IP and Ethernet Communication Technologies and Topologies for IED networks

Authors

Navindra Yadav	Eruch Kapadia
Principal Engineer,	Sr. Solutions Architect/Technical Leader
Cisco Systems, Inc.	Cisco Systems Inc.
nyadav@cisco.com	ekapadia@cisco.com

History

Version 1.0 October 20, 2010

Contents

1 Introduction 3
2 Ethernet Topologies with IEC 61850 based Substations 3
2.1 Physical Ethernet Topology choices for the Substation 4
2.1.1 Redundant Tree Topology 5
2.1.2 Ring Topology 6
2.1.3 Comparison of the two leading Network Topologies for IEC 61850 on Ethernet 7
3 Challenges with IEC 61850 traffic on an Ethernet Profile 8
3.1 IEC 61850 Traffic contained Inside a Substation 10
3.2 IEC 61850 Traffic outside the Substation 11
4 IEC 62439 and High Availability 12
4.1 PRP end point structure 14
4.2 Interoperability Issues with Parallel Redundancy Protocol of IEC 62439 15
4.2.1 Time synchronization 15
4.2.2 Trace Route 16
5 References 16

1 Introduction

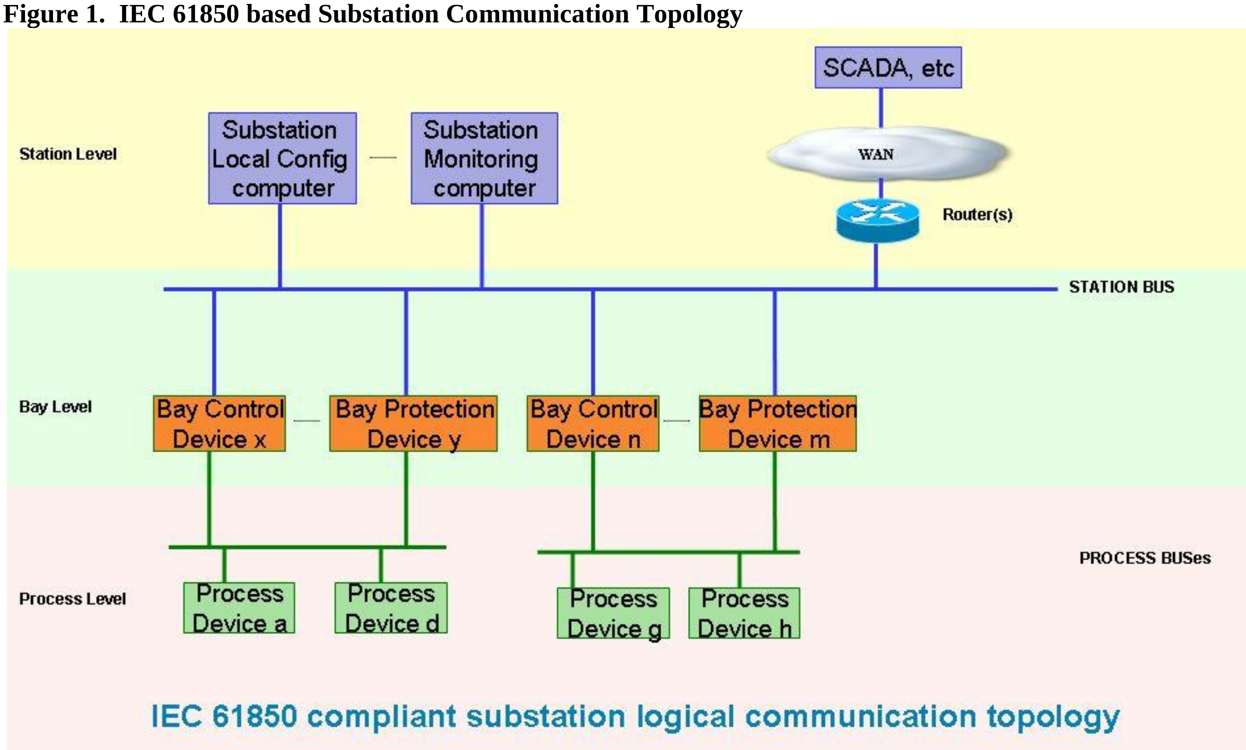
IED networks and associated applications require high degree of reliability, dependability, and provide deterministic behavior. In this paper we examine different network topologies and related technologies that provide different degrees of reliability and scalability for sub-station connectivity. We extend our analysis to both intra substation communication and communication between control center and substation using IEC 61850 networks and present the case for IEC 61850 profiles on top of IP for inter-operability with communication networks. Finally we examine the deployment of high-availability IEC 62439 parallel redundancy protocol in sub-stations and discusses the inter-operability challenges with IP network by examining time synchronization deployment challenges and traceroute behavior.

2 Ethernet Topologies with IEC 61850 based Substations

The choice of using Ethernet within IEC 61850 was made to leverage the cost effective and high speed Ethernet technologies used in the computer networking industries. In addition, Ethernet is a simple layer 2 protocol and makes use of common and familiar visibility tools and devices such as routers and switches that already exist within utility business networks.

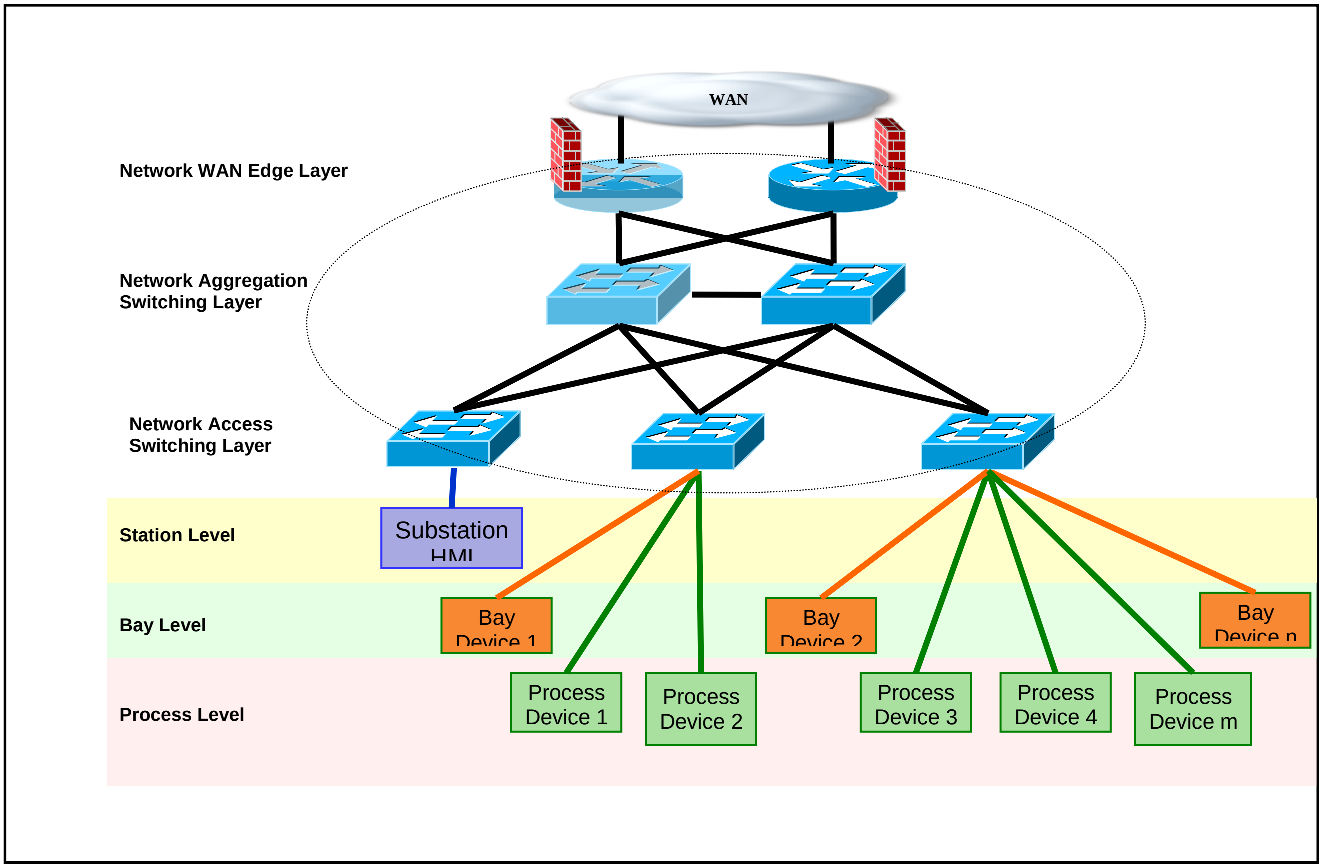
Ethernet based IEC 61850 Substations have a Station Bus and a Process Bus. The devices which interact with the Process equipment (aka primary equipment) sit on the Process Bus. The Figure 1. below shows an example of the IEC 61850 based logical topology.

Figure 1. IEC 61850 based Substation Communication Topology



The physical topology for the above logical topology is shown in Figure 2. below:

Figure 2. Substation Physical Topology



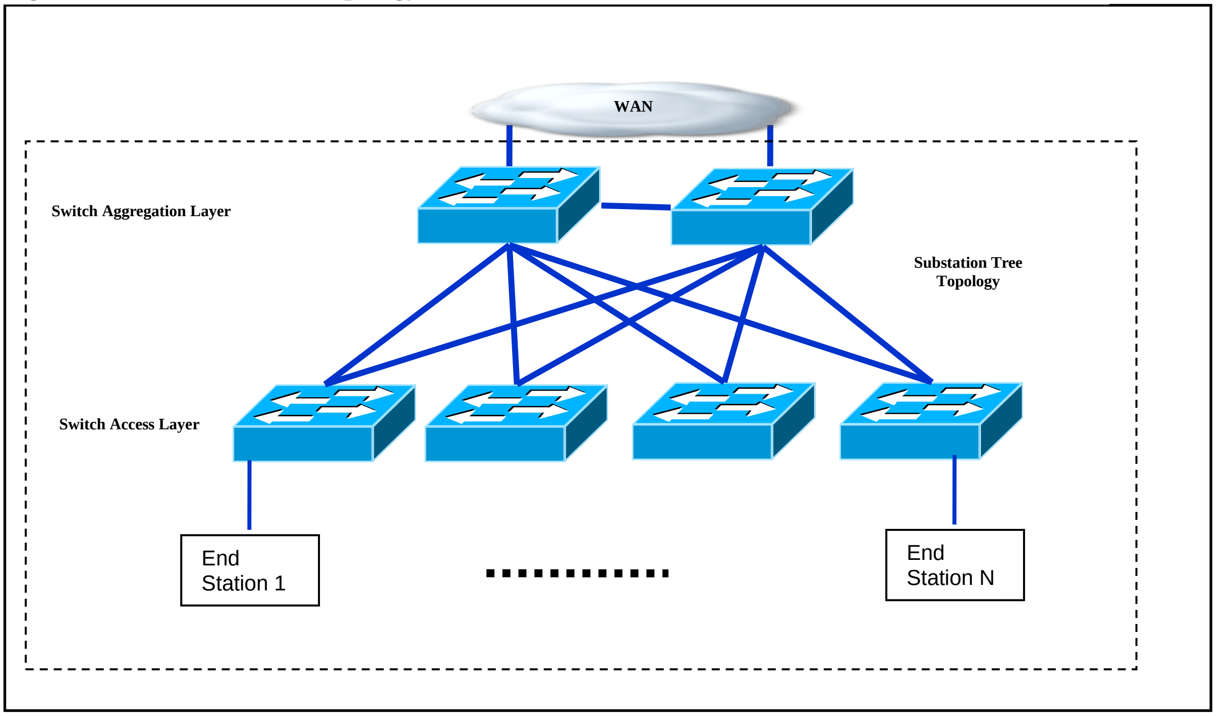
2.1 Physical Ethernet Topology choices for the Substation

The concepts presented in this paper are independent of the physical Ethernet Topology and can apply to any topology. We have selected a dual redundant tree based topology as shown in Figure 3. for the basis of discussion in this paper. The choice of redundant tree based Ethernet topologies is based on the experience Cisco has gathered in multiple industry segments. Before we delve deeper into the main topic of this paper we briefly establish the reasons in using redundant trees.

Ethernet LANs could be built using multiple physical topologies like redundant trees, rings, stars, etc. Multiple logical topologies can be over laid on top of the physical topologies. Logical topologies can be built using the concept of virtual LANs (VLANs). Almost all modern Ethernet LANs (ranging from simple enterprise campus network designs to complex data centers) are implemented by using full duplex Ethernet (100 Mbps or higher), and in majority of these deployments a redundant tree topology is used. The discussion below highlights the pros and cons of the two main competing Ethernet Topologies – Ethernet Rings and Ethernet Redundant Trees.

2.1.1 Redundant Tree Topology

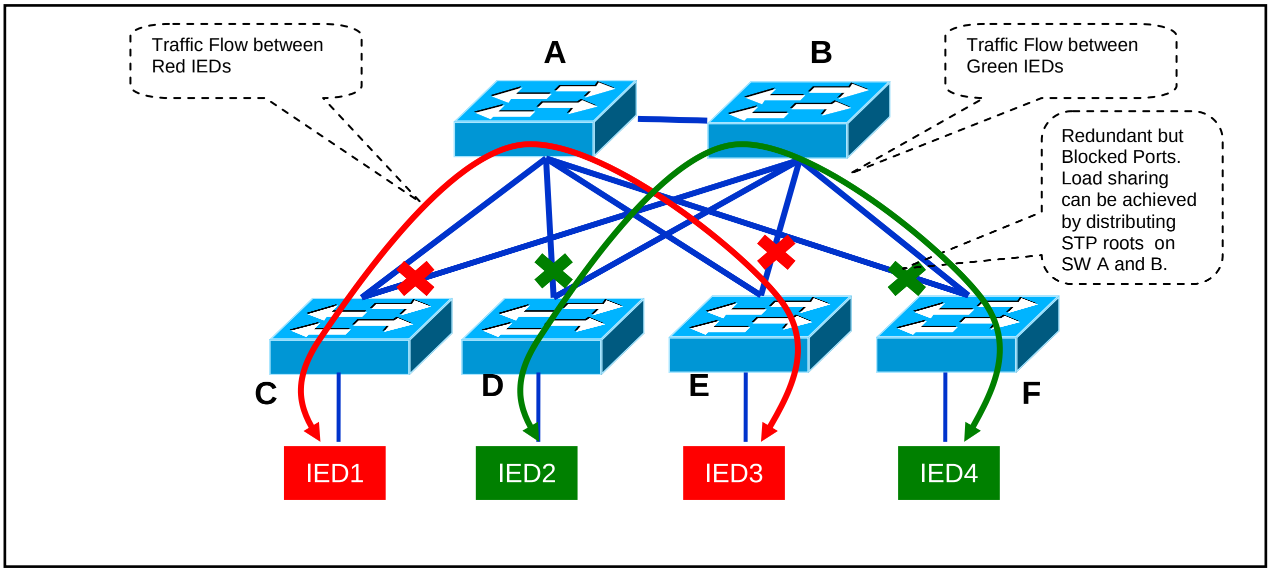
Figure 3. Redundant Tree Topology



The Figure 3. above show a redundant tree physical topology which could be used for the station connectivity and process bus networks inside a substation. Every switch has a redundant path to every other switch in this network. The topology above has Ethernet loops which for reasons mentioned in IEEE 802.1d have to be broken. The Figure 4. below shows the redundant tree topology with the redundant links blocked for a specific VLAN to prevent traffic loops. This topology offers multiple advantages over a ring topology as listed in section 2.1.3.

The redundant tree topology as shown in the Figure 4. allows for traffic separation between the IEDs. This can be achieved by pruning VLANs on the switches that do not require to carry that VLAN to the attached IEDs. In the figure above IED1 and IED2 are part of the red group and therefore the access Switch C and E have pruned the access to Green VLAN. Similarly IEDs attached to access Switch D and F does not have access to Red VLAN. In comparison to the ring topology, the hierarchical tree topology also offers better bandwidth utilization and scalability. By distributing the spanning roots for different sets of VLANs on the switch A and B traffic is load shared on both the uplinks from the access switches. The tree topology also offers superior QoS as the jitter is lowered because of lower traffic interference due to the ability to achieve traffic separation. The smaller broadcast and multicast flooding domains also improve security and resiliency.

Figure 4. Multi Vlan Redundant Tree Topology



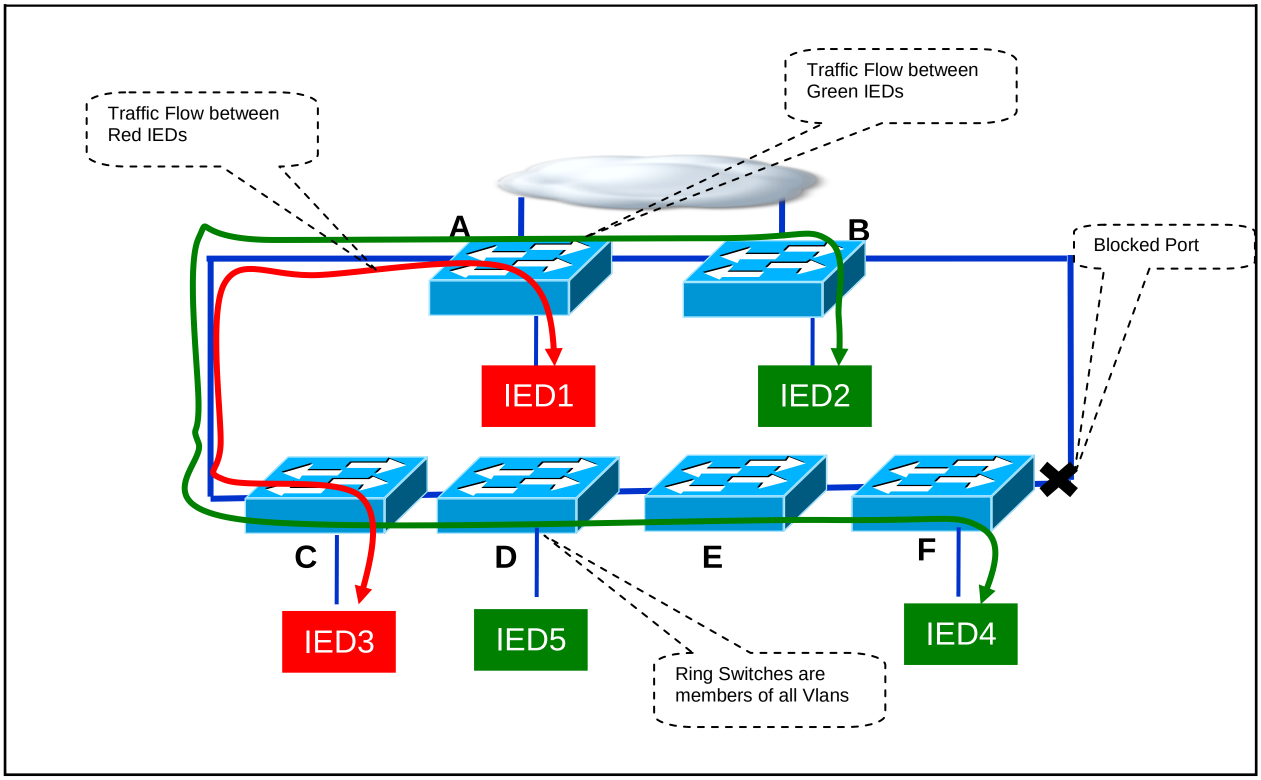
The Spanning tree protocol is typically used in the redundant tree topology shown above to block the redundant links to prevent traffic loops. The IEEE 802.1d spanning tree protocol with default timers can result in 50 sec or more timeout of traffic when it transitions high availability links from blocking to forwarding states. Ethernet Rapid Spanning Tree Protocol (RSTP) (aka IEEE 802.1w) is a faster variant of IEEE 802.1D standard which allows faster convergence of spanning tree resulting in faster recovery from failure of links in the redundant environment. While RSTP convergence in the order of a few seconds is sufficient for most environments it does not provide the milli second recovery required in substations. In order to achieve faster recovery the recommendation is to deploy the Flex Link feature when deploying Ethernet trees in substations. Flexlinks typically converges the network in sub 100 milli second. For more information on Flexlink check refer to Cisco Systems, Inc. documentation on Flexlink [Ref. 10].

2.1.2 Ring Topology

The ring topology as show on in Figure 5. while simple from a physical perspective does not provide clean traffic separation. All VLANs must be present on all switches. The presence of all VLANs on all switches leads to poor segmentation and cyber security profile. It also results in a bad design for traffic flood containment. The ring topology also leads to

- Inefficient bandwidth utilization
- Non deterministic latencies through the Ring as changes in the location of the block port impacts the switching latency
- Inferior QoS as compared to a Tree
  - Traffic from IED4 to IED2 competes with traffic of the same or higher priority at each of the 6 switches on the way.
  - However traffic between IED4 and IED5 competes with traffic of the same or higher priority at each of the 3 switches on the way
  - If the block port moves between Switch D and E, the situation reverses.
  - So Traffic behavior depends on the block point in the ring

Figure 5. Ring Topology



In the ring topology also Ethernet spanning tree loops are also broken by using a Spanning tree protocol IEEE 802.1d which can cause a traffic timeout of 50 seconds or more to recovery from a failed link . Rapid Spanning Tree Protocol (RSTP) (aka IEEE 802.1w) is a faster variant of IEEE 802.1D.   However RSTP converges in the order of a few seconds. The recommendation is therefore to deploy the Resilient Ethernet Protocol (REP) when deploying Ethernet rings in substation networks. REP typically converges the network in sub 100 milli second. For more information on REP check out Cisco Systems, Inc. documentation on the REP protocol [Ref. 9].

2.1.3 Comparison of the two leading Network Topologies for IEC 61850 on Ethernet

Table 1. Comparison of Networking Topologies

Areas	Redundant Trees	Rings
Physical Redundancy	Yes	Yes
Connectivity/Topology	Trees – Hierarchical	Ring – Simple
Predictable Latency	Superior (Fixed and deterministic latency. Tree depth determines the number of hops.)	Inferior (Latency varies. The number of hops between the source and the destination depends on where the loop in the ring is broken. When the blocking point changes the

		latency also changes.)
Smaller Fault Domain	Superior (Smaller Fault Domain)	Inferior (the whole ring is the fault domain.)
Bandwidth Efficiency	Superior	Inferior (all inter switch traffic contends for the ring bandwidth)
Scalability	Superior (only the leaf switches and the root switches through which traffic is exchanged learn about the switch)	Inferior (all switches have to learn about all end points. Least capable switch determines the capacity of the ring)
Multicast and Broadcast Containment	Superior	Inferior
Maintenance and serviceability	Superior (no downtime to the network to add a new leaf switch)	Inferior (downtime seen by the network to add a new switch to the network)
Fairness	Superior	Inferior (traffic sent by the edge switches has to compete with similar class of traffic at every hop on the ring)
Fast Convergence	Faster convergence (order of sub 100msecs) can be achieved by using some proprietary techniques like FlexLinks and other protocols from 62439.	Faster convergence (order of sub 100msecs) can be achieved by using some proprietary techniques like FlexLinks and other protocols from 62439.
Cyber Security	Superior (not all switches have to have all vlans, also the flooding domains are smaller)	Inferiror (all switches have to have all vlans)

3 Challenges with IEC 61850 traffic on an Ethernet Profile

In the current definition of IEC 61850 all the 61850 based GOOSE and sampled value traffic rides directly on top of Ethernet. There is no well developed network layer or transport layer in the middle. This section focuses on highlighting the issues that are created due to this decision. We highlights the pain points caused because of the above decision inside the substation in section 3.1, and discuss the pain points caused when 61850 based GOOSE and Sampled value traffic have to be sent outside the substation in section 3.2.

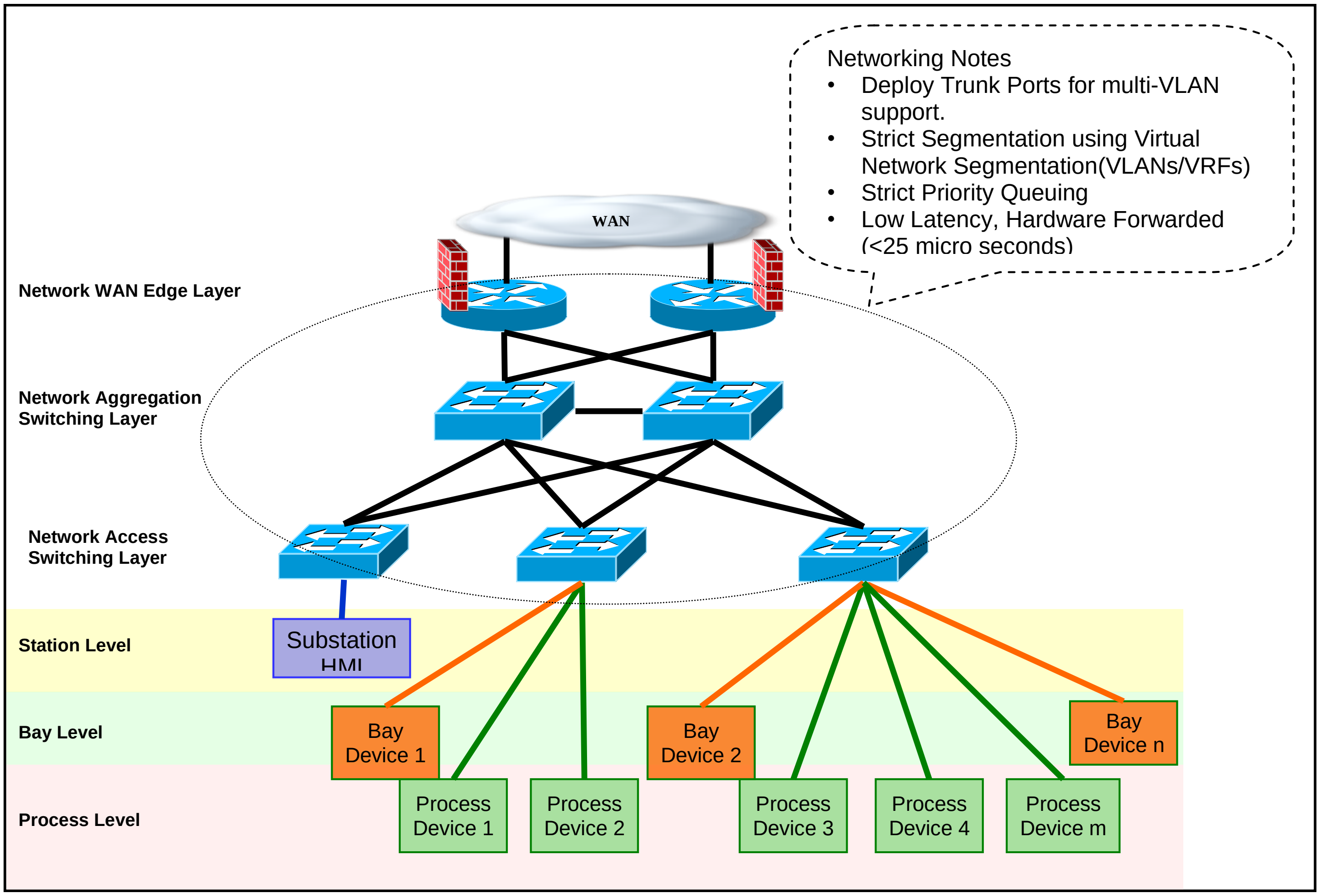
At the time of writing this paper the authors have come across drafts where an IPv4 and IPv6 profile for IEC 61850 are being defined for carrying Phasor Measurement Traffic. This discussion is in drafts of IEC 61850-90-5 Technical Report. The authors strongly recommend developing and standardizing these IPv4 and IPv6 based profiles (along with the TCP and UDP *transport) for carrying not just Phasor measurement unit traffic (C37.118), but also non PMU, Relay traffic. By moving IEC 61850 GOOSE and SV messaging on top of an TCP/IP protocol stack the issues the authors present below can be resolved.*

Talking to various people in the industry who are also part of the 61850 (TC 57 working group) the authors gathered that the reason for not sending GOOSE messages over TCP/IP were:

- No need for IP as there was the feeling that the 61850 GOOSE/SV traffic would be contained inside a substation. This is no longer true, for example 61850-90-1, inter substation Tele protection, PMU traffic streaming, etc are examples of traffic not contained inside a substation.
- There is the impression that adding IP headers increases the latency of messaging in the network.
    - This point too is no longer true with the commercially available off the shelf Ethernet technology. Since the late 1990s Cisco Switches have been forwarding Ethernet and Ethernet+IPv4/IPv6 packets at the same wire speed forwarding rates and with the same switching latencies.
    - The Cisco Connected Grid Switch has forwarding latencies between 8 micro seconds to 25 micro seconds irrespective of whether the packets have an Ethernet or IPv4/v6 or TCP or UDP headers on it. The latency range comes from the size of the Ethernet Packet which can range from a frame size of 64 bytes to 1518 bytes.

3.1 IEC 61850 Traffic contained Inside a Substation

Figure 6. IEC 61850 Inside a substation



The issues caused due to the absence of an IP (IPv4 or IPv6) layer in IEC 61850 based GOOSE and SV messages (which predominantly use multicast) are:

- Inefficient Multicast Traffic Distribution
  - Non IP based Ethernet Multicast inside a substation results in flooding the packets inside the vlan (flooding domain).
  - Pruning layer 2 based Ethernet multicast with configuration is not a scalable and maintainable option in the long term.

  If an IP layer was present in these GOOSE and SV packets then these problems would have been avoided. The multicast would only have been delivered to the appropriate switches and end devices. This would be achieved using dynamic multicast group membership protocols like IGMP for IPv4 and MLD for IPv6. These protocols are widely deployed in the IT industry and work both at the layer 3 and layer 2 level. At layer 2 these features are called IGMP snooping or MLD snooping. Refer to the reference for more details. The network learns which access Ethernet ports and which switches have an end point which is interested in receiving traffic for a specific multicast group (and even from which source). Based on this they can deliver traffic optimally.

- Cyber Security
    - Unnecessary flooding of multicast traffic to unneeded end points increases the security risks around information leakage.

    If a protocol like IGMP or MLD is used the access switches can tie that with multicast traffic authorization feature to determine if the end point is authorized to receive the traffic. When using source specific multicast, the access switches can prevent denial of service attacks by enforcing the source bindings at the access of the network.
- Scaling broadcast domains

3.2 IEC 61850 Traffic outside the Substation

Currently use cases like distance protection, tele protection, phasor measure measurement units, CRAS, etc. send IEC 61850 based GOOSE and Sampled Value messages outside the substation. The message exchanges may be between multiple peer substations or between the substation and the control center. The big challenge in either of these scenarios is backhauling Ethernet traffic across the WAN network for the following reasons:

- Ethernet is not and was not built for Wide area communication.
- Ethernet is not a routable protocol.
- If one examines the current Ethernet and IP based networks, the reason these networks scale to the size of the internet and even larger is because of the key attributes of containment and hierarchy.
- In a layer 2 domain (Ethernet bridge domain), traffic reaches the destination by getting forwarded through bridge tables (which are populated based on mac address learning) or flooding the traffic when the destination is unknown. Large bridge domains can potentially de-stabilize the network and recovery times are very difficult to determine.

Some of the techniques used to connect the different Substations networks over a WAN are (some of them are discussed in the IEC 61850-90-1 TR):

- Tunneling
    - GRE Tunnel. Tunneling the Ethernet frames on top of an IP GRE tunnel
    - Layer 2 Tunneling Protocol
- Protocol Translation/Gateways
    - Isn't a practical solution for low latency applications which need inter substation communications
- Encapsulations
    - Multi Protocol Label Switching (MPLS)
        - Pseudo Wires
        - Virtual Private LAN Service (VPLS)
        - Other similar protocols

It should be noted that Encapsulations are different from tunnels in that they do not require tunnel interfaces and do not add extra latency to packet forwarding (no packet re-circulation in the forwarding plane is required for the features configured on the tunnel interface)

However, some of these technologies that allow layer 2 network extensions across the WAN still suffer from the issues of flat large layer 2 networks. Flat Ethernet network work well for smaller networks but are not an ideal technology to use for large flat layer 2 Ethernet network. Following are some of the reasons that argue against large flat layer 2 networks:

- A topology change would flush the forwarding state on all the bridges in the layer two domain
- A fault in the layer two domain propagates to the entire layer two domain. This becomes especially significant when multiple substations get connected with over the WAN and share the same VLAN, etc.
- If multiple substations are tied together in a layer two domain them a fault in one substation will propagate to other substations too. Example a spanning tree loop in one substation will spill over to other substation networks.
- The flooding domains become large and wasteful. Example GOOSE and SV based multicast messages will flood to all the substations with the VLAN, irrespective of whether the substation is interested in the GOOSE message or not.
- To limit the flooding domains, if the number of VLAN is increased, then that creates Vlan proliferation issues and adds to the operational and management expense.

Hence we need profiles for IEC 61850 GOOSE and Sampled Value messages on top of IPv4 and IPv6.

4 IEC 62439 and High Availability

In this section we examine the IEC 62439 in particular the Parallel Redundancy Protocol(PRP).

IEC 62439 discusses the different options available to construct highly available networks as shown in the Table 2 below. We take the best possible option viz. PRP in terms of minimal network down time and look at the inter-operability issues of running Ethernet and IPv4/IPv6 on top of an IEC 62439 network.

The Table below shows a comparison of the different protocols.

Table 2. IEC 62439 Protocol comparison

Protocol	Re-convergence Time	Packet Loss Observed	Where does the Protocol Run	Network Topology	End Node Network attachment
IP Routing	>30 sec Non deterministic	Yes	Network	Any	One
STP (802.1d)	>20 sec Non deterministic	Yes	Network	Any	One

RSTP	>2 sec Non deterministic	Yes	Network	Any	One
FlexLink	<100 msec	Yes	Network	Tree	One
REP (Resilient Ethernet Protocol)	<100 msec	Yes	Network	Ring	One
1+1 or PRP	**0 secs**	**No**	**End Point (MAC layer)**	**Two or more Network Any Topology**	**2 or more links**

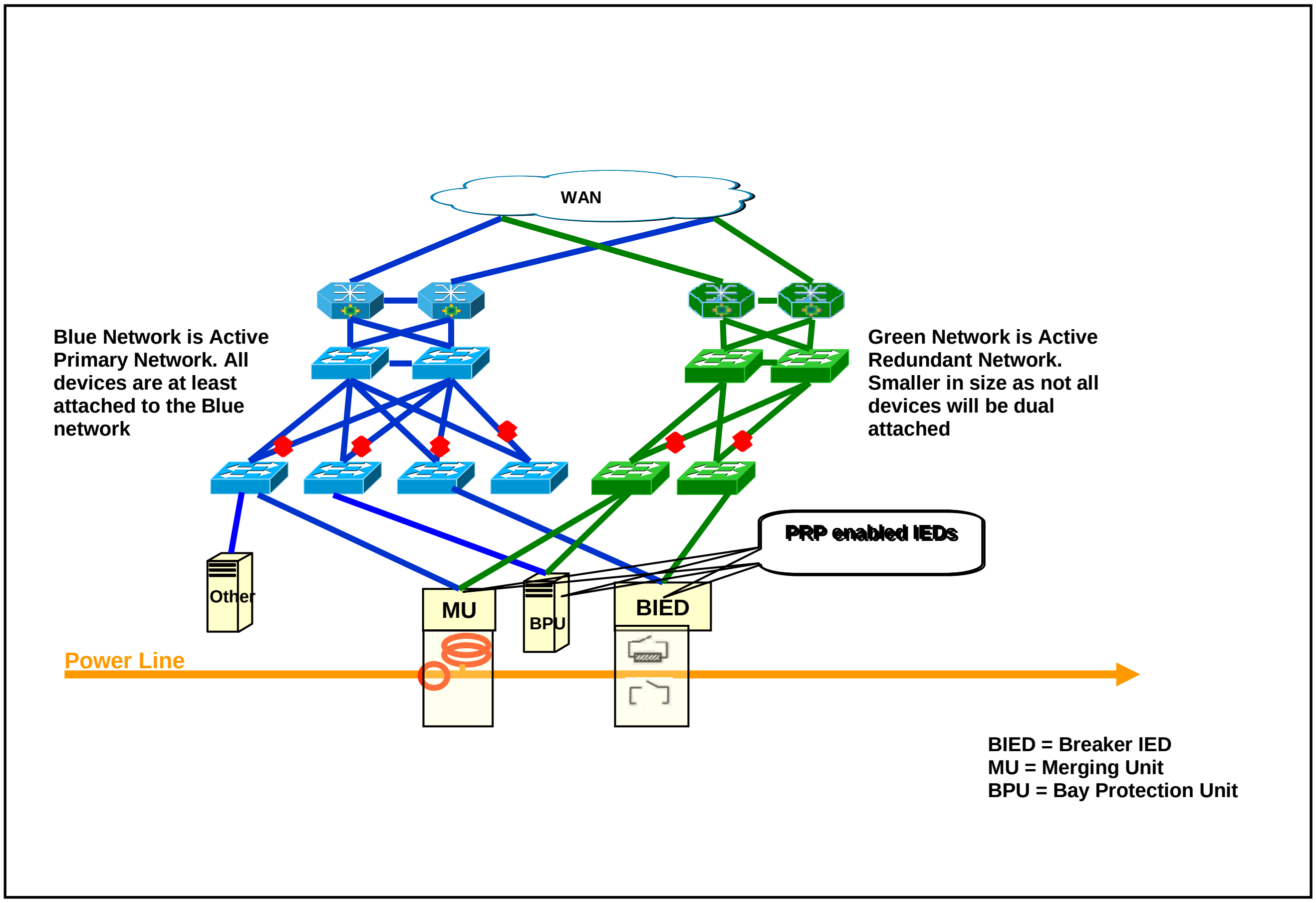
Parallel Redundancy Protocol(PRP) is a protocol defined to improve the high availability of a network [Ref. 4]. The principle of operation for PRP is very simple. Two (usually just two, though the there could be more than two networks) physically independent and uncorrelated networks are present. The PRP aware end points connect to both the networks. The two network LANs are identical in protocol at the MAC-LLC level, but they can differ in performance and topology. Transmission delays may also be different. The LANs have no direct connection between them and they are assumed to be fail-independent.

A doubly attached node implementing PRP (DANP) is attached to both network LANs. The figure blow shows the Blue and the Green LAN. Singly attached nodes (SANs) can be attached in two ways:

1. SANs can be attached direct to one of the networks only. SANs can only communicate with other SANs on the same LAN. For instance, the device named "Other" could be a sub-station HMI device. This HMI device may not require extremely high degree of availability is connected to the blue network and can communicate with other devices inside the Blue network, but not with the devices in the green network.
2. SANs can communicate with all DANPs in their own network. SANs can be attached via a redundancy box which implements PRP and which is connected to both the blue and green network LANs. Such SANs can communicate with all SANs and DANPs.

The Figure below shows the an Intra substation network with a mix of PRP enabled end points and non PRP enabled end points. The blue topology is the primary topology to which both SAN and DNAP end points are connected. The green network is the topology to which only the DNAPs are connected, hence the Green topology may be somewhat smaller as compared to the blue topology.

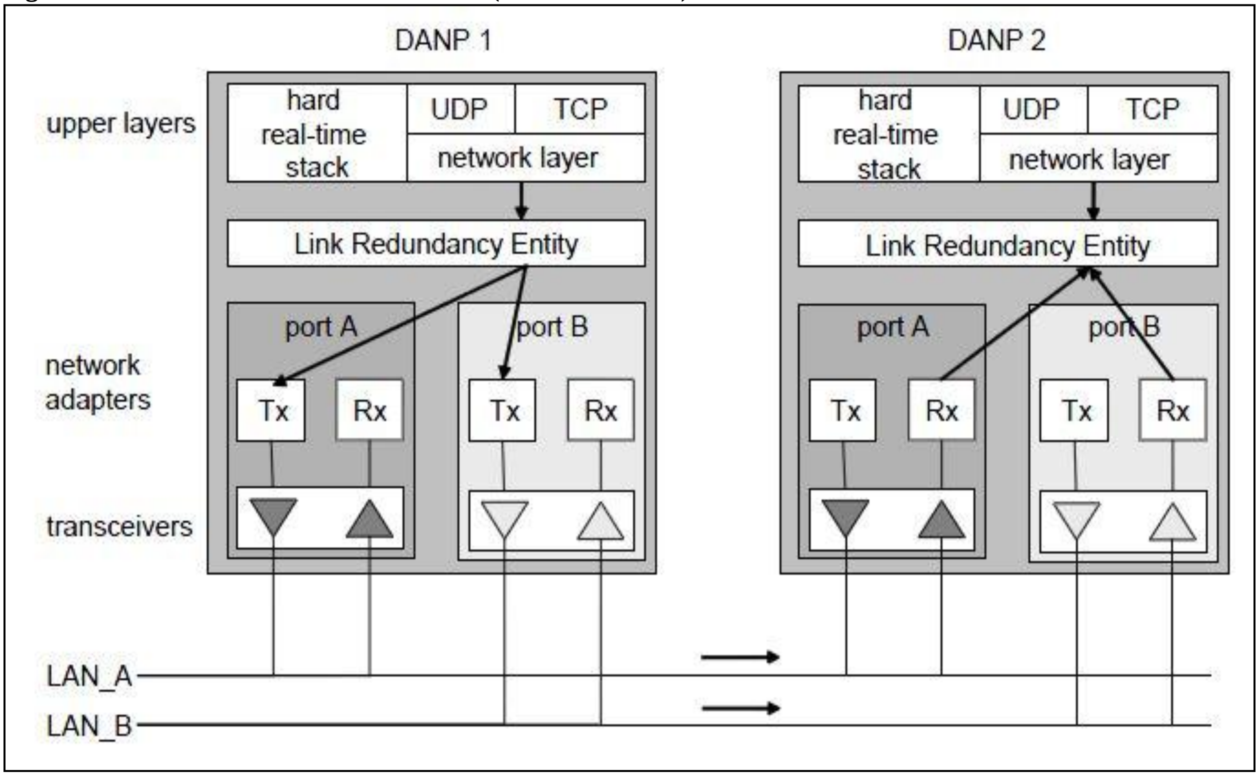
Figure 7. Intra-substation Redundant tree topology with PRP endpoints



4.1 PRP end point structure

Each PRP end point node has two physical ports that operate in parallel and that are attached to the same upper layers of the communication stack through the Link Redundancy Entity (LRE), as shown in the figure below:

Figure 8. PRP two DANPs communication (REF. IEC62439)



The Link Redundancy Entity(LRE) has two goals. It duplicates frames and performs redundancy management. This layer presents to its upper layers the same interface as the network adapter of a non-redundant adapter. When receiving a frame from the end points upper layers, the LRE replicates the frame and sends it through both its physical ports to both the green and blue networks at nearly the same time. The two frames transit through the two LANs with different delays, ideally they arrive at around the same time at the destination node.

When receiving frames from the network, the LRE forwards the first received frame of a pair to the node's upper layers and discards the duplicate frame (if it arrives). For management of redundancy, the LRE appends a redundancy check trailer (RCT) including a sequence number to the frames it sends to keep track of duplicates. In addition, the LRE periodically sends PRP_Supervision frames and evaluates the PRP_Supervision frames of the other DANPs. If a frame is received without the RCT the LRE sends the frame to the higher layer, this is to handle the case when a SAN sends a frame to the DNAP end node. SANs would not know how to add a RCT to the frame, and this is how the DNAP would distinguish this case.

4.2 Interoperability Issues with Parallel Redundancy Protocol of IEC 62439

4.2.1 Time synchronization

Networks with high availability also require a precision timing support. Time protocols like IEEE 1588 (aka Precision Time Protocol - PTP), Network Time Protocol, etc measure the path delay between the time slave and the time master. Based on this delay they compute the offset to add or subtract from the time synchronization messages. The problem is if the time protocol does not know of the PRP networks below it at the physical layer, it will get varying delay estimates from the duplicate packets travelling the two different networks and this will lead it to calculate an incorrect estimate.

The solution to this problem resides in exposing the fact that there are two physical networks to the time protocol. For further details on the deployment options refer to the document IEEE 1588 [Ref. 5].

4.2.2 Trace Route

A different problem manifests itself at Layer 3. When a trace route command is issued, then the responses to the trace route will be seen from the two physical networks. This will clobber the result of the trace route and the results will not make sense.

One solution to this problem, is to either teach trace route about the two physical networks. However this defeats most of the value of PRP, because the higher in the stack one has to teach the protocol about PRP, the lesser would be the value of using PRP.

So another solution is to create the same routing topology and assign the same IP addresses to the equivalent routers in the two topologies. The next challenge is how do we enable the two overlapping IP address spaces to coexist. The solution to the second problem is to leverage the Virtual Route and Forwarding (VRF) technology along with Route Distinguishers in the WAN and router space. VRFs gracefully support overlapping address spaces and carry routes from differ spaces, and keep them unique by adding route distinguishers.

5 References

1. IEC 61850 Standards [http://webstore.iec.ch/]
2. IEC 61850-90-1 [http://webstore.iec.ch/webstore/webstore.nsf/artnum/043864]
3. IEC 61850-90-5 Draft Technical Report [http://www.iec.ch/cgi-bin/procgi.pl/www/iecwww.p?wwwlang=e&wwwprog=sea1112.p&progdb=db1&css_color=purple&committee=95&class=&refno=&type=&date=]
4. IEC 62439 Standard – Chapter 6 on Parallel Redundancy Protocol. [http://infosys2008.iec.ch/cgi-bin/procgi.pl/www/iecwww.p?wwwlang=english&wwwprog=pro-det.p&progdb=db1&He=IEC&Pu=62439&Pa=3&Se=&Am=&Fr=&TR=&Ed=1]
5. IEEE 1588 [http://ieee1588.nist.gov/]
6. Hans Weibel, Sven Meier, "IEEE 1588 applied in the environment of high availability LANs", International Symposium on Precision Clock Synchronization for Measurement, Control and Communication, October 2007, Vienna.
7. Cisco MPLS/VPLS [http://www.cisco.com/en/US/tech/tk436/tk891/technologies_q_and_a_item09186a00801ed3bf.shtml]

8. Cisco L2TPv3 [http://www.cisco.com/univercd/cc/td/doc/product/software/ios120/120newft/120limit/120s/120s23/l2tpv3.htm]
9. Cisco REP [http://www.cisco.com/en/US/prod/collateral/switches/ps6568/ps6580/prod_white_paper0900aecd806ec6fa.pdf]
10. Cisco Flexlinks [http://www.cisco.com/en/US/docs/switches/lan/catalyst3550/software/release/12.2_25_see/configuration/guide/swflink.html]
11. Cisco Connected Grid Router [http://www.cisco.com/en/US/prod/collateral/routers/ps10967/ps10977/data_sheet_c78_593509.pdf]
12. Cisco Connected Grid Switch [http://www.cisco.com/en/US/prod/collateral/switches/ps10968/ps10978/data_sheet_c78_593672.pdf]

End of Document