	0	*38*	8792	*8804*	0
	itr key	0	2	1	0	0	0	0	0	571	0	5	1772	*1773*	0
	itr value	0	0	0	0	0	0	0	0	0	0	*33*	36	*47*	0
	itr entry	0	0	*9*	0	0	0	0	0	0	0	0	6984	*6985*	0
	clone	0	0	0	0	0	0	0	0	0	0	0	0	0	0

6 August, 2012

James Noble:

For each benchmark (14 different benchmarks), for each "test" (6 different tests, standard, owners, memento, memento dynamic, no proxy, and no proxy dynamic), the time taken was recorded for 25 runs. We wanted to test whether the mean time taken differed between standard and each of the other tests, while controlling for the different times needed for different benchmarks. The actual means ± standard deviations were:

Descriptive Statistics

Dependent Variable:Times (ms)

test	Bench	Mean	Std. Deviation	N
standard	avrora	23002.80	299.673	25
	batik	2515.56	34.420	25
	eclipse	53793.08	1030.629	25
	fop	392.52	27.157	25
	h2	24133.48	580.308	25
	jython	15041.32	215.103	25
	luindex	705.32	18.416	25
	lusearch	7250.92	184.212	25
	pmd	3943.96	46.644	25
	sunflow	22655.68	517.895	25
	tomcat	7575.60	108.105	25
	tradebeans	27951.68	408.908	25
	tradesoap	64475.84	1251.264	25
	xalan	26604.28	383.737	25
	Total	20003.00	18783.548	350
owners	avrora	22781.04	415.494	25
	batik	2517.16	19.146	25
	eclipse	53480.28	936.434	25
	fop	396.84	29.054	25
	h2	24238.08	593.387	25
	jython	15475.64	99.539	25
	luindex	687.40	22.873	25
	lusearch	7334.04	104.926	25
	pmd	3992.04	38.828	25
	sunflow	22559.80	364.581	25
	tomcat	7640.52	134.604	25
	tradebeans	27556.40	493.813	25
	tradesoap	65193.32	1548.789	25

	xalan	26692.32	318.249	25
	Total	20038.92	18838.867	350
memento	avrora	22815.60	236.415	25
	batik	2519.76	25.186	25
	eclipse	53716.24	1009.544	25
	fop	393.72	20.126	25
	h2	24141.40	517.154	25
	jython	15725.00	106.620	25
	luindex	713.92	22.057	25
	lusearch	7181.08	189.265	25
	pmd	4046.20	59.333	25
	sunflow	23365.40	126.337	25
	tomcat	7687.12	133.845	25
	tradebeans	28258.32	505.726	25
	tradesoap	65041.92	1712.479	25
	xalan	26383.48	247.099	25
	Total	20142.08	18860.387	350
mem dynamic	avrora	22867.28	178.693	25
	batik	2518.96	19.182	25
	eclipse	53811.68	1328.635	25
	fop	397.00	23.071	25
	h2	24188.28	379.774	25
	jython	15719.44	53.046	25
	luindex	712.52	19.339	25
	lusearch	7120.20	197.800	25
	pmd	4064.64	72.432	25
	sunflow	22970.00	710.582	25
	tomcat	7735.64	110.630	25
	tradebeans	28142.32	274.980	25
	tradesoap	65119.48	1462.600	25
	xalan	26173.08	257.892	25
	Total	20110.04	18873.821	350
no proxy	avrora	22947.88	389.192	25
	batik	2519.40	25.426	25
	eclipse	53553.80	1405.639	25
	fop	396.12	22.720	25
	h2	23967.12	319.725	25
	jython	15837.28	109.988	25
	luindex	710.48	40.050	25
	lusearch	7226.04	298.639	25

	pmd	4005.32	64.303	25
	sunflow	22851.36	522.582	25
	tomcat	7733.08	88.124	25
	tradebeans	27998.20	327.537	25
	tradesoap	64389.96	1377.639	25
	xalan	26124.64	251.450	25
	Total	20018.62	18703.768	350
no proxy dynamic	avrora	22873.32	322.942	25
	batik	2528.40	30.092	25
	eclipse	53608.12	738.360	25
	fop	399.16	19.861	25
	h2	23934.00	374.680	25
	jython	17050.08	145.138	25
	luindex	718.36	50.717	25
	lusearch	7324.84	78.230	25
	pmd	4053.52	67.293	25
	sunflow	22330.64	207.213	25
	tomcat	7660.68	117.922	25
	tradebeans	28020.00	339.848	25
	tradesoap	65110.96	1498.743	25
	xalan	26310.32	290.756	25
	Total	20137.31	18807.353	350
Total	avrora	22881.32	321.387	150
	batik	2519.87	26.039	150
	eclipse	53660.53	1087.214	150
	fop	395.89	23.607	150
	h2	24100.39	478.450	150
	jython	15808.13	628.730	150
	luindex	708.00	32.418	150
	lusearch	7239.52	200.989	150
	pmd	4017.61	71.907	150
	sunflow	22788.81	555.301	150
	tomcat	7672.11	127.433	150
	tradebeans	27987.82	450.989	150
	tradesoap	64888.58	1493.342	150
	xalan	26381.35	357.949	150
	Total	20075.00	18789.044	2100

Once again, we had a problem in that the data did not meet the traditional assumptions of normal distribution and equality of variances between groups. We therefore used the Mann-Whitney test to compare "tests" (i.e., standard vs each of the others) within each benchmark. This test essentially does a two-sample t test on the ranks of the data instead of the data itself.

Bench = avrora

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	30.34	758.50
	owners	25	20.66	516.50
	Total	50		

Mann-Whitney U = 191.5, Z = -2.348, p = 0.019. We conclude that owners has a significantly lower mean than standard for benchmark avrora.

Standard vs Memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	30.32	758.00
	memento	25	20.68	517.00
	Total	50		

Mann-Whitney U = 192.0, Z = -2.338, p = 0.019. We conclude that memento has a significantly lower mean than standard for benchmark avrora.

Standard vs Mem Dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	29.40	735.00
	mem dynamic	25	21.60	540.00
	Total	50		

Mann-Whitney U = 215.0, Z = -1.892, p = 0.059. We conclude that the mean difference in time for Mem dynamic vs standard is not statistically significant, but shows a trend towards statistical significance (since p < 0.10).

Standard vs No Proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	26.08	652.00
	no proxy	25	24.92	623.00
	Total	50		

Mann-Whitney U = 298.0, Z = -0.281, p = 0.778. We conclude that there is no difference in the mean time for no proxy vs standard.

Standard vs No proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	28.00	700.00
	no proxy dynamic	25	23.00	575.00
	Total	50		

Mann-Whitney U = 250.0, Z = -1.213, p = 0.225. We conclude that there is no difference in the mean time for no proxy dynamic vs standard.

Bench = batik

Standard vs Owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.68	592.00
	owners	25	27.32	683.00
	Total	50		

Mann-Whitney U = 267.0, Z = -0.883, p = 0.377. We conclude that there is no difference in the mean time for owners vs standard.

Standard vs Memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.84	596.00
	memento	25	27.16	679.00
	Total	50		

Mann-Whitney U = 271.0, Z = -0.806, p = 0.420. We conclude that there is no difference in the mean time for memento vs standard.

Standard vs Mem Dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.20	580.00
	mem dynamic	25	27.80	695.00
	Total	50		

Mann-Whitney U = 255.0, Z = -1.116, p = 0.264. We conclude that there is no difference in the mean time for mem dynamic vs standard.

Standard vs No Proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.72	593.00
	no proxy	25	27.28	682.00
	Total	50		

Mann-Whitney U = 268.0, Z = -0.864, p = 0.388. We conclude that there is no difference in the mean time for no proxy vs standard.

Standard vs No proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	21.36	534.00
	no proxy dynamic	25	29.64	741.00
	Total	50		

Mann-Whitney U = 209.0, Z = - 2.010, p = 0.044. We conclude that there is a statistically significant difference in the mean time for no proxy dynamic vs standard.

Bench = eclipse

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	27.80	695.00
	owners	25	23.20	580.00
	Total	50		

Mann-Whitney U = 255.0, Z = - 1.116, p = 0.265. We conclude that there is not a statistically significant difference in the mean time for owners vs standard.

Standard vs Memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	25.92	648.00
	memento	25	25.08	627.00
	Total	50		

Mann-Whitney U = 302.0, Z = - 0.204, p = 0.839. We conclude that there is not a statistically significant difference in the mean time for memento vs standard.

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	26.40	660.00
	mem dynamic	25	24.60	615.00
	Total	50		

Mann-Whitney U = 290.0, Z = - 0.437, p = 0.662. We conclude that there is not a statistically significant difference in the mean time for mem dynamic vs standard.

Standard vs No proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	29.76	744.00
	no proxy	25	21.24	531.00
	Total	50		

Mann-Whitney U = 206.0, Z = - 2.067, p = 0.039. We conclude that there is a statistically significant difference in the mean time for no proxy vs standard.

Standard vs No proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	26.06	651.50
	no proxy dynamic	25	24.94	623.50
	Total	50		

Mann-Whitney U = 298.5, Z = - 0.272, p = 0.786. We conclude that there is no a statistically significant difference in the mean time for no proxy dynamic vs standard.

Bench = fop

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	24.56	614.00
	owners	25	26.44	661.00
	Total	50		

Mann-Whitney U = 289.0, Z = - 0.456, p = 0.648. We conclude that there is no a statistically significant difference in the mean time for owners vs standard.

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	24.80	620.00
	memento	25	26.20	655.00
	Total	50		

Mann-Whitney U = 295.0, Z = - 0.340, p = 0.734. We conclude that there is not a statistically significant difference in the mean time for memento vs standard.

Standard vs Mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.70	592.50
	mem dynamic	25	27.30	682.50
	Total	50		

Mann-Whitney U = 267.0, Z = - 0.873, p = 0.384. We conclude that there is not a statistically significant difference in the mean time for mem dynamic vs standard

Standard vs No Proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	24.08	602.00
	no proxy	25	26.92	673.00
	Total	50		

Mann-Whitney U = 277.0, Z = - 0.689, p = 0.491. We conclude that there is not a statistically significant difference in the mean time for no proxy vs standard

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.56	564.00
	no proxy dynamic	25	28.44	711.00
	Total	50		

Mann-Whitney U = 239.0, Z = - 1.427, p = 0.154. We conclude that there is not a statistically significant difference in the mean time for no proxy dynamic vs standard

Bench = h2

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	24.16	604.00
	owners	25	26.84	671.00
	Total	50		

Mann-Whitney U = 279.0, Z = - 0.650, p = 0.516. We conclude that there is not a statistically significant difference in the mean time for owners vs standard

Standard vs Memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	25.00	625.00
	memento	25	26.00	650.00
	Total	50		

Mann-Whitney U =300.0, Z = - 0.243, p = 0.808. We conclude that there is not a statistically significant difference in the mean time for memento vs standard

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.60	590.00
	mem dynamic	25	27.40	685.00
	Total	50		

Mann-Whitney U = 265.0, Z = - 0.922, p = 0.357. We conclude that there is not a statistically significant difference in the mean time for mem dynamic vs standard

Standard vs No proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	26.50	662.50
	no proxy	25	24.50	612.50
	Total	50		

Mann-Whitney U = 287.5, Z = - 0.485, p = 0.628. We conclude that there is not a statistically significant difference in the mean time for no proxy vs standard

Standard vs No proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	27.32	683.00
	no proxy dynamic	25	23.68	592.00
	Total	50		

Mann-Whitney U = 267.0, Z = - 0.883, p = 0.377. We conclude that there is not a statistically significant difference in the mean time for no proxy dynamic vs standard

Bench = jython

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	14.00	350.00
	owners	25	37.00	925.00
	Total	50		

Mann-Whitney U = 25.0, Z = - 5.579, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for owners vs standard.

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	13.00	325.00
	memento	25	38.00	950.00
	Total	50		

Mann-Whitney U = 0.000, Z = - 6.064, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for memento vs standard.

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	13.00	325.00
	mem dynamic	25	38.00	950.00
	Total	50		

Mann-Whitney U = 0.000, Z = - 6.065, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for mem dynamic vs standard

Standard vs No Proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	13.00	325.00
	no proxy	25	38.00	950.00
	Total	50		

Mann-Whitney U = 0.000, Z = - 6.064, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for no proxy vs standard

Standard vs No proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	13.00	325.00
	no proxy dynamic	25	38.00	950.00
	Total	50		

Mann-Whitney U = 0.000, Z = - 6.065, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for no proxy dynamic vs standard

Bench = luindex

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	31.82	795.50
	owners	25	19.18	479.50
	Total	50		

Mann-Whitney U = 154.5, Z = - 3.067, p =0.002. We conclude that there is a statistically significant difference in the mean time for owners vs standard

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.78	569.50
	memento	25	28.22	705.50
	Total	50		

Mann-Whitney U = 244.5, Z = - 1.320, p =0.187. We conclude that there is not a statistically significant difference in the mean time for memento vs standard

Standard vs Mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.96	574.00
	mem dynamic	25	28.04	701.00
	Total	50		

Mann-Whitney U = 249.0, Z = - 1.233, p =0.218. We conclude that there is not a statistically significant difference in the mean time for mem dynamic vs standard

Standard vs no proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	25.58	639.50
	no proxy	25	25.42	635.50
	Total	50		

Mann-Whitney U = 310.5, Z = - 0.039, p =0.969. We conclude that there is not a statistically significant difference in the mean time for no proxy vs standard

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	24.06	601.50
	no proxy dynamic	25	26.94	673.50
	Total	50		

Mann-Whitney U = 376.5, Z = - 0.699, p =0.485. We conclude that there is not a statistically significant difference in the mean time for no proxy dynamic vs standard

Bench = lusearch

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.06	576.50
	owners	25	27.94	698.50
	Total	50		

Mann-Whitney U = 251.5, Z = - 1.184, p =0.237. We conclude that there is not a statistically significant difference in the mean time for owners vs standard

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	30.32	758.00
	memento	25	20.68	517.00
	Total	50		

Mann-Whitney U = 192.0, Z = - 2.338, p =0.019. We conclude that there is a statistically significant difference in the mean time for memento vs standard

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	29.96	749.00
	mem dynamic	25	21.04	526.00
	Total	50		

Mann-Whitney U = 201.0, Z = - 2.163, p =0.031. We conclude that there is a statistically significant difference in the mean time for mem dynamic vs standard

Standard vs no proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	26.52	663.00
	no proxy	25	24.48	612.00
	Total	50		

Mann-Whitney U = 287.0, Z = - 0.495, p =0.631. We conclude that there is not a statistically significant difference in the mean time for no proxy vs standard

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	23.80	595.00
	no proxy dynamic	25	27.20	680.00
	Total	50		

Mann-Whitney U = 270.0, Z = - 0.825, p =0.410. We conclude that there is not a statistically significant difference in the mean time for no proxy dynamic vs standard

Bench = pmd

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	18.48	462.00
	owners	25	32.52	813.00
	Total	50		

Mann-Whitney U = 137.0, Z = - 3.406, p =0.001. We conclude that there is a statistically significant difference in the mean time for owners vs standard

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	15.12	378.00
	memento	25	35.88	897.00
	Total	50		

Mann-Whitney U = 53.0, Z = - 5.036, p < 0.0005. We conclude that there is a statistically significant difference in the mean time for memento vs standard

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	14.64	366.00
	mem dynamic	25	36.36	909.00
	Total	50		

Mann-Whitney U = 41.0, Z = - 5.268, p < 0.0005. We conclude that there is a statistically significant difference in the mean time for mem dynamic vs standard

Standard vs no proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	18.46	461.50
	no proxy	25	32.54	813.50
	Total	50		

Mann-Whitney U = 136.5, Z = - 3.415, p =0.001. We conclude that there is a statistically significant difference in the mean time for no proxy vs standard

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	15.70	392.50
	no proxy dynamic	25	35.30	882.50
	Total	50		

Mann-Whitney U = 67.5, Z = - 4.755, p < 0.0005. We conclude that there is a statistically significant difference in the mean time for no proxy dynamic vs standard

Bench = sunflow

Standard vs Owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	26.48	662.00
	owners	25	24.52	613.00
	Total	50		

Mann-Whitney U = 288.0, Z = - 0.475, p =0.635. We conclude that there is not a statistically significant difference in the mean time for owners vs standard

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	15.42	385.50
	memento	25	35.58	889.50
	Total	50		

Mann-Whitney U = 60.5, Z = - 4.890, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for memento vs standard

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	21.60	540.00
	mem dynamic	25	29.40	735.00
	Total	50		

Mann-Whitney U = 215.0, Z = - 1.892, p =0.059. We conclude that there is not a statistically significant difference in the mean time for mem dynamic vs standard, although there is a trend towards statistical significance ($p < 0.10$).

Standard vs No proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.52	563.00
	no proxy	25	28.48	712.00
	Total	50		

Mann-Whitney U = 238.0, Z = - 1.446, p =0.148. We conclude that there is not a statistically significant difference in the mean time for no proxy vs standard.

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	30.24	756.00
	no proxy dynamic	25	20.76	519.00
	Total	50		

Mann-Whitney U = 194.0, Z = - 2.299, p =0.021. We conclude that there is a statistically significant difference in the mean time for no proxy dynamic vs standard.

Bench = tomcat

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.66	566.50
	owners	25	28.34	708.50
	Total	50		

Mann-Whitney U = 241.5, Z = - 1.378, p =0.168. We conclude that there is not a statistically significant difference in the mean time for owners vs standard.

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	19.46	486.50
	memento	25	31.54	788.50
	Total	50		

Mann-Whitney U = 161.5, Z = - 2.930, p =0.003. We conclude that there is a statistically significant difference in the mean time for memento vs standard.

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	16.84	421.00
	mem dynamic	25	34.16	854.00
	Total	50		

Mann-Whitney U = 96.0, Z = - 4.201, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for mem dynamic vs standard.

Standard vs no proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	16.24	406.00
	no proxy	25	34.76	869.00
	Total	50		

Mann-Whitney U = 81.0, Z = - 4.492, $p < 0.0005$. We conclude that there is a statistically significant difference in the mean time for no proxy vs standard.

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	20.94	523.50
	no proxy dynamic	25	30.06	751.50
	Total	50		

Mann-Whitney U = 198.5, Z = - 2.212, p =0.027. We conclude that there is a statistically significant difference in the mean time for no proxy dynamic vs standard.

Bench = tradebeans

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	31.32	783.00
	owners	25	19.68	492.00
	Total	50		

Mann-Whitney U = 167.0, Z = - 2.823, p =0.005. We conclude that there is a statistically significant difference in the mean time for owners vs standard.

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	20.96	524.00
	memento	25	30.04	751.00
	Total	50		

Mann-Whitney U = 199.0, Z = - 2.202, p =0.028. We conclude that there is a statistically significant difference in the mean time for memento vs standard.

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.44	561.00
	mem dynamic	25	28.56	714.00
	Total	50		

Mann-Whitney U = 236.0, Z = - 1.484, p =0.138. We conclude that there is not a statistically significant difference in the mean time for mem dynamic vs standard.

Standard vs no proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	25.08	627.00
	no proxy	25	25.92	648.00
	Total	50		

Mann-Whitney U = 302.0, Z = - 0.204, p =0.839. We conclude that there is not a statistically significant difference in the mean time for no proxy vs standard

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	24.74	618.50
	no proxy dynamic	25	26.26	656.50
	Total	50		

Mann-Whitney U = 293.5, Z = - 0.369, p =0.712. We conclude that there is not a statistically significant difference in the mean time for no proxy dynamic vs standard

Bench = tradesoap

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.60	565.00
	owners	25	28.40	710.00
	Total	50		

Mann-Whitney U = 240.0, Z = - 1.407, p =0.160. We conclude that there is not a statistically significant difference in the mean time for owners vs standard

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.60	565.00
	memento	25	28.40	710.00
	Total	50		

Mann-Whitney U = 240.0, Z = - 1.407, p =0.160. We conclude that there is not a statistically significant difference in the mean time for memento vs standard. (Note that the stats are exactly the same for owners and memento – I double checked it).

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.18	554.50
	mem dynamic	25	28.82	720.50
	Total	50		

Mann-Whitney U = 229.5, Z = - 1.610, p =0.107. We conclude that there is not a statistically significant difference in the mean time for mem dynamic vs standard.

Standard vs no proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	25.74	643.50
	no proxy	25	25.26	631.50
	Total	50		

Mann-Whitney U = 306.5, Z = - 0.116, p =0.907. We conclude that there is not a statistically significant difference in the mean time for no proxy vs standard.

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	22.20	555.00
	no proxy dynamic	25	28.80	720.00
	Total	50		

Mann-Whitney U = 230.5, Z = - 1.601, p =0.109. We conclude that there is not a statistically significant difference in the mean time for no proxy dynamic vs standard.

Bench = xalan

Standard vs owners

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	24.08	602.00
	owners	25	26.92	673.00
	Total	50		

Mann-Whitney U = 277.0, Z = - 0.689, p =0.491. We conclude that there is not a statistically significant difference in the mean time for owners vs standard.

Standard vs memento

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	30.56	764.00
	memento	25	20.44	511.00
	Total	50		

Mann-Whitney U = 186.0, Z = - 2.455, p =0.014. We conclude that there is a statistically significant difference in the mean time for memento vs standard.

Standard vs mem dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	33.52	838.00
	mem dynamic	25	17.48	437.00
	Total	50		

Mann-Whitney U = 112.0, Z = - 3.890, p < 0.0005. We conclude that there is a statistically significant difference in the mean time for mem dynamic vs standard.

Standard vs no proxy

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	33.92	848.00
	no proxy	25	17.08	427.00
	Total	50		

Mann-Whitney U = 102.0, Z = - 4.084, p < 0.0005. We conclude that there is a statistically significant difference in the mean time for no proxy vs standard.

Standard vs no proxy dynamic

Ranks

	test	N	Mean Rank	Sum of Ranks
Times (ms)	standard	25	31.48	787.00
	no proxy dynamic	25	19.52	488.00
	Total	50		

Mann-Whitney U = 163.0, Z = - 2.901, p =0.004. We conclude that there is a statistically significant difference in the mean time for no proxy dynamic vs standard.

17 August, 2012

James Noble:

SPECjbb2005 data.

For each "test" (6 different tests, standard, OasD, mem, memdyn, noproxy, noprodyn), the time taken was recorded for 25 runs. We wanted to test whether the mean time taken differed between standard and each of the other tests The actual means $\pm$ standard deviations were:

Report

bops (the higher the better)

1=std 2=OasD 3=mem 4=mem dyn 5=no proxy 6=no pro dyn	Mean	N	Std. Deviation	Median
1	29597.96	25	404.619	29557.00
2	14061.68	25	180.795	14047.00
3	28825.24	25	859.751	28724.00
4	28540.36	25	619.492	28561.00
5	28959.00	25	764.066	29017.00
6	28393.68	25	640.569	28326.00
Total	26396.32	150	5581.431	28615.00

The first analysis we did was a one-way Analysis of Variance. This tests the hypothesis that the mean bops are not different for the different "tests". It assumes the data are normally distributed, which they probably are not, but enables us to look at pairwise differences.

The ANOVA table is:

Tests of Between-Subjects Effects

Dependent Variable:bops (the higher the better)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	4.586E9	5	9.172E8	2378.857	.000
Intercept	1.045E11	1	1.045E11	271059.943	.000
@1std2OasD3mem4memdyn5noproxy6noprodyn	4.586E9	5	9.172E8	2378.857	.000
Error	55523288.160	144	385578.390		
Total	1.092E11	150			
Corrected Total	4.642E9	149			

a. R Squared = .988 (Adjusted R Squared = .988)

Thus there is a highly significant difference in mean bops by "test" ($F(5, 144) = 2378.857$, $p < 0.0005$). We conclude that at least one group mean is different from the others.

We did Post Hoc tests to compare between the "tests", to see which means were different from which. These are best summarized by the Table of homogeneous subsets, defined by the Tukey multiple comparisons test.

bops (the higher the better)

	1=std 2=OasD 3=mem 4=mem dyn 5=no proxy 6=no pro dyn	N	Subset 1	Subset 2	Subset 3	Subset 4
Tukey HSD[a,b]	2	25	14061.68			
	6	25		28393.68		
	4	25		28540.36	28540.36	
	3	25		28825.24	28825.24	
	5	25			28959.00	
	1	25				29597.96
	Sig.		1.000	.144	.169	1.000

This table tells us that test 1 (std) has the highest mean bops (29597.96), and that mean is significantly different from the means of all the other tests (at least at the p = 0.05) level. Also, test 2 (OasD) has the lowest mean number of bops (14061.68), and that mean is significantly lower than all the other means. Tests 6, 4, and 3 have significantly higher means that test 2, and tests 4, 3, and 5 have significantly lower means than test 1, but these two groups overlap somewhat.

Since we are not sure that the data are normally distributed, we also did a Mann Whitney U test to compare test 1 (std) to each of the other tests with respect to mean bops. The results were:

Comparison	Mann-Whitney Z statistic	p-value
Std vs #2	-6.063	P< 0.0005
Std vs #3	-3.764	P< 0.0005
Std vs #4	-5.268	P< 0.0005
Std vs #5	-3.580	P< 0.0005
Std vs #6	-5.151	P< 0.0005

Therefore, we conclude that std is highly significantly different from the other "tests". As can be seen from the previous table, std has a significantly higher mean bops than an of the other "tests".

17 August, 2012

James Noble:

SPECjvm2008 data.

For each benchmark, for each "test" (6 different tests, standard, OasD, mem, memdyn, noproxy, noprodyn), the time taken was recorded for 20 runs. We wanted to test whether the mean time taken differed between standard and each of the other tests, while controlling for the different times needed for different benchmarks. The actual means ± standard deviations were:

Descriptive Statistics

Dependent Variable:ops / m

Benchmarks	1=std 2=OasD 3=mem 4=mem dyn 5=no proxy 6=no pro dyn	Mean	Std. Deviation	N
composite	1	28.8960	.24977	20
	2	28.8215	.23723	20
	3	28.9465	.18624	20
	4	28.9565	.21137	20
	5	28.8520	.23869	20
	6	28.9355	.22795	20
	Total	28.9013	.22704	120
compress	1	46.5630	.81382	20
	2	46.7700	.58969	20
	3	46.7125	.58660	20
	4	46.8180	.41703	20
	5	46.8295	.42384	20
	6	46.8425	.56901	20
	Total	46.7559	.57773	120
crypto.aes	1	18.4875	.18293	20
	2	18.4005	.12668	20
	3	18.4580	.10309	20
	4	18.4790	.18373	20
	5	18.4180	.10108	20
	6	18.4780	.19061	20
	Total	18.4535	.15335	120
crypto.rsa	1	35.0405	.25287	20
	2	34.8915	.28470	20
	3	34.9205	.30734	20
	4	34.9470	.35726	20

	5	34.9350	.25353	20
	6	34.9495	.27113	20
	Total	34.9473	.28765	120
crypto.signverify	1	53.0635	.27031	20
	2	52.9720	.31095	20
	3	53.1110	.29483	20
	4	53.1590	.36172	20
	5	53.0855	.37848	20
	6	53.1160	.29204	20
	Total	53.0845	.31903	120
derby	1	21.5525	.47931	20
	2	21.6325	.39820	20
	3	21.5610	.49668	20
	4	21.7260	.42820	20
	5	21.5930	.37649	20
	6	21.6060	.32716	20
	Total	21.6118	.41682	120
mpegaudio	1	15.0840	.05394	20
	2	15.1010	.05360	20
	3	15.0845	.05969	20
	4	15.1065	.05244	20
	5	15.0985	.06243	20
	6	15.0740	.05734	20
	Total	15.0914	.05665	120
scimark.fft.large	1	23.6140	.27118	20
	2	23.5255	.30384	20
	3	23.4870	.26956	20
	4	23.4270	.27995	20
	5	23.5515	.24195	20
	6	23.5095	.37904	20
	Total	23.5191	.29359	120
scimark.fft.small	1	82.7520	4.49351	20
	2	84.2425	5.20987	20
	3	84.9800	3.93504	20
	4	83.3460	4.00923	20
	5	84.4565	4.70209	20
	6	83.5695	4.26497	20
	Total	83.8911	4.42605	120
scimark.lu.large	1	6.9805	1.43579	20
	2	6.6895	1.23541	20

	3	6.7240	1.22455	20
	4	7.0515	1.38741	20
	5	6.3365	.95588	20
	6	7.0200	1.41898	20
	Total	6.8003	1.28465	120
scimark.lu.small	1	107.9805	.87217	20
	2	107.6050	.91727	20
	3	107.4815	.75228	20
	4	107.5785	.84388	20
	5	107.9375	.69587	20
	6	107.8180	.85447	20
	Total	107.7335	.83047	120
scimark.monte_carlo	1	15.3275	1.48834	20
	2	15.3345	1.48320	20
	3	15.6495	.09556	20
	4	15.6725	.11130	20
	5	15.6310	.03059	20
	6	15.2920	1.47193	20
	Total	15.4845	1.04053	120
scimark.sor.large	1	13.0100	.01522	20
	2	13.0095	.01791	20
	3	13.0135	.01424	20
	4	13.0190	.01651	20
	5	12.9920	.05207	20
	6	13.0130	.01490	20
	Total	13.0095	.02653	120
scimark.sor.small	1	57.4675	.10770	20
	2	57.4665	.10287	20
	3	57.4855	.10385	20
	4	57.4650	.12275	20
	5	57.4340	.07177	20
	6	57.4405	.09665	20
	Total	57.4598	.10144	120
scimark.sparse.large	1	11.5115	.22951	20
	2	11.9650	1.21889	20
	3	11.7015	.28268	20
	4	11.8150	.80475	20
	5	11.7080	.34080	20
	6	11.7715	.35921	20
	Total	11.7454	.64785	120

scimark.sparse.small	1	43.8685	.10835	20
	2	43.7890	.12363	20
	3	43.7950	.18280	20
	4	43.8275	.11097	20
	5	43.8450	.18205	20
	6	43.8045	.14684	20
	Total	43.8216	.14548	120
serial	1	33.2225	.97814	20
	2	32.4590	.95790	20
	3	33.2670	.80898	20
	4	33.0365	.91651	20
	5	32.8775	1.30508	20
	6	33.3110	.89162	20
	Total	33.0289	1.01212	120
sunflow	1	20.5120	.50088	20
	2	20.4835	.49570	20
	3	20.5525	.31598	20
	4	20.4190	.52098	20
	5	20.2170	.63115	20
	6	20.5085	.47015	20
	Total	20.4487	.49992	120
xml.validation	1	63.3805	1.09099	20
	2	63.0110	.95878	20
	3	63.3640	1.02952	20
	4	63.1920	1.41898	20
	5	63.7475	1.39470	20
	6	63.2260	1.02589	20
	Total	63.3202	1.16537	120
Total	1	36.7534	26.05233	380
	2	36.7458	26.13193	380
	3	36.8576	26.18187	380
	4	36.7917	26.01006	380
	5	36.8182	26.26357	380
	6	36.8045	26.08697	380
	Total	36.7952	26.09261	2280

Once again, we had a problem in that the data did not meet the traditional assumptions of normal distribution and equality of variances between groups. We therefore used the Mann-Whitney test to compare "tests" (i.e., standard vs each of the others) within each benchmark. This test essentially does a two-sample t test on the ranks of the data instead of the data itself. The results are summarized in the following table:

Benchmark	**"test"**	**ANOVA, F**	**Mann-Whitney U**	
composite	overall	F = 1.180, p = 0.324		
	2 vs std	T = -0.075, p = 0.300	Z = -0.650, p = 0.516	
	3 vs std	T = 0.050, p = 0.482	Z = -0.771, p = 0.441	
	4 vs std	T = 0.061, p = 0.399	Z = -0.812, p = 0.417	
	5 vs std	T = -0.044, p = 0.540	Z = -0.406, p = 0.685	
	6 vs std	T = 0.039, p = 0.582	Z = -0.717, p = 0.473	
compress	overall	F = 0.662, p = 0.653		
	2 vs std	T = 0.207, p = 0.263	Z = -0.717, p = 0.473	
	3 vs std	T = 0.150, p = 0.418	Z = -0.108, p = 0.914	
	4 vs std	T = 0.255, p = 0.169	Z = -0.555, p = 0.579	
	5 vs std	T = 0.267, p = 0.150	Z = -0.920, p = 0.357	
	6 vs std	T = 0.280, p = 0.132	Z = -1.231, p = 0.218	
Crypto.aes	overall	F = 1.110, p = 0.359		
	2 vs std	T = -0.087, p = 0.075	Z = -1.450, p = 0.147	
	3 vs std	T = -0.029, p = 0.543	Z = -0.163, p = 0.871	
	4 vs std	T = -0.009, p = 0.861	Z = -0.041, p = 0.968	
	5 vs std	T = -0.069, p = 0.048	Z = -1.097, p = 0.273	
	6 vs std	T = -0.009, p = 0.845	Z = -0.122, p = 0.903	
Crypto.rsa	overall	F = 0.602, p = 0.698		
	2 vs std	T = -0.149, p = 0.107	Z = -1.664, p = 0.096	
	3 vs std	T = -0.120, p = 0.193	Z = -1.055, p = 0.291	
	4 vs std	T = -0.094, p = 0.310	Z = -0.988 p = 0.323	
	5 vs std	T = -0.105, p = 0.253	Z = -1.448, p = 0.148	
	6 vs std	T = -0.091, p = 0.323	Z = -1.083, p = 0.279	
Crypto.signverify	overall	F = 0.793, p = 0.557		
	2 vs std	T = -0.091, p = 0.368	Z = -0.839, p = 0.402	
	3 vs std	T = 0.048, p = 0.640	Z = -0.812, p = 0.417	
	4 vs std	T = 0.096, p = 0.348	Z = -0.974, p = 0.330	
	5 vs std	T = 0.022, p = 0.829	Z = -0.108, p = 0.914	
	6 vs std	T = 0.052, p = 0.605	Z = -0.433, p = 0.665	
derby	overall	F = 0.449, p = 0.813		
	2 vs std	T = 0.080, p = 0.550	Z = -0.798, p = 0.429	
	3 vs std	T = 0.008, p = 0.949	Z = -0.271, p = 0.787	
	4 vs std	T = 0.173, p = 0.196	Z = -1.434, p = 0.152	
	5 vs std	T = 0.040, p = 0.762	Z = -0.460, p = 0.646	
	6 vs std	T = 0.054, p = 0.689	Z = -0.555, p = 0.579	
mpegaudio	overall	F = 0.966, p = 0.442		
	2 vs std	T = 0.017, p = 0.345	Z = -1.044, p = 0.296	
	3 vs std	T =0.001, p = 0.978	Z = -0.122, p = 0.903	
	4 vs std	T = 0.022, p = 0.212	Z = -1.234, p = 0.217	
	5 vs std	T = 0.014, p = 0.420	Z = -0.759, p = 0.448	
	6 vs std	T = -0.010, p = 0.578	Z = -0.911, p = 0.362	
Scimark.fft.large	overall	F = 0.911, p = 0.477		
	2 vs std	T = -0.088, p = 0.343	Z = -1.001, p = 0.317	
	3 vs std	T = -0.127, p = 0.175	Z = -1.569, p = 0.117	

	4 vs std	T = -0.187, p = 0.047	Z = -2.097, p = 0.036	
	5 vs std	T = -0.062, p = 0.503	Z = -0.622, p = 0.534	
	6 vs std	T = -0.105, p = 0.264	Z = -0.920, p = 0.358	
Scimark.fft.small	overall	F = 0.670, p = 0.647		
	2 vs std	T = 1.490, p = 0.293	Z = -1.109 p = 0.267	
	3 vs std	T = 2.228, p = 0.117	Z = -1.718 p = 0.086	
	4 vs std	T = 0.594, p = 0.674	Z = -0.568, p = 0.570	
	5 vs std	T = 1.704, p = 0.229	Z = -0.947, p = 0.344	
	6 vs std	T = 0.817, p = 0.563	Z = -0.730, p = 0.465	
Scimark.lu.large	overall	F = 0.910, p = 0.477		
	2 vs std	T = -0.291, p = 0.476	Z = -0.677, p = 0.499	
	3 vs std	T = -0.256, p = 0.530	Z = -0.311, p = 0.756	
	4 vs std	T = 0.071, p = 0.862	Z = -0.257, p = 0.797	
	5 vs std	T = -0.644, p = 0.116	Z = -1.597, p = 0.110	
	6 vs std	T = 0.040, p = 0.923	Z = -0.421, p = 0.674	
Scimark.lu.small	overall	F = 1.253, p = 0.289		
	2 vs std	T = -0.376, p = 0.153	Z = -1.109, p = 0.267	
	3 vs std	T = -0.499, p = 0.059	Z = -1.935, p = 0.053	
	4 vs std	T = -0.402, p = 0.127	Z = -1.299, p = 0.194	
	5 vs std	T = -0.043, p = 0.870	Z = -0.460, p = 0.646	
	6 vs std	T = -0.162, p = 0.535	Z = -0.676, p = 0.499	
Scimark.monte_carlo	overall	F = 0.611, p = 0.691		
	2 vs std	T = 0.007, p = 0.983	Z = -0.095, p = 0.924	
	3 vs std	T = 0.322, p = 0.334	Z = -0.354 p = 0.723	
	4 vs std	T = 0.345, p = 0.301	Z = -1.404, p = 0.160	
	5 vs std	T = 0.304, p = 0.362	Z = -0.476, p = 0.634	
	6 vs std	T = -0.035, p = 0.915	Z = -1.413, p = 0.158	
Scimark.sor.large	overall	F = 2.575, p = 0.030		
	2 vs std	T = 0.000, p = 0.951	Z = -0.138, p = 0.891	
	3 vs std	T = 0.003, p = 0.667	Z = -0.625, p = 0.532	
	4 vs std	T = 0.009, p = 0.270	Z = -2.065, p = 0.039	
	5 vs std	T = -0.018, p = 0.029	Z = -1.720, p = 0.085	
	6 vs std	T = 0.003, p = 0.713	Z = -0.636, p = 0.525	
Scimark.sor.small	overall	F = 0.702, p = 0.623		
	2 vs std	T = -0.001, p = 0.975	Z = -0.068, p = 0.946	
	3 vs std	T = 0.018, p = 0.578	Z = -0.636, p = 0.525	
	4 vs std	T = -0.002, p = 0.938	Z = -0.108, p = 0.914	
	5 vs std	T = -0.033, p = 0.302	Z = -0.853, p = 0.394	
	6 vs std	T = -0.027, p = 0.405	Z = -0.474, p = 0.635	
Scimark.sparse.large	overall	F = 1.068, p = 0.382		
	2 vs std	T = 0.453, p = 0.029	Z = -2.301, p = 0.021	
	3 vs std	T = 0.190, p = 0.355	Z = -2.570, p = 0.010	
	4 vs std	T = 0.303, p = 0.141	Z = -1.773 p = 0.076	
	5 vs std	T = 0.197, p = 0.339	Z = -1.908, p = 0.056	
	6 vs std	T = 0.260, p = 0.206	Z = -2.408, p = 0.016	
Scimark.sparse.small	overall	F = 0.912, p = 0.476		
	2 vs std	T = -0.079, p = 0.087	Z = -2.194, p = 0.028	
	3 vs std	T = -0.073, p = 0.114	Z = -1.245, p = 0.213	
	4 vs std	T = -0.041, p = 0.376	Z = -1.286, p = 0.198	
	5 vs std	T = -0.023, p = 0.611	Z = -0.054, p = 0.957	
	6 vs std	T = -0.064, p = 0.168	Z = -1.624, p = 0.104	
serial	overall	F = 2.133, p = 0.066		
	2 vs std	T = -0.763, p = 0.016	Z = -2.083, p = 0.037	

	3 vs std	T = 0.044, p = 0.887	Z = -0.365, p = 0.715	
	4 vs std	T = -0.186, p = 0.553	Z = -0.555, p = 0.579	
	5 vs std	T = -0.345 p = 0.272	Z = -0.785, p = 0.433	
	6 vs std	T = 0.088, p = 0.778	Z = -0.108, p = 0.914	
sunflow	overall	F = 1.196 p = 0.315		
	2 vs std	T = -0.028, p = 0.857	Z = -0.474, p = 0.636	
	3 vs std	T = 0.041, p = 0.797	Z = -0.487, p = 0.626	
	4 vs std	T = -0.093, p = 0.556	Z = -0.839, p = 0.402	
	5 vs std	T = -0.295, p = 0.064	Z = -1.488, p = 0.137	
	6 vs std	T = -0.003, p = 0.982	Z = -0.419, p = 0.675	
Xml.validation	overall	F = 0.907, p = 0.479		
	2 vs std	T = -0.370, p = 0.319	Z = -1.461, p = 0.144	
	3 vs std	T = -0.017, p = 0.964	Z = -0.406, p = 0.685	
	4 vs std	T = -0.188, p = 0.611	Z = -0.406, p = 0.685	
	5 vs std	T = 0.367, p = 0.369	Z = -1.055, p = 0.291	
	6 vs std	T = -0.155, p = 0.676	Z = -0.730, p = 0.465	

So, in this table, for each benchmark, we have the results of two analyses.

The first one, Analysis of Variance (ANOVA), assumes the data are normally distributed. It tests, first, whether the mean bops differs at all between any of the different "tests" (std, etc). that is the F statistic, with its p-value. Next, I did a contrast between std and each of the other "tests". For each of these comparisons, we have a t statistic and its p-value. So, for xml.validation, the overall F test shows no difference between means ($p = 0.479 > 0.05$), and none of the pairwise comparisons with std are significant.

The last column gives the results of the non-parametric tests (not assuming normality). We did the Mann-Whitney U test, which is also called the Wilcoxon rank sum test. Here we don't have an overall comparison (like the F test), we just have the 5 pairwise comparisons, each with a Z statistic and its p-value.

There're a lot of p-values here. The best thing is to look for patterns. Where there are differences are they more likely to occur with different benchmarks, or with different "tests".