Configuring High Availability (128T Router)

This document contains the steps required for configuring support for high availability (HA) on a 128T router. Unlike traditional routers, where deploying high availability involved deploying two separate routers and using a protocol such as VRRP or HSRP to provide failover protection, the 128T deploys software instances (referred to as "nodes") in pairs, but are collectively referred to as a single, logical router.

Requirements

Configuring high availability requires that two 128T routing nodes have at least one device-interface that is shared between them (referred to in this document as a *shared interface*). Shared interfaces are configured on both nodes, but are active on only one node at a time. These shared interfaces **must** be in the same L2 broadcast domain; this is because the 128T uses "gratuitous ARP" messages to announce an interface failover, so that it may receive packets in place of its counterpart.

The two 128T router nodes that make up a high availability pair must be collocated due to latency sensitivities for the information that they synchronize between themselves.

Before You Begin

There are several things to be mindful of before configuring HA; the two nodes must be informed that they are part of a high availability set, and they must have a dedicated interface between themselves for synchronizing state information about active sessions. These steps will be covered in this section.

Clock Synchronization

Because highly available nodes synchronize time-series data, it is critical that the two nodes that comprise an HA pair have synchronized clocks. It is sufficient to manually synchronize the clocks until 128T software is installed, after which point NTP (Network Time Protocol) can be used to automatically synchronize the clocks.

Use the `timedatectl` application to set the clock and the system's time zone:

```
[root@labsystem1 ~]# chronyd -q 'pool pool.ntp.org iburst'
[root@labsystem1 ~]# hwclock -w
```

(Use the command `timedatectl list-timezones` to get the available time zone options for your region.)

Migrating from Standalone to HA

For an established standalone router of one node, converting it to be highly available requires configuring a second node within the 128T configuration (PCLI or GUI) at the outset.

> **Note**: converting an existing router from standalone to HA will require downtime, and is therefore only to be undertaken during a maintenance window, as applicable.

Adding a second node is simply a matter of configuring another *node* container within the router. Eventually, this node will contain one or more *shared interfaces*, which will protect the router from failure modes if/when interfaces or links fail. Configuring shared interfaces is covered later in this document.

Configuring a Teamed Interface

Each 128T router node must have at least one interface dedicated to synchronizing state to its counterpart. These interfaces are not configured through the 128T software configuration, but are instead configured within the host operating system; typically this is done using a Linux command line tool called `nmtui` . While it is possible (and common) to use a directly connected cable between two physically adjacent host platforms that are acting as an HA router, when connecting two systems' interfaces for state synchronization, 128 Technology recommends the use of a Linux *teamed interface*. This is to prevent a so-called "split brain" scenario when an interface or cable fails, causing the two nodes to lose communication with its mate and each become active.

Teamed Interfaces, comprised of individual *TeamPort interfaces*, represent the evolution of the Linux legacy bond interface. The key capability is that Team interfaces preserve the IP address regardless of interface carrier state. For more information, refer to the Understanding Network Teaming section of Red Hat's documentation.

In our example here, the interface selected for HA sync is named "enp0s20f3" within the host operating system. (To identify the name of your interface, use the command `ip a` in a Linux shell.)

```
[root@node1 network-scripts]# cat ifcfg-team-slave-enp0s20f3
NAME=team-slave-enp0s20f3
DEVICE=enp0s20f3
USERCTL=no
BOOTPROTO=none
ONBOOT=yes
NM_CONTROLLED=no
TYPE=TeamPort
TEAM_MASTER=team128sync
TEAM_PORT_CONFIG='{"prio": 2000}'
```

Configure the address (IPADDR), prefix length (PREFIX), and gateway (GATEWAY) on the team interface. The IP addresses, prefix, and gateways listed here are recommended for use in all 128T HA deployments; the 169.254/16 prefix is reserved for *link-local* addresses, and is ideally suited to this application. **Important**: make sure this interface is in a firewalld zone capable of permitting HA sync traffic (e.g., *trusted*).

Example for node1:

```
[root@node1 ~]# cat /etc/sysconfig/network-scripts/ifcfg-team-team128sync
NAME=team-team128sync
DEVICE=team128sync
BOOTPROTO=none
ONBOOT=yes
DEFROUTE=no
NM_CONTROLLED=no
DEVICETYPE=Team
TEAM_CONFIG='{"runner": {"name": "activebackup", "min_ports": 0}, "link_watch":
{"name": "ethtool"}}'
IPADDR=169.254.253.129
PREFIX=29
GATEWAY=169.254.253.130
ZONE=trusted
```

Example for node2:

```
[root@node2 ~]# cat /etc/sysconfig/network-scripts/ifcfg-team-team128sync
NAME=team-team128sync
DEVICE=team128sync
BOOTPROTO=none
ONBOOT=yes
DEFROUTE=no
NM_CONTROLLED=no
DEVICETYPE=Team
TEAM_CONFIG='{"runner": {"name": "activebackup", "min_ports": 0}, "link_watch":
{"name": "ethtool"}}'
IPADDR=169.254.253.130
PREFIX=29
GATEWAY=169.254.253.129
ZONE=trusted
```

Note the addresses and gateways for node1 and node2 are transposed in the sample configuration.

Configuring the Shared Interface(s)

A highly available router is comprised of exactly two routing nodes within the same *router* container. (Configuring two routers, each comprised of one node, cannot be made highly available.) Additionally, as mentioned previously, these routers must have at least one shared interface in common.

Configuring the basic properties of the two nodes is described elsewhere in this documentation. For high availability, the crucial step is identifying to the 128T the interfaces that are to be shared between them. This is done by establishing a common Layer 2 address, known as a MAC address, that is maintained by the active node in the pair. (I.e., when node1 of the pair has active control over the interface, it will respond to ARP requests for the addresses on that interface with the shared MAC address, whereas node2 will not.) The configuration element for this MAC address is the *shared-phys-address*, within the device-interface element.

The shared-phys-address is simply a series of six octets, where the only requirement is that it is unique on a given broadcast domain. (The 128T Conductor also enforces that the shared-phys-address be unique among all routers within an Authority.) There are no hardfast rules for creating "globally unique" MAC addresses; there are, however, many websites available that will generate random values. Again, since these MAC addresses are only used on a broadcast domain, they do not need to be globally unique to suit the 128T router's needs. Irrespective of how you choose to generate the value, the shared-phys-address is configured using the format "00:00:00:00:00:00."

Configuring the same shared-phys-address on two different interfaces (one per node in the high availability pair) informs the 128T that you wish to have the interfaces protect one another. This in turn causes the 128T to assign all corresponding pairs of network-interfaces that belong to this shared interface the same common *global ID*. (I.e., each network-interface on a node will have a unique global ID, but each counterpart network-interface on a highly available node will have the same global ID.) The global ID is an internal identifier, used by the 128T, to refer to the shared interface.

About the Global ID

Each network-interface within a 128T configuration has a global ID assigned to it. Much as the name suggests, these IDs are unique among all interfaces within an Authority – with one exception. When two nodes share an interface for high availability, each network-interface pair, one per node, that fails over to another network-interface on the paired node is assigned the same global ID.

```
admin@node1.router1 (device-interface[name=wan])# show
name                 wan
description          "WAN interface, port 0"
type                 ethernet
pci-address          0000:00:14.0
link-settings        auto
enabled              true
forwarding           true
shared-phys-address  00:00:5e:00:00:00

network-interface    vlan0
    name                   vlan0
    global-id              1
```

This value is also present in the output of *show rib*, where it is the trailing value within each RIB entry:

```
admin@node1.router1# show rib
Mon 2019-01-07 10:53:19 EST
Codes: K - kernel route, C - connected, S - static, R - RIP,
       O - OSPF, I - IS-IS, B - BGP, E - EIGRP, N - NHRP,
       T - Table, v - VNC, V - VNC-Direct, A - Babel, D - SHARP,
       F - PBR,
       > - selected route, * - FIB route

C>* 10.0.128.0/17 is directly connected, g2, 00:46:05
C * 169.254.127.126/31 is directly connected, g4294967293, 00:43:58
C>* 169.254.127.126/31 is directly connected, g4294967294, 00:46:06
C>* 169.254.255.2/31 is directly connected, g3, 00:44:15
C>* 192.0.2.0/24 is directly connected, g1, 00:46:05
```

The *g1* value in the line above refers to the interface assigned with global-id value 1.

Network Interface Consistency

When configuring shared interfaces, it is crucial that the network-interface elements within a shared device-interface are mirror images of one another. This is to prevent any behavioral changes when ownership of a shared interface changes from one node to its counterpart. The configuration validation step will prevent committing configuration changes when the network-interface elements are not identical.

Confirming that Interfaces are Shared

Once you've configured two device-interface elements on individual nodes within a router for high availability, the `show device-interface summary` command will identify which devices are redundant (shared) within the pair, as well as whether the interface is *active* or *standby* (or *non-redundant*, for interfaces that do not have a counterpart).

```
admin@node1.router1# show device-interface summary
Mon 2019-01-07 10:45:11 EST

================= ============== ============= ===================
===================
 Name              Admin Status   Oper Status   Redundancy Status   MAC Address
================= ============== ============= ===================
===================
 node1:wan         up             up            active
00:90:0b:54:f6:86
 node1:lan         up             up            active
00:90:0b:54:f6:87

Completed in 0.36 seconds
```

In this sample output, the interfaces on `node1` are active from a redundancy standpoint. Adding the optional argument `node all` to the command will show all interfaces on the nodes that comprise the router:

```
admin@node1.router1# show device-interface node all summary
Mon 2019-01-07 10:49:09 EST

================= ============== ============= ===================
===================
 Name              Admin Status   Oper Status   Redundancy Status   MAC Address
================= ============== ============= ===================
===================
 node1:wan         up             up            active
00:90:0b:54:f6:86
 node1:lan         up             up            active
00:90:0b:54:f6:87
 node2:wan         up             up            standby
00:90:0b:73:88:40
 node2:lan         up             up            standby
00:90:0b:73:88:41

Completed in 0.66 seconds
```

Configuring the Fabric Interface

An optional, but common inclusion in highly available routers is a fabric interface, also known as a "dogleg" interface. Named to evoke the imagery of a fabric backplane or midplane of a chassis-based router, the fabric interface is a forwarding interface between two nodes in a router, and is used when the ingress interface and egress interface for a given session are active on different nodes.

Fabric interfaces are not required for simple active/standby deployments where the two nodes are mirror images of one another (e.g., each WAN interface and LAN interface is protected using shared interfaces). It does offer an additional protection against failure even in these active/standby setups: the double failure of a LAN port on node 1 and a WAN port on node 2. For deployments where Ethernet ports are not at a premium, a fabric interface is strongly recommended.

```
device-interface  internode
    name              internode
    description       "Direct connect between nodes, port 2"
    type              ethernet
    pci-address       0000:00:14.2
```

```
    forwarding         true

    network-interface  fabric
        name         fabric
        global-id    3
        description  "Fabric link between nodes"
        type         fabric

        address      169.254.255.2
            ip-address     169.254.255.2
            prefix-length  31
        exit
    exit
exit
```

Configuring Redundancy Groups

Redundancy groups are sets of interfaces that *share fate*, such that if one of the interfaces in the group fails, mastership of all interfaces in the group will be relinquished to the counterpart node in the router. Redundancy groups are required when the two nodes in a router do not have a fabric interface between them; otherwise, you could end up in a situation where the active LAN interface is on node 1 and the active WAN interface is on node 2, with no way to transit packets from node 1 to node 2.

While redundancy groups are most commonly found in legacy deployments (i.e., those that predate 128 Technology's introduction of the fabric interface), they are still useful in simple HA deployments. Furthermore, the redundancy group affords administrators the ability to assert a preference for which node is active in an HA pair in the "sunny day" scenario where no interfaces are administratively or operationally down.

Generally, you will configure two nodes that each has a set of forwarding interfaces (for illustrative purposes, assume an interface on an internal network named *lan* and an interface on an external network named *wan*). Each node will require a *redundancy-group* that contains its pair of internal and external interfaces, as is seen in the following example:

```
redundancy-group      grp-node1
    name      grp-node1

    member    node1 wan
        node       node1
        device-id  wan
    exit

    member    node1 lan
        node       node1
        device-id  lan
    exit
    priority  50
exit

redundancy-group      grp-node2
    name      grp-node2

    member    node2 wan
        node       node2
        device-id  wan
    exit

    member    node2 lan
        node       node2
        device-id  lan
    exit
```

```
    priority  25
exit
```

In this example, our two redundant nodes (node1 and node2) each have two interfaces contained within part of the `redundancy-group` . Note that each group collects the interfaces for a node, *not interfaces that share a global-id*.

The *priority* value indicates, all things being otherwise equal, an administrative preference for which group should be active. When configuring two redundancy-groups with differing *priority* values, the failover of the systems is said to be "revertive" – that is, the group with the higher priority will be active unless it experiences a failure, but when that failure is restored it will become active again.

> **Note**: when configuring two redundancy-groups with the same *priority* value, the 128T router will select an active member using an internal election algorithm, which is not guaranteed to be revertive in the event of a failure – but is neither guaranteed to be non-revertive. For this reason, it is suggested that you configure redundancy-group elements with different *priority* values.

Confirm NTP

To confirm that you have NTP configured, use the command `show config running` as shown here:

```
admin@labsystem2.newton# show config running authority router newton system ntp

config
    authority
        router  newton
            name    newton
            system
                ntp
                    server  time.nist.gov
                        ip-address  time.nist.gov
                    exit
                exit
            exit
        exit
    exit
exit
```

To confirm that NTP is synchronized, use the `show ntp` command and confirm that at least one NTP server is in the `active` state (some columns have been removed for display purposes):

```
admin@labsystem2.newton# show ntp
Sat 2019-01-26 06:54:29 EST

Node: labsystem2

======== ================== ========= ========= ====== ======== ======== ========
 Status   Time Source         Ref. ID   Stratum   Poll    Delay   Offset   Jitter
======== ================== ========= ========= ====== ======== ======== ========
 active   *time-b-wwv.nist   .NIST.          1   1024   68.905   -0.981    2.524

Completed in 0.19 seconds
```

##Sample Configuration

Below is a sample, minimal configuration which shows the inclusion of both a fabric interface as well as redundancy-groups.

```
config

    authority
        name              128technology
        dynamic-hostname  interface-{interface-id}.{router-name}.{authority-name}

        router            router1
            name                  router1
            location-coordinates  +42.35972+116.17917/
            description           "HA branch office router, Lanner 7573B"

            system
                contact    admin@128technology.com
                log-level  info

                ntp

                    server  132.163.97.1
                        ip-address  132.163.97.1
                    exit
                exit
            exit

            node                  node1
                name              node1
                description       "Node 1 of HA pair"

                device-interface  wan
                    name                  wan
                    description           "WAN interface, port 0"
                    type                  ethernet
                    pci-address           0000:00:14.0
                    link-settings         auto
                    enabled               true
                    forwarding            true
                    shared-phys-address   00:00:5e:00:00:00

                    network-interface    vlan0
                        name                    vlan0
                        global-id               1

                        neighborhood            internet
                            name      internet
                            topology  spoke
                        exit
                        inter-router-security  internal

                        address                 192.0.2.1
                            ip-address     192.0.2.1
                            prefix-length  24
                        exit
                    exit
                exit

                device-interface  lan
                    name                  lan
                    description           "LAN interface, port 1"
                    type                  ethernet
                    pci-address           0000:00:14.1
                    link-settings         auto
                    enabled               true
                    forwarding            true
                    shared-phys-address   00:00:5e:00:00:01

                    network-interface    vlan100
```

```
                name                    vlan100
                global-id               2
                vlan                    100
                type                    external
                inter-router-security   internal

                address                 10.0.128.1
                    ip-address     10.0.128.1
                    prefix-length  17
                exit
            exit
        exit

        device-interface  internode
            name               internode
            description        "Direct connect between nodes, port 2"
            type               ethernet
            pci-address        0000:00:14.2
            forwarding         true

            network-interface  fabric
                name         fabric
                global-id    3
                description  "Fabric link between nodes"
                type         fabric

                address      169.254.255.2
                    ip-address     169.254.255.2
                    prefix-length  31
                exit
            exit
        exit
    exit

    node                  node2
        name              node2
        description       "Node 2 of the HA pair"

        device-interface  wan
            name                  wan
            description           "WAN interface, port 0"
            type                  ethernet
            pci-address           0000:00:14.0
            link-settings         auto
            enabled               true
            forwarding            true
            shared-phys-address   00:00:5e:00:00:00

            network-interface    vlan0
                name                    vlan0
                global-id               1

                neighborhood            internet
                    name      internet
                    topology  spoke
                exit
                inter-router-security   internal

                address                 192.0.2.1
                    ip-address     192.0.2.1
                    prefix-length  24
                exit
            exit
        exit
```

```
        device-interface  lan
            name                 lan
            description          "LAN interface, port 1"
            type                 ethernet
            pci-address          0000:00:14.1
            link-settings        auto
            enabled              true
            forwarding           true
            shared-phys-address  00:00:5e:00:00:01

            network-interface    vlan100
                name                   vlan100
                global-id              2
                vlan                   100
                type                   external
                inter-router-security  internal

                address                10.0.128.1
                    ip-address     10.0.128.1
                    prefix-length  17
                exit
            exit
        exit

        device-interface  internode
            name               internode
            description        "Direct connect between nodes, port 2"
            type               ethernet
            pci-address        0000:00:14.2
            forwarding         true

            network-interface  fabric
                name         fabric
                global-id    3
                description  "Fabric link between nodes"
                type         fabric

                address      169.254.255.3
                    ip-address     169.254.255.3
                    prefix-length  31
                exit
            exit
        exit
    exit

    redundancy-group      grp-node1
        name      grp-node1

        member    node1 wan
            node       node1
            device-id  wan
        exit

        member    node1 lan
            node       node1
            device-id  lan
        exit
        priority  50
    exit

    redundancy-group      grp-node2
        name      grp-node2

        member    node2 wan
            node       node2
```

```
                    device-id  wan
                exit

                member    node2 lan
                    node       node2
                    device-id  lan
                exit
                priority  25
            exit

            service-route         rte_default-route
                name          rte_default-route
                service-name  default-route

                next-hop      node1 vlan0
                    node-name  node1
                    interface  vlan0
                exit
            exit
        exit

        service           default-route
            name         default-route
            description  "Default route"
            scope        public
            address      0.0.0.0/0
        exit
    exit
exit
```