Technical Disclosure Commons

Defensive Publications Series

March 19, 2019

IN-BAND REMOTE FAILURE DETECTION

Karthik babu Haricandra Babu

Ananthakrishnan Rajamani

Frank Brockners

Lionel Florit

IN-BAND REMOTE FAILURE DETECTION

AUTHORS:
Karthik babu Harichandra Babu
Ananthakrishnan Rajamani
Frank Brockners
Lionel Florit

ABSTRACT

Techniques are described herein for an alternate option to enable delivery in the in-band Remote Failure Indication (RFI). This may provide guaranteed delivery in the data packet, and may be complementary to existing Operations, Administration, and Maintenance (OAM) mechanisms.

DETAILED DESCRIPTION

One of the Operations, Administration, and Maintenance (OAM) features as defined by Institute of Electrical and Electronics Engineers (IEEE) 802.3ah is Remote Failure Indication (RFI), which helps in detecting faults in Ethernet connectivity that are caused by slowly deteriorating quality. Ethernet OAM provides a mechanism for an OAM entity to convey these failure conditions to its peer via specific flags in the OAM Protocol Data Unit (PDU). One of the failure condition methods to communicate is Dying Gasp (DG), which indicates that an unrecoverable condition has occurred (e.g., when an interface is shut down). This type of condition is vendor specific. A notification about the condition may be sent immediately and continuously.

In one example, DG resides on a hardware component on the High-performance Wide Area Network (WAN) Interface Card (HWIC) and supports the Fast Ethernet and Gigabit Ethernet interfaces. It will be appreciated, however, that the techniques described herein need not only apply to hosts which use Network Interface Cards (NICs). For example, these techniques may apply to networking devices such as bridges, switches, routers, etc. The networking devices typically rely on a temporary back-up power supply using a capacitor that allows for a graceful shutdown and the generation of the DG message. This temporary power supply is typically designed to last from 10 to 20 milliseconds to perform these tasks.

However, this usually does not work as designed as the device is usually unable to construct and send out the packet within the miniscule time window available. This causes a delay in identifying the node that went down which can cause significant loss in traffic for service provider customers. Without the DG message, it becomes difficult for service providers to differentiate between a power failure at the customer premise and an equipment or facility failure.

The techniques described herein use any existing field in a data packet to indicate that a failure has occurred. This saves a significant amount of time as compared to constructing a new packet for indicating failure. With this, the device may be able to indicate that it has failed using the RFI with guaranteed delivery and may be extracted in the receiving peer node. The peer node may then react to any failure and take necessary actions.

One example is in-band Layer 2 (L2) OAM RFI in the data/customer packet.

Introducing RFI in-band in data packets makes the detection / monitoring / troubleshooting of critical events possible with a high level of reliability. In-band OAM RFI may be enabled for the following conditions which are critical events that could otherwise cause loss of traffic: reload command (warm reboot); boot system flash primary/secondary command (warm reboot); failure on the box (cold reboot); when the temperature of the box breaches the warning/shutdown threshold; and fan failure.

In existing DG, once the network equipment shuts down, a DG packet must be created to fill it with all the required details and send it to the remote node. The techniques described herein involve processing the in-band L2 OAM RFI function by the remote node rather than the dying node as in the 802.3ah mechanism. The remote node is thus responsible for fetching the details (e.g., Internet Protocol (IP) address, interface, hostname of the dying node, etc.). Existing protocols like Link Layer Discovery Protocol (LLDP) may be employed to derive the needed information of the dying node and send out the RFI syslog from the received remote node to the syslog/SNMP servers. All the members including the neighbor node find the device that is going to shut down using the source Media Access Control (MAC) address that is being sent with this data packet.

This in-band L2 OAM RFI indication is sent across the data packet which can be used by the upper layers to: trigger rerouting of the data traffic with lower convergence time; take the necessary Quality of Service (QoS) action; signal link/device characteristics; compute an alternate path; and provide a clear indication of the device that went down using the hostname/IPv[4/6] address/interface.

The problem addressed is not only about the broken RFI DG packet construct/transmit. The techniques described herein also transmit the same through an in-band mechanism efficiently and cheaply. In one example, the customer packet may have a new reserved MAC address with no other changes to the packet. This may be a bit slower. In another example, the customer packet has the new reserved MAC and the existing payload is removed. This may require computing a new cyclic redundancy check, and may be faster.

The solution to overcome the existing problem is to assign a reserved multicast RFI MAC address which will be used by the dying node to notify the neighbor device of its dying status instead of generating the existing OAM PDU DG packet, which is time consuming and a failure in some platforms.

When the Ethernet-OAM is configured on the device they become an automatic member of this multicast group (e.g., 01:80:xx:00:00:00). When the data traffic is flowing in the network and one of the devices is about to die, the dying node changes the destination MAC address to the reserved multicast RFI MAC address. In case of L2 switches, the source address is replaced as well. This ensures the neighboring node finds the device that is going to shut down using the source MAC address that is being sent with this data packet. Now the neighbor devices take the responsibility of constructing the DG packet (which was previously constructed by the dying node) and notifying the required nodes/server, as the existing functionality would have done.

Figure 1 below illustrates an example overview.

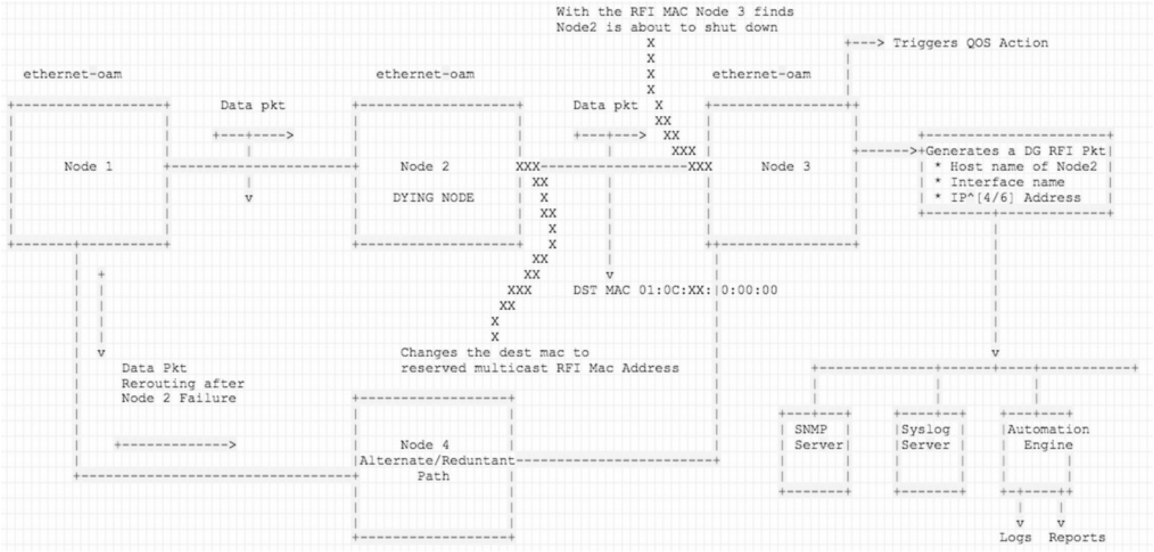

Figure 1

When a multicast IPv4 packet is encapsulated into an Ethernet frame, the destination MAC address is constructed using the Organizationally Unique Identifier (OUI) prefix of 01:00:5E and the lowmost 23 bits of the destination IPv4 address are directly copied to the lowmost 23 bits of the destination MAC address (the topmost bit in the fourth MAC octet is set to 0). In certain cases, switches and routers may receive a multicast frame as transit devices, but the destination MAC may not correspond to (is not constructed from) the destination IPv4 address of the packet in the frame.

For switches that implement these techniques, the multicast packet sent to this group may be taken up the control plane and processed. If there are switches that do not understand or implement this recommendation, there may be no impact because an L2-only Ethernet switch would not care about the discrepancy between the destination IP and MAC addresses. To a basic L2 switch, any destination MAC address with its Individual/Group (I/G) bit set to 1 (a group destination address) is essentially unknown and unlearnable because no host would use such an address as its source MAC address. As a result, any simple L2 switch may flood all multicast frames indiscriminately, and may not care about any discrepancy between the destination MAC address and the destination IP address of the packet inside the frame.

If the dying device is an L2 switch, the source MAC address is not the dying device’s unless it is made as such. Doing so would give any receiver the means to identify the dying node. If the source MAC address is left unchanged, this DG frame should not be forwarded beyond its neighbor because other nodes will not know where it is coming from. The techniques described herein enable changing the source MAC address if the dying node is an L2 switch or if the source MAC addressed is left unchanged to ensure it does not travel beyond the neighbor node.

For routers which implement these techniques, the multicast packet sent to this group may be taken up the control plane and processed. If there are routers that do not understand or implement these techniques, there will be no impact for the following reason. With respect to the operation of a router, the destination MAC address is important only to the degree of the router's incoming NIC treating that address as one of the addresses it is listening to, because that is what makes the NIC receive the frame and pass its contents to the corresponding driver for further processing. If it is not, the packet may be ignored.

Although examples described herein relate to using a dedicated multicast address to signal DG, other in-band methods may be utilized as well. For example, if the packet has room for carrying the DG as additional metadata, or the operational domain allows for insertion of the DG as metadata into the existing packet (e.g., repurposing the v6 flow label if it is not used in a domain), the DG may be carried in places other than the MAC address. Any suitable protocol that allows for carrying metadata natively may be used (e.g., Generic Network Virtualization Encapsulation (GENEVE), Network Service Header (NSH), v6 (via extension header), v4 (via option header), etc.). Other approaches that add metadata to customer packets and as such create room for additional metadata, such as In-situ OAM (IOAM), may also be used.

The techniques described herein enable bringing in the RFI with the data packet itself. They may be implemented by the ASIC and not overlooked by the CPU every time, as it works now. This ASIC-based implementation makes this one of the fastest RFIs in the field. These techniques may use existing infrastructure and have greater efficiency.

In summary, techniques are described herein for an alternate option to enable delivery in the in-band RFI. This may provide guaranteed delivery in the data packet, and may be complementary to existing OAM mechanisms.