Ruby master - Bug #18141
Marshal load with proc yield objects before they are fully initialized
09/01/2021 08:17 AM - byroot (Jean Boussier)

Status: Closed
Priority: Normal
Assignee: 
Target version: 
ruby -v:
Backport: 2.6: REQUIRED, 2.7: DONE, 3.0: DONE

Description
I assume this is a bug because I can't find any spec or test for this behaviour:

Consider the following script:

```ruby
payload = Marshal.dump("foo")

Marshal.load(payload, -> (obj) {
  if obj.is_a?(String)
    p [obj, obj.encoding]
  end
  obj
})
p [:final, string, string.encoding]
```

outputs:

```
["foo", #<Encoding:ASCII-8BIT>]
[:final, "foo", #<Encoding:UTF-8>]
```

So Marshal call the proc before the string get its encoding assigned, this is because the encoding is stored alongside as a TYPE_IVAR. I think in such cases Marshal should delay calling the proc until the object is fully restored.

A corollary to this behaviour is that the following code:

```ruby
Marshal.load(payload, :freeze.to_proc)
```

raises with can't modify frozen String: “foo” (FrozenError).

The same happens with any instance variable on Array or Hash

```ruby
foo = {}
foo.instance_variable_set(:@bar, 42)
```

```ruby
object = Marshal.load(payload, ->(obj) {
  if obj.is_a?(Hash)
    p [obj, obj.instance_variable_get(:@bar)]
    obj.freeze
  end
  obj
})
```

```ruby
[{}, nil]
```

```
/tmp/marshal.rb:6:in `load': can't modify frozen Hash: {} (FrozenError)
from /tmp/marshal.rb:6:
```

Associated revisions
Revision fe9d33be - 10/09/2021 06:05 AM - nagachika (Tomoyuki Chikanaga)

merge revision(s) 89242279e61b023a81c58065c62a82de8829d0b3,529fc204a84f825f9883c34b004acbaa802615: [Backport #18141]
Marshall.load: do not call the proc until strings have their encoding

Ref: https://bugs.ruby-lang.org/issues/18141

---
marshal.c | 7 +++-
spec/ruby/core/marshal/shared/load.rb | 62 ++++++++++++++++++++++++++-
test/ruby/test_marshal.rb | 17 ++++++
3 files changed, 64 insertions(+), 22 deletions(-)

marshal.c: don't call the proc with partially initialized objects.

(#4866)

For cyclic objects, it requires to keep a st_table of the partially initialized objects.

---
marshal.c | 75 +++++++++++++++++++++---------------
spec/ruby/core/marshal/shared/load.rb | 75 +++++++++++++++++++++---------------
test/ruby/test_marshal.rb | 12 ++++
3 files changed, 97 insertions(+), 65 deletions(-)

Revision 419266d4 - 11/24/2021 10:36 AM - NAKAMURA Usaku
merge revision(s) 8924279e61b023a81c58065c62a82de8829d0b3529fc204a84f25f98f83c34b004acbaa802615: [Backport #18141]

Marshall.load: do not call the proc until strings have their encoding

---
marshal.c | 7 +++-
spec/ruby/core/marshal/shared/load.rb | 62 ++++++++++++++++++++++++++-
test/ruby/test_marshal.rb | 17 ++++++
3 files changed, 64 insertions(+), 22 deletions(-)

marshal.c: don't call the proc with partially initialized objects.

(#4866)

For cyclic objects, it requires to keep a st_table of the partially initialized objects.

---
marshal.c | 75 +++++++++++++++++++++---------------
spec/ruby/core/marshal/shared/load.rb | 75 +++++++++++++++++++++---------------
test/ruby/test_marshal.rb | 12 ++++
3 files changed, 97 insertions(+), 65 deletions(-)

History

#1 - 09/01/2021 05:57 PM - byroot (Jean Boussier)
I potentially have a fix: https://github.com/ruby/ruby/pull/4797

#2 - 09/02/2021 12:40 AM - nobu (Nobuyoshi Nakada)
Should use ruby_bug "#18141", "..."3.1" instead of ruby_version_is.

#3 - 09/14/2021 11:05 PM - nobu (Nobuyoshi Nakada)
- Backport changed from 2.6: UNKNOWN, 2.7: UNKNOWN, 3.0: UNKNOWN to 2.6: REQUIRED, 2.7: REQUIRED, 3.0: REQUIRED
- Status changed from Open to Closed

#4 - 09/17/2021 02:19 AM - byroot (Jean Boussier)
So while working on https://bugs.ruby-lang.org/issues/18148, I discovered that many other types of objects are impacted.

Just a few examples:

def round_trip(obj, proc = ->(o) { o.freeze })
  Marshal.load(Marshal.dump(obj), proc)
end

h = {}
h.instance_variable_set(:@foo, 42) #=>FrozenError: can't modify frozen Hash
round_trip(h) rescue p $!

a = []
a.instance_variable_set(:@foo, 42) #=>FrozenError: can't modify frozen Array

03/24/2022
round_trip(a) rescue p !$

Also, probably by design, but since you can replace the object by what the proc returns:

```ruby
a = {}
a.instance_variable_set(:@foo, 42)
round_trip(a, proc { 24 }) rescue p !$ #<FrozenError: can't modify frozen Integer>
```

I fixed most cases in [https://github.com/ruby/ruby/pull/4859](https://github.com/ruby/ruby/pull/4859), which is my current attempt at implementing [https://bugs.ruby-lang.org/issues/18148](https://bugs.ruby-lang.org/issues/18148), but since I just noticed this was marked for backport, I might need to split the bug fix from the new feature. No?

#5 - 09/18/2021 07:34 AM - nagachika (Tomoyuki Chikanaga)
Hello byroot,
Thank you for the investigation about the issue. Yes, the patch with only the bug fix is very helpful to maintain stable branches.

#6 - 09/18/2021 01:45 PM - byroot (Jean Boussier)
I made a followup patch: [https://github.com/ruby/ruby/pull/4866](https://github.com/ruby/ruby/pull/4866)
It now handle similar bugs with Array, Hash and other mutable objects. It also handle circular objects.

#7 - 09/28/2021 08:42 AM - byroot (Jean Boussier)
- Status changed from Closed to Open
- Description updated
- Subject changed from Marshal load with proc yield strings before they are fully initialized to Marshal load with proc yield objects before they are fully initialized

I took the liberty to re-open this issue and to rewrite it to be more generic.

I wonder if it wouldn't be simpler to revert the string only fix ([https://github.com/ruby/ruby/pull/4797](https://github.com/ruby/ruby/pull/4797)), and then to merge the more general one ([https://github.com/ruby/ruby/pull/4866](https://github.com/ruby/ruby/pull/4866)), this way it would be simpler to backport.

Any opinions?

#8 - 09/30/2021 03:33 PM - byroot (Jean Boussier)
- Status changed from Open to Closed

[https://github.com/ruby/ruby/pull/4866](https://github.com/ruby/ruby/pull/4866) was merged as 529fc204af84f825f98f83c34b004acbaa802615, closing.

#9 - 10/09/2021 06:37 AM - nagachika (Tomoyuki Chikanaga)
- Backport changed from 2.6: REQUIRED, 2.7: REQUIRED, 3.0: REQUIRED to 2.6: REQUIRED, 2.7: REQUIRED, 3.0: DONE

ruby_3_0_fe9d33beb78d5c7932a5c2ca3953045c0aa751d5 merged revision(s)
89242279e61b023a81c58065c62a82de8829d0b3,529fc204af84f825f98f83c34b004acbaa802615.

#10 - 11/24/2021 10:36 AM - usa (Usaku NAKAMURA)
- Backport changed from 2.6: REQUIRED, 2.7: REQUIRED, 3.0: DONE to 2.6: REQUIRED, 2.7: DONE, 3.0: DONE

ruby_2_7_419266d44c54c8b75f1e824f060c8b388f7a405b merged revision(s)
89242279e61b023a81c58065c62a82de8829d0b3,529fc204af84f825f98f83c34b004acbaa802615.