Subject: Re: [PATCH net-2.6.26 2/6][NETNS][SOCK]: Introduce per-net inuse counters.
Posted by Eric Dumazet on Fri, 28 Mar 2008 07:36:15 GMT
View Forum Message <> Reply to Message

> Eric Dumazet wrote:
>

>>
>>> This is probably the most controversial part of the set.
>>>
>>> The counters are stored in a per-cpu array on a struct net. To
>>> index in this array the prot->inuse is declared as int and used.
>>>
>>> Numbers (indices) to protos are generated with the appropriate
>>> enum. I though about using some existing IPPROTO_XXX numbers for
>>> protocols but they were too large (IPPROTO_RAW is 255) and did
>>> not differ for ipv4 and ipv6 (there's no IP6PROTO_RAW, etc).
>>>
>>> The sock_prot_inuse_(add|get) now use the net argument to
>>> get the counter, but this all hides under CONFIG_NET_NS.
>>>
>>> The sock_prot_inuse_(init|fini) are no-ops. DEFINE_PROTO_INUSE
>>> is empty and REF_PROTO_INUSE assigns an index to a proto.
>>>
>>>
>>>
>> Given that :
>>
>> 1) pcounter should really go away from kernel, since Andrew disagree
>> with the implementation.
>>
>
> Does this and ... (below)
>
>
>> 2) the need to enumerate all protocols in your enum, it seems ... ugly :)
>>
>
> Yup :(
>
>
>> 3) alloc_percpu(struct net_prot_inuse) per net is nice because we dont
>> waste memory (if we had to use percpu_counters for each proto for example)
>>
>> I suggest to :
>>

>> 1) not use pcounter anymore
>>
>
> ... this mean that I can rework the inuse accounting in order not
> to use pcounters at all even with CONFIG_NET_NS=n? :)
>
>
Absolutely

I had to do it eventually but my paid work is currently taking me 10-12
hours per day, so please be my guest :)

reference : http://kerneltrap.org/mailarchive/linux-kernel/2008/2/16/873754

>> 2) change 'inuse' field to 'inuse_idx' or 'prot_num' that is
>> automatically allocated at proto_register time, instead statically at
>> compile time.
>>
>
> Hm... I like this approach. Will do.
>
>
>> Just provide a big enough NET_INUSE_NR (might depend on IPV6 present or
>> not, static or module) to take into account all possible protocols.
>>
>
> Well, I though about this, but wasn't sure whether such heuristics
> would be accepted.
>
>
>> struct net_prot_inuse {
>> int val[NET_INUSE_NR];
>> };
>>
>>
>>
>
>
>

Thank you