Ruby - Feature #5106

Is MurmurHash overkill?

07/27/2011 12:54 PM - kstephens (Kurt Stephens)

Status:	Rejected	
Priority:	Normal	
Assignee:	mame (Yusuke Endoh)	
Target version:		

Description

st.c implements MurmurHash to compute hash table indexes (#hash).

Simpler hash functions may be appropriate for hash tables, esp. small tables.

Is there a particular reason this hash function was chosen? Is MurmurHash typically used for check-summing purposes?

Anybody positively adverse to changing it? If so I won't bother. Otherwise I might take a crack at it.

-- KAS

History

#1 - 07/27/2011 04:17 PM - naruse (Yui NARUSE)

- Priority changed from Normal to 3

#2 - 03/25/2012 05:22 PM - mame (Yusuke Endoh)

- Status changed from Open to Assigned

- Assignee set to mame (Yusuke Endoh)

#3 - 03/27/2012 01:44 AM - mame (Yusuke Endoh)

- Status changed from Assigned to Feedback

Hello,

> Simpler hash functions may be appropriate for hash tables, esp. small
> tables.

Please make it quantitative when talking about performance.

Please show a benchmark first.
How slow is MurmurHash? How fast will st.c be if the algoritm
is replaced with another one?
And how much the drawback? That is, how many collision will
increase in actual cases?

> Anybody positively adverse to changing it? If so I won't bother. Otherwise
> I might take a crack at it.

Please make your move first.

We cannot answer your question until you show a benchmark.
Even after that, we cannot promise to import your patch before
you create it. The chicken or the egg :-)

Thanks,

--
Yusuke Endoh mame@tsg.ne.jp

#4 - 03/29/2012 05:22 AM - MartinBosslet (Martin Bosslet)

With the recent changes since hashDoS I'm not sure whether
we should really aim for simplicity. It's quite easy to

become a target with a function that is algebraically as
simple as for example DJB33X. But speed on the other hand
couldn't hurt. We use Murmur 2, there is a version 3 by
the same author who claims it is generally faster than 2.
Maybe we should give that a try first?

#5 - 11/20/2012 02:32 AM - mame (Yusuke Endoh)

- Status changed from Feedback to Rejected

I'm closing this ticket. Feel free to reopen with benchmark.

BTW: MurmurHash was replaced with SipHash which is a bit slower than MurmurHash.

--
Yusuke Endoh mame@tsg.ne.jp