

kubernetes [#69758](#) issue [#69597](#)

commit 141b55abf5848f23848c433ff7ecd3f9ff1338cc

根据代码，nt.exhaustedZones记录了zone的集合。当zone中的所有node被删除后，zone也会被nt (NodeTree) 删除掉，但是nt.exhaustedZones没有记录该事件以至于原来的zone仍然在nt.exhaustedZones中，因此在判断exhausted的时候有len(nt.exhaustedZones) > len(nt.zones)，从而无法进入到resetExhausted。

修复的方式是去掉了不必要的nt.exhaustedZones，使用计数的方式，一旦numExhaustedZones大于了当前的nt.zones，就认为此次遍历已经耗尽。

```
for {
    if nt.zoneIndex >= len(nt.zones) {
        nt.zoneIndex = 0
    }
    zone := nt.zones[nt.zoneIndex]
    nt.zoneIndex++
-    // We do not check the set of exhausted zones before calling next() on the zone.
This ensures
+    // We do not check the exhausted zones before calling next() on the zone. This
ensures
    // that if more nodes are added to a zone after it is exhausted, we iterate over the
new nodes.
    nodeName, exhausted := nt.tree[zone].next()
    if exhausted {
-        nt.exhaustedZones.Insert(zone)
-        if len(nt.exhaustedZones) == len(nt.zones) { // all zones are exhausted. we
should reset.
+        numExhaustedZones++
+        if numExhaustedZones >= len(nt.zones) { // all zones are exhausted. we
should reset.
            nt.resetExhausted()
        }
    } else {
```