Dynamic Extent Objects


CHRISTIAN QUEINNEC
UUCP: ...!mcvax!inria!ltp!queinnec
INTERNET: queinnec@ltp.unip67.fr
Greco de Programmation

Abstract

Lisp objects are heap- rather than stack-allocated because their extent is generally indefinite. Since stack-deallocation is performed without running the garbage collector, speed improvements are expected. Dedicated hardware can stack-allocate objects associated with reference counters or microcode daemons such that one can exactly know the status of the object to be deallocated and perform whatever appropriate treatment (usually a copy in the heap) according to its reachability. However such a solution is not efficient on stock hardware.

This paper presents and analyzes a new technique to efficiently solve this problem. It creates first class dynamic extent objects (DEO) that are stack-allocated, permits to access them only while they are in the stack and prevent to follow up dangling pointers. Every usual indefinite extent object has its counterpart as a DEO associated with the same set of operators. Finally DEO do not require dedicated hardware and may be properly compiled. The technique can be used within other programming languages. Some performance figures are discussed at the end of the paper.

1 Introduction

Lisp world is two-fold: there is a stack that contains continuations and pieces of environments, and a heap where are allocated almost all objects (cons-cells, strings, arrays, code ...). Since objects extent is usually indefinite and cannot be simply bounded, they are heap-allocated rather than stack-allocated; deallocation is then left to the Garbage Collector. A typical let form such as

(let ((temp (foo)))
  (bar temp temp) )

seems to introduce a stack discipline but that depends on the meaning of bar. For example:

(define (foo) ; computes the 10^{10^th} decimal of \pi
  ...
)
(define (bar n p)
  (print n)(print p)
  t
)
(let ((temp (foo)))
  (bar temp temp) )

is stack obedient while

(define (foo)
  (list 1 2 3 4) )
(define (bar n p)
  (nconc n p) )
(let ((temp (foo))
     (bar temp temp))

is definitely not. In the first case, after exiting the let form, the value of temp can be scavenged since nothing can point on it: temp can then be stack-allocated. In the second case since the value of temp is part of the value of the let form, its extent cannot be known from the sole inspection of the let body nor from the body of bar, it must probably be heap-allocated. One cannot stack-allocate it since subsequent growth of the stack will overwrite the value. Moreover the recently freed part of the stack cannot be guaranteed to remain unchanged since garbage collector or interrupts often use the complement of the active part of the stack.

However stack-allocation is efficient and widely used in classical programming languages as C or Pascal where the prevention of dangling pointers is left to the user: a fact Lisp cannot agree with! Nevertheless some compilers can stack-allocate objects which can be proved to obey a strict stack discipline. Among others ORBIT [3] can stack-allocate contexts of some closures (typically lambda in map forms) while Symbolics Common Lisp [11] offers stack-lists and stack-allocated variant of &rest [9]. Moreover dynamic extent entities exist in Lisp. These are special bindings from identifiers to values or from labels to continuations. One can hope to extend this behaviour to more tangible objects.

Compared to incremental or concurrent GC, using dynamic extent object is rather cheap and provides a valuable alternative. Moreover stack discipline is approximatively respected except that scope and size of resources needed for a computation are difficult to foresee.

Our aim is to describe an efficient way to access, stack-allocate and stack-deallocate first class dynamic extent objects providing a good compilability, avoiding birth of dangling pointers and, substantially better, detecting any attempt to follow a reference to obsolete object. DEO must support the same set of operations that their indefinite extent counterpart support and access to DEO must not be restricted to a lexical scope which restrict their use.

To present our proposition, we first define a denotational semantics for a tiny Lisp Kernel offering a new special form, similar to let, that allocates a cons-cell in the stack, binds it to a local name and evaluates the body in that new environment. The cons-cell is deallocated at the end of the body without any assumptions to the final value of this body. This scheme is not very interesting with cons-cells but can easily be extended to vectors, structures or bounded strings . . . The second part presents a raw implementation of this new special form that will be enhanced in a third part. In the fourth part will be discussed some performances.

2 The Lisp Kernel

To simply express the technique we will use a small Lisp Kernel in the spirit of Scheme [8]. It will be defined by a denotational semantics [10]. The equations will use an environment ρ mapping identifiers to locations, a store σ mapping locations to values and a continuation κ which encodes the rest of computations. We do not give full equations and restrict ourselves to what is really necessary. The notation is summarised below

\[ \mathcal{E}(\pi) \quad \text{The meaning of the form } \pi \]
\[ \rho[x \leftarrow i] \quad \text{Substitution } \rho \text{ with } x \text{ for } i \]
\[ \langle c_1, c_2 \rangle \quad \text{Binary sequence formation} \]

The function \( \mathcal{E} \) is not explicitly defined. \( \mathcal{E}[\pi] \) invokes after the syntactic category of \( \pi \) the correct semantical function explained below. Thus \( \mathcal{E}[v] \) invokes identifier while \( \mathcal{E}[\lambda(v)\pi] \) calls abstraction.

The primitive domains are

\[ v \in \text{Id} \]
\[ \pi \in \text{Form} \]
\[ \text{Cons} = \text{Loc} \times \text{Loc} \]
\[ \epsilon \in \text{Val} = \text{Cons} + \text{Symbol} \]
\[ \rho \in \text{Env} = \text{Id} \rightarrow \text{Loc} \]
\(\sigma \in \text{Store} = \text{Loc} \rightarrow \text{Val}\)

\(\kappa \in \text{Cont} = \text{Val} \times \text{Store} \rightarrow \text{Val}\)

\(\phi \in \text{Func} = \text{Val} \times \text{Store} \times \text{Cont} \rightarrow \text{Val}\)

The current environment \(\rho\) converts an identifier \(v\) into a location which, in turn, given to a store \(\sigma\), is converted into a value which is returned to the continuation \(\kappa\) along with the unmodified store.

\[\text{identifier}(v) = \lambda \sigma \rho \kappa. \quad \kappa(\sigma(\rho(v)), \sigma)\]

To simplify, a function is restricted to a single variable \(v\) and a single body \(\pi\). This Lisp kernel is a lexical Lisp, \texttt{lambda} creates a closure containing the definition environment \(\rho\). \texttt{newlocation} is an implementation-dependent function which given a store finds an unused location \(\alpha\). In this equation, \(\sigma_1\) is the definition store while \(\sigma_1\) is an application store.

\[\text{abstraction}(v, \pi) = \lambda \sigma \rho \kappa. \quad \kappa(\lambda \sigma_1 \kappa_1. \quad \mathcal{E}(\pi)\quad (\sigma_1[\alpha \leftarrow \epsilon], \rho[v \leftarrow \alpha], \kappa_1) \quad \text{where} \quad \alpha = \text{newlocation}(\sigma_1), \sigma)\]

A form is evaluated by applying the value of its first term \(\pi\) to the value of its second term \(\pi_1\). The value of the first term \(\phi\) is assumed to be a function.

\[\text{combination}(\pi, \pi_1) = \lambda \sigma \rho \kappa. \quad \mathcal{E}(\pi)\quad (\sigma, \rho, \lambda \phi \sigma_1. \quad \mathcal{E}(\pi_1)\quad (\sigma_1, \rho, \lambda \epsilon \sigma_2. \quad \phi(\epsilon, \sigma_2, \kappa)))\]

Let us now give the definition of the predefined function \texttt{cons} which takes two parameters \(\pi\) and \(\pi_1\).

\[\text{subr-cons}(\pi, \pi_1) = \lambda \sigma \rho \kappa. \quad \mathcal{E}(\pi)\quad (\sigma, \rho, \lambda \epsilon \sigma_1. \quad \mathcal{E}(\pi_1)\quad (\sigma_1, \rho, \lambda \epsilon_1 \sigma_2. \quad \kappa(<\epsilon, \epsilon_1, \sigma_2)))\]

We can now deduce (by \(\beta\)-reductions) the meaning of

\[(\text{let} (v (\text{cons } \pi \pi_1)) \pi_2)\]

\[\lambda \sigma \rho \kappa. \quad \mathcal{E}(\pi)(\sigma, \rho, \lambda \epsilon \sigma_1. \quad \mathcal{E}(\pi_1)(\sigma_1, \rho, \lambda \epsilon_1 \sigma_2. \quad \mathcal{E}(\pi_2)(\sigma_2[\alpha \leftarrow <\epsilon, \epsilon_1 >], \rho[v \leftarrow \alpha], \kappa) \quad \text{where} \quad \alpha = \text{newlocation}(\sigma_2))\]

The new special form is called \texttt{dynamic-extent-cons-let} (we shall abbreviate it to \texttt{Dconslet}). We restrict ourselves to only one variable, one cons-cell allocation and a one-form body. Its syntax is

\[(\text{Dconslet } (v \pi \pi_1) \pi_2)\]

and its formal definition is
\[ f_{\text{sub-dconslet}}(v, \pi, \pi_1, \pi_2) = \]
\[ \lambda \rho \kappa. \]
\[ \mathcal{E}(\pi) \]
\[ (\sigma, \rho, \lambda \epsilon_1 \sigma_1). \]
\[ \mathcal{E}(\pi_1) \]
\[ (\sigma_1, \rho, \lambda \epsilon_1 \sigma_2). \]
\[ \mathcal{E}(\pi_2) \]
\[ (\sigma_2[\alpha < \epsilon, \epsilon_1 >], \rho[v \leftarrow \alpha], \lambda \epsilon_2 \sigma_3). \]
\[ \kappa(\epsilon_2, \sigma_3[\alpha \leftarrow \text{obsolete-dcons-object}]) \]

where \( \alpha = \text{newlocation}(\sigma_2) \)

\text{obsolete-dcons-object} is the bottom element of domain \( \text{Cons} \) that is to say \( \bot_{\text{Cons}} \). The difference between the two previous forms is the final store. The equation for \text{dconslet} suggests a stack-allocation of the freshly created cons-cell associated to a deallocation when exiting the body of the \text{dconslet}. Although written for a dynamic extent cons-cell, the latest equation can also deal with any other DEO such as vectors, arrays, structures or bounded strings.

3 Related Implementations

The first implementation of dynamic extent objects that come to mind is to allocate them in the heap combined with a flag which indicates their validity. Deallocation is simply achieved by changing the flag to mean “obsolete”. The GC will be responsible for scavenging them.

Since garbage collection of varisized objects is painful, one can choose to stack-allocate the object but to leave a handle with a flag in the heap. Something like

The GC is still responsible for recycling the handles but since they have a fixed size they can be grouped in a common area and recycled there.

4 Feasibility

We present a first raw implementation to explain the theory of operation. We shall tune it in the next section.

If we roughly allocate an DEO in the stack, the problem is to construct an indefinite extent reference to it with the following properties:

- until being deallocated, the reference to the object must lead to its stack-address,
• after being deallocated and whatever the stack can be (shrunk or grown) the reference must forbid any
access to the fields of the former object.

The reference cannot only be the stack-address since it would violates the second requirement. The
reference must provide a way to check if the stack still contains the object. We then propose to uniquely
number the DEO. We then record the association between these numbers and stack-addresses in a hash-array
indexed by these numbers. Access to a DEO is done through

1. access to the hash-array, with the number of the wanted DEO,
2. check if the DEO is still in the stack and get its stack-address.
3. extract from the stack the desired field of the DEO

It is the responsibility of the deallocation process to remove the corresponding entry from the hash-array
thus forbidding, given a reference, to obtain the associated stack-address.

If we used tagged pointers and dynamic extent cons-cells then the previous technique will look like

![Diagram of DCONS and DH array](image)

The deallocation process just removes the entry thus making the reference now erroneous. car, cdr, 
\texttt{rplaca} and \texttt{rplacd} will then raise an error. The hash-array looks like a set of handles from references to
objects, maintained by allocation and deallocation processes. More precisely we can name DH this hash-
array and \texttt{Dcons}\footnote{We chose to prefix usual cons-cells operators by a \texttt{D} but these specialized methods can be added to preexisting generic functions such as \texttt{first}, \texttt{rest}...} (for allocation), \texttt{Dcar}, \texttt{Dcdr}, \texttt{Dplaca}, \texttt{Dplacd}, \texttt{Dcons} (to check if the object is still accessible), \texttt{Duncons} (for deallocation) the operators to deal with dynamic extent cons-cells. Except \texttt{Dcons}
and \texttt{Duncons}, all of them can be directly invoked by the user. \texttt{Dcons} and \texttt{Duncons} are only part of the \texttt{opaque}
definition of the \texttt{Dconslet macro}, but are simpler to be described as functions.

\begin{verbatim}
(defun dcons (car-val cdr-val)
  (let* ((p (new-number))
         (alpha (push-frame 'Dcons-frame
                          p car-val cdr-val )))
      (setf (gethash p DH) (get-DynExt-address alpha))
      (make-reference 'Dcons p )))
\end{verbatim}

\texttt{new-number} returns a new number for a new dynamic extent object. It can be implemented by a global
counter regularly incremented as
(let ((DynExt-counter -1))
   (define (new-number)
     (setq DynExt-counter (1+ DynExt-counter))
     DynExt-counter ))

(push-frame type . arguments-of-the-frame) pushes a new frame onto the stack and returns the stack-address of the frame. To this address get-DynExt-address adds an offset to get the direct stack address of the dynamic extent cons-cell.

make-reference builds a pointer onto the $p$th DEO which type is Dcons. In a tagged architecture, the reference is compound of a tag (Dcons) and an information (p, the dynamic extent cons-cell number).

Dcar and others are are all built on the same model. We assume references to be regular Dcons references.

(defun Dcar (reference)
  (let* ((p (extract-DynExt-number reference))
         (alpha (gethash p DH))
         (if alpha (take-car alpha)
               (error "Obsolete Object") ) ) )

extract-Dynext-number is a reciprocal to make-reference. It extracts the number of the dynamic extent object from a valid reference built by make-reference. Dcar checks if the object is still in the stack (alpha is not nil) and performs a car on the obtained stack-address (by take-car). Note that errors are only raised when one tries to access part of a dynamic extent object. It is therefore legal and safe, but not very useful (and contrary to dynamic extent spirit) to return an obsolete reference outside the Dconslet scope as can be seen in

(Dconslet (L 1 2) (equal L (Dconslet (L 1 2) L)))

The result is obviously false since one compares two different references, one of which being obsolete.

We slightly lie in the opaque macro definition for Dconslet where, given a reference, the deallocation is explicitly done by the user. A real implementation deallocates the DEO while unwinding the stack, that is to say invokes Duncons on a frame-address and not on a reference. This latter version follows. Duncons deallocates the frame and cleans up DH.

(defun Duncons (frame-address)
  (let ((p (extract-DynExt-number-from-frame frame-address)))
    (remhash p DH)
    (pop-frame frame-address)
    t ) )

Given a Dcons-frame address, extract-Dynext-number-from-frame extracts the Dcons-number from the frame.

If we want to be more precise, we must exhibit the way such references behave under garbage collection. Marking is simple since we have only to follow the references only when permissible.

(defun mark (reference)
  (set-mark reference)
  (let* ((p (extract-DynExt-number reference))
         (alpha (gethash p DH))
         (if alpha (mark alpha)
               nil ) ) )

and naturally we can collect every unmarked references. Remark that it is not interesting to collect unmarked but valid dynamic extent objects ! They will be automatically deallocated when the stack is unwound.

The cost of this raw implementation is high since Dcons and Duncons mainly involves updates of DH, while Dcar and others require one access through DH. The next section will lessen it.
5 Enhancements

We discuss in this section some variations in order to lessen the price of access to DEO. The main cost is due to the DH hash-array which involves index computation and collision handling. The first cost can be reduced if we arrange DH to be an array with some power of two entries. Therefore the hash function can be as simple as a logical AND: a simple instruction on mainly all computers. The collision handling can be avoided if we manage to avoid any collisions! This is possible since new-number can choose a good number which corresponds to a free entry in DH. For example, with the previous scheme, any number greater than DynExt-counter is possible and new-number can choose the first one which leads to the next free entry of DH.

The access now looks like

\[ \text{DH} \text{ is no more an hash-array but a set of two vectors DHp and DHalpha. DHp records the dynamic extent objects numbers while DHalpha contains the associated stack-addresses. Dcar and similarly Dcdr, Drplaca and Drplacd become}\]

\[
\begin{align*}
\text{(defun Dcar (reference)} \\
\text{ (let* (,(p (extract-DynExt-number reference)))} \\
\text{ (dh *current-DH*)} \\
\text{ (hp (mod p [DHsize dh]))} \\
\text{ (if (= (aref [DHp dh] hp) p) } \\
\text{ (take-car (aref [DHalpha dh] hp))} \\
\text{ (DError "Obsolete Object"))) ) ) )
\end{align*}
\]

An access is then compound of a mask operation (mod), an indexed load (aref), a test (=) and another indexed load followed by the standard car operation. As we said, this cost is not so high and is only a maximal cost which can be reduced by compilers in good situations. However we can estimate the maximum cost of Dcar to approximately five to seven times the cost of a car on an average stack hardware computer.

Two problems can arise:

- DH is exhausted,
- new-number fails to find a good number.

The size of DH (DHsize) is the total number of permitted DEO that can simultaneously reside in the stack. DHsize must then be related to the length of the stack. When DH is exhausted, some solutions must be taken. The cheapest way is probably to map Dcons on cons thus allocating in the heap rather than in the stack according to one of the two first ways given in section 3. The other solution is to double the size of DH and to rehash all entries.

The implementation of new-number follows directly from the previous section: a global counter DynExt-counter is maintained for the whole DH table. The problem is to minimize the growth of the counter, that is to say, to

---

\( ^2 \)For reasons that will soon appear clearly we must use the current DH table: we then write [DHsize dh] instead of accessing the global variable DHsize. In fact, DH can be modeled by a structure with at least slots DHsize, DHp and DHalpha. Structures accesses are emphasized by square brackets.
choose the lowest number greater than `DynExt-counter` and corresponding to a free entry of DH. We name DHmu a pointer corresponding to the most recent used entry in DH. Then `new-number` is

```lisp
(defun new-number ()
  (block found
    (repeat [DHashize *current-DH*]
      (setq DynExt-counter (+ DynExt-counter 1))
      (setq DHmu (mod (+ 1 DHmu) [DHashize *current-DH*]))
      (if (equal (aref DHp DHmu) DHfree-entry)
        (return-from found DynExt-counter) ) )
  (error "DH exhausted") )
```

The algorithm is at its best when DH is sparse since then `DynExt-counter` will not grow too quickly. But a new problem arises: `DynExt-counter` can overflows. This problem will be soon addressed.

6 Where to put DH

If the DEO stack allocation is to work well, everybody will use it and it will increase references to the stack and to DH. The stack will be bigger since it will contain environments, continuations but also local data. Dedicated hardware with the top of the stack in the main processor will have considerably more stack page default but this will be partially corrected by the fact they have a stack cache. On the other hand dedicated hardware do not often have a heap cache, thus DH must not reside in the heap. The best way is then to allocate DH in the bottom of the stack but we can fortunately remark that DH is itself a DEO!

Suppose DH initially in the bottom of the stack. We want to create a new DEO and we must make room for its entry in the current DH. If DH is saturated then we must use a greater DH which extent is precisely the extent of the new DEO, so

- We stack allocate a new DH of greater size (two or four times the previous size),
- and we initialize the new DH with the old one and make the new one current. The hash-function is now a new mask operation. If we double the size of DH, we then take one more bit in the reference number to get the entry number.
  - We now allocate the new DEO,
  - resume computation ... and on return
  - we deallocate the entry
- we synchronize the old DH table with the current one (`DynExt-counter` has been incremented and DHmu must respect DHmu = DynExt-counter mod DHashize),
- we deallocate the current DH table,
- and we make the old DH table again current.

The greatest possible number of DH entries is bounded by the greatest possible reference number. This number is approximatively *most-positive-fixnum* and rarely under $2^{15}$!

7 Cooperation with the GC

When `DynExt-counter` overflows (ie. can be no more a fixnum) we cannot use bignums since they are heap allocated! We are thus constrained to reuse the same set of reference numbers. But at a given time, at most DHashize reference numbers are valid since DH can contain at most DHashize DEO simultaneously. Other reference numbers obviously correspond to obsolete objects. A solution is to invoke the GC (or only the sweep phasis) to renumber all references. If a reference number p designate a living DEO we can then renumber it as (p mod DHashize) and in case of obsolescence we can just turn it into ↓ Dcons so future access will be more quickly erroneous.
8 Performance Analysis

Performance analysis is quite difficult since we have to compare prices of heap-cons operations versus stack-cons operations. The former are very dependent of the performance of the GC while the latter cannot be approximated by an implementation in Lisp. A proper implementation of Dconslet requires access to the virtual machine below the Lisp system and particularly the ability to push and pop frames onto the stack. Nevertheless the DEO mechanism presented so far does not induce runtime overhead when one does not use it.

However we can examine the behaviour of the algorithm managing the DH array. Since new-number guesses a number corresponding to a free entry in DHp, bad cases may occur incrementing DynExt-counter too quickly and thus forcing a garbage collection when DynExt-counter overflows. new-number implements a linear probing algorithm which performances are analysed in [2], page 539. First, if we want the median of ∆DynExt-counter to slowly advance, say by 1.5 (resp. 2.5) for each allocation, the load average of DH must not exceed 30% (resp 50%). Thus to double DH is better triggered by the load average rather than by complete saturation. Second, since doubling DH is not costless, the greater is the initial DH and the better are the performances. This is particularly sensible in case of a bounded stack where the maximum number of simultaneous DEO can be known.

9 Extensions

Dconslet is restricted in several ways. Dconslet can only handle one variable and allocate a single cons-cell. To deal with multiple variables is straightforward. Let just have

(Dconslet ((L1 car1 cdr1) (L2 car2 cdr2) ... )
body )

A single frame can be allocated for multiple dynamic extent objects but the local variables L1 L2 ... must be given as many different references as there are objects. The multiple variables Dconslet form is just a syntactic convenience.

Following declaration style[9], one can prefer to introduce a new declaration specifier and write

(let ((L1 car1 cdr1)
(L2 car2 cdr2) ...
)(declare (dynamic-extent L1 L2 ...))
body )

We can also extend the form to dynamic extent objects other than cons-cells, such as lists, strings, vectors (any bounded resource): for example

(Dvectorlet ((V1 term1 term2 ...) (V2 term1 term2 ...) ... )
body )

Dvectorlet is superior to Dconslet since it is not reduced to a cons-cell with just two pointers. A whole structure with numerous fields can be stack-allocated under a single reference.

10 Conclusion

We have presented an implementation technique for a new kind of first class objects with a dynamic extent which are allocated in the stack rather than in the heap. Every indefinite extent object has its counterpart as a DEO and support the same set of operators without restriction. DEO represent local and temporary resources which will be released without overhead. The proposed implementation respects the safety of references to such objects avoiding to create dangling pointers when they leave the stack. The access cost is
not so high and the offered safety can be valuable during the debugging process. Surprisingly tables needed to access DEO are themselves DEO and thus reside in the stack. The technique is operated by new special forms or declarations and offers some speed improvements since the GC is not involved.

References


