Appropriate Testing Reveals New Best-in-Class Topology for Ethernet Networks

D. Dolezilek, J. Dearien, A. Kalra, and J. Needs
Schweitzer Engineering Laboratories, Inc.

Presented at the
13th International Conference on Developments in Power System Protection
Edinburgh, United Kingdom
March 7–10, 2016

Appropriate testing reveals new best-in-class topology for Ethernet networks

D. Dolezilek*, J. Dearien*, A. Kalra*, J. Needs*

**Schweitzer Engineering Laboratories, Inc., 2350 NE Hopkins Court, Pullman, WA 99163 USA, dave_dolezilek@selinc.com*

Keywords: Ethernet, testing, performance, reconfiguration, conformance.

Abstract

This paper (an adaptation of [1]) describes the results of a research effort to discover which Ethernet switch topology is the most reliable and resilient for peer-to-peer protection communications. The results are based on many thousands of network failure tests performed in order to arrive at statistically significant results. In the end, by adding a few additional cables and making a few settings changes, a new Ethernet network topology was engineered that has n-5 and n-11 switch- and cable-failure resilience, respectively. Also, the failures that do take place are corrected so quickly that they do not impact the delivery of protection signals.

1 Introduction

Protection pilot schemes rely on the exchange of protection signals between intelligent protective relays to perform interlocking and teleprotection. Newer methods that exchange digital messages that contain the signal statuses, rather than older methods such as tone gear or pilot line carrier, have fundamentally changed the measures of reliability, dependability, and security. Dependability now not only refers to the ability of relays to react when needed but also to the ability of communications channels to deliver every signal instantly and to refrain from dropping messages. Security no longer simply refers to relays refraining from performing unintended operations but also to communications channels refraining from delivering unwanted, corrupted, and repeated messages. Reliability not only refers to the availability of the protection scheme but also to the ability to prohibit mistaken commands and cyberintrusions from disrupting protection application communications.

International standards, such as IEC 60834-1, Teleprotection Equipment of Power Systems—Performance and Testing—Part 1: Command Systems, describe very specific reliability, security, and dependability requirements for protection communications. However, traditional Ethernet switch topologies (such as the ring topology) have been designed for convenient installation, and the behavior of relays has been forced to change to minimize the unintended consequences of using Ethernet for protection. Therefore, it was necessary to evaluate the first principles of the fundamental behavior of Ethernet in order to correctly engineer an Ethernet network topology for use within protection applications.

The Ethernet communications system is a switched network with several physical cable paths or loops, like cables in an electrical distribution system where one is active and the others may act as hot standby. Ethernet packets are prevented from traveling in a loop back toward their intelligent electronic device (IED) of origin where they would be rebroadcast into the network each time they are received. This unending rebroadcast of messages creates a packet storm of unwanted data and high-bandwidth consumption that jeopardizes network performance. The Spanning Tree Algorithm (STA) detects these physical cable loops and uses an Ethernet switch mechanism to disable one or more cables and prevent looping. This suspends packet flow on the cables that are acting as hot standby. This works much the same way electric energy is prevented from looping back toward its source by a power system switch that physically opens the circuit and stops energy flow. And, similar to an automated energy distribution network, when there is a failure in the network, the system detects and isolates it and then reconfigures among the hot-standby paths to quickly begin delivering packets again. When a portion of the Ethernet network is unavailable to deliver packets, we refer to it as being dark. Therefore, the period of time a network channel is interrupted and cannot deliver packets between perimeter ports where IEDs are connected is referred to as network darkness. Similar to energy distribution systems, it is extremely important for signaling to keep periods of network darkness short and infrequent in Ethernet packet distribution networks.

For energy distribution, the system interruption duration includes both network reconfiguration and reestablishment of energy delivery to the consumer. For Ethernet packet delivery systems, the interruption duration similarly includes network reconfiguration plus the subsequent reestablishment of the channel and packet delivery to the consumer. The IEC 61850 Communication Networks and Systems for Power Utility Automation—Part 90-4: Network Engineering Guidelines Technical Report echoes common best engineering practices in requiring that the system average interruption duration, or period of darkness, for each possible Ethernet packet delivery channel failure mode be tested and measured [2]. Best engineering practice is to design systems where the median time of darkness is brief enough to be within the maximum signal delay time. The statistical distribution indicates the worst-case signal channel delay caused by the failure of an Ethernet switch. The duration and probability of the worst-case delay must be understood and mitigated via the selection

of network devices with sufficient availability, measured as mean time between failures (MTBF) in years.

The time duration to perform protection, control, and monitoring (PCM) signaling includes processing within both the source and destination IEDs, as well as propagation of the digital message through the network. Overall application reliability is maximized via dual primary PCM applications, each with its own digital messaging network. The testing methods presented in this paper are equally applicable to testing individual or dual primary networks. Even though both serial and Ethernet networks can be deployed individually or redundantly, it is not possible to answer questions about Ethernet network behavior the same way it has been possible with serial networks. For example, multiservice Ethernet shares the available bandwidth with signaling and other protocols, which may affect message delivery behavior. Also, message parameters in the Ethernet packets work in concert with switch settings to control signal channel paths, and therefore delivery performance, through the network. Perhaps the most useful difference Ethernet provides is the ability to reconfigure after a cable or switch failure to use the hot-standby path. Once reconfiguration is completed, signaling proceeds normally; however, periods of darkness during the reconfiguration may impact the signaling during a power system event. These differences, which make Ethernet networks flexible for reconfiguration after failures, create a challenge for understanding Ethernet signal channel behavior.

2 Signal transmission, transfer, and transit time

The transfer time specified for an application is the time allowed for a signal or data exchange through a communications system. Transfer time is shown in Fig. 1 (which is from IEC 61850-5) as the time duration between the action of communicating a value from the logic processing of one device to the logic processing within a second device as part of an application. The time duration to publish signal information from Physical Device 1 (PD1), deliver it via a protocol message, and act on it in Physical Device 2 (PD2) is the transmission time of the signal or information. This transmission time duration represents actually performing an action as part of a communications-assisted automation or protection scheme. The transit time, t_b, is the time duration for the message to travel through the communications network.

Enhancements to IEC 61850 that are documented in Edition 2 numerate different types of messages and their associated transfer times, as shown in Table 1.

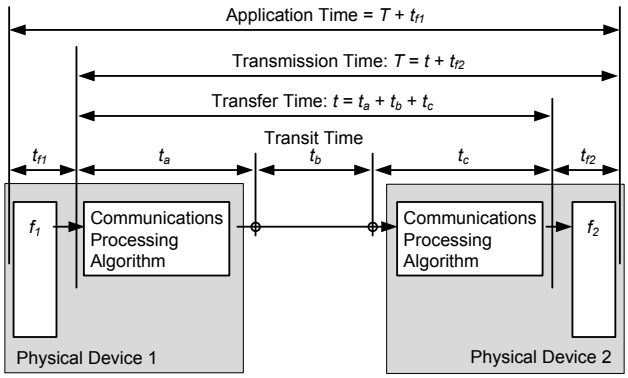

Fig. 1. Transmission time and transfer time based on IEC 61850-5.

Transfer Time Class	Transfer Time (ms)	Application Example
TT0	>1,000	Files, events, log contents
TT1	1,000	Events, alarms
TT2	500	Operator commands
TT3	100	Slow automatic interactions
TT4	20	Fast automatic interactions
TT5	10	Releases, status changes
TT6	3	Trips, blockings

Table 1: IEC 61850 transfer time requirements [2].

3 IEC 61850 GOOSE and Ethernet network test criteria

The IEC 61850 Part 90-4 Technical Report provides advice on network engineering and commissioning. Section 5.3.17 describes testing and recommends the following: "Once the network has been designed, its compliance to the requirements needs to be tested, first as a design verification, then during factory acceptance tests and finally at site acceptance" [2]. This technical report also requires that during operation, an appropriate subset of the tests continue to monitor the network so as to detect and mitigate failures.

However, it is very difficult to cause the worst-case event for t_a, t_b, and t_c from Fig. 1 simultaneously. Therefore, best engineering practice requires that we test and measure the worst case for each time individually and calculate the total worst case as the aggregate ($t_a + t_b + t_c$). Experience shows that Ethernet switches designed for Ethernet Generic Object-Oriented Substation Event (GOOSE) signaling typically deliver packets in a normally operating network in well under 1 millisecond. For this paper, we used IEDs and Ethernet network switches that together meet a packet transfer time for protection signal messages of 3 milliseconds. IEC 61850 identifies messages capable of meeting the 3-millisecond signal transfer time to be message Type 1A, Performance Class P2/P3 [3]. Other message types that satisfy less critical timing are also described by IEC 61850. However, the other message types are not appropriate for protection signal exchange for communications-assisted protection and interlocking. Also, these IEDs synchronize to an IRIG-B

source with microsecond accuracy, so IED clock synchronization error is negligible and ignored. The multiple devices are synchronized to the same time source and therefore share the same absolute time. Time synchronization and accuracy are important because we use IED time stamps to calculate time durations for test results.

IEDs perform each task, such as detecting a power system change and subsequently performing logic and logging a Sequential Events Recorder (SER) report, at some point during each operating cycle. The IEDs do not monitor inputs or processing logic continuously but rather are designed with specific processing intervals that determine how often they scan their inputs and process their logic. For the IEDs used in tests with a 2-millisecond operating cycle, a binary input and a Boolean logic variable change of state are recognized only once every 2 milliseconds. Signal trigger events, such as digital input contact closure and incoming or outgoing GOOSE signal bit changes and their subsequent SERs, will happen at asynchronous points during the PD1 operating cycle f_1 shown in Fig. 1. Signal reaction events, such as logic equations, digital output contact closure, and incoming or outgoing GOOSE signal bit changes and their subsequent SERs, will happen at asynchronous points uniformly distributed throughout the PD2 operating cycle f_2 shown in Fig. 1. Using the range rule for standard deviation, we approximate the standard deviation for time-stamp error to be one-fourth of the operating cycle time and the mean as one-half. Therefore, we approximate the typical time for f_1 or f_2 reaction processing to be one-half of the operating cycle duration. Also, the time-stamp values of the SER are accurate to 0 + 0.5 operating cycle duration. Ethernet network failures are tested to validate how they impact time t_b shown in Fig. 1.

4 Ethernet network reconfiguration

There are several standardized and proprietary algorithms and protocols used to determine primary and failover paths and the rules of how to change between them. There are two types of network failures: switch failures (bridge death) and link failures (link loss). Understanding the behavior of the network reconfiguration algorithm is crucial for engineering a network suitable for critical messaging.

The STA is a standard, widely used method that uses the Rapid Spanning Tree Protocol (RSTP) to communicate among switches. When a failure occurs, the STA is solved to determine how the network should reconfigure, and then RSTP is used to trigger reconfiguration. Parts of the network that are affected by this reconfiguration may be unavailable to deliver packets during the transition or period of network darkness.

The STA chooses to always keep the network in the optimal configuration for message delivery, and when a failure occurs, a new optimal configuration is determined and the network transitions to that new configuration. RSTP, by default, chooses active paths (and, in turn, inactive paths) such that the length of all paths among switches between end devices is minimized and uses the highest-bandwidth links possible. It is possible to control these decisions and force specific paths to be active (and others inactive) if required to satisfy engineering needs. If the failure condition is resolved, either by restoring the link that was lost (link restoration) or replacing or fixing the switch that failed (bridge life), the network will revert to the previous configuration that was optimal according to the STA. This restorative event will also cause brief network darkness for the same sections of the network that experience darkness during the original failure. It is important to physically wire and properly configure the switches in the network to provide the performance required by the application that will use the network. RSTP allows us to control which switch commands the STA of a network by choosing the root bridge using the bridge priority setting. The root bridge of an RSTP-controlled network, often considered the logical center of the network, is very important because all other decisions about active and inactive paths are based on its location. It is recommended that a device with a very high MTBF be chosen for this device and its backup. The backup root is the device that will become the logical center of the network in charge of STA decisions in the event that the root device fails. A root bridge failure is very traumatic to an RSTP network because all path decisions must be recalculated to use the backup root device.

The transmission and transfer time tests were performed by coordinating the change of state with the network failure to confirm typical times. Then the failures were tested separately in an automated fashion to obtain a statistically significant number of samples in order to understand the statistical distribution, mean, median, and standard deviation.

The transit test requires injecting specific test packets and should not be continuously run on in-service systems. However, the transmission and transfer time tests can be performed in a laboratory, during a factory system test, during a site acceptance test, and continuously as a system self-test function. They are performed in devices executing the control and automation applications and in surrogate devices added to the system specifically for test and validation. Once tested for a specific application, IEDs perform similarly in the laboratory and in service. However, the times change with changes in network traffic and path failures. During these tests, the transfer time duration between the synchronized logic IEDs (SLIs) and physical devices (PDs) is simultaneously calculated. Typical transfer time is calculated to be (transmission time – (t_{f2})/2) with an accuracy of 0 + 0.5 operating cycle duration. Additionally, a ping-pong test can be performed, where the second IED reacts by returning a change of state in another GOOSE message and the first IED measures the roundtrip time. Typical transmission time was calculated to be ≤2 milliseconds for SLIs and ≤3 milliseconds for protective relays. In each test, the network behaves the same way and the duration difference is a result of measurement accuracy differences.

For test criteria, we chose to satisfy the mission-critical signal application of direct trip or control with a typical signal transfer time of less than 3 milliseconds per IEC 61850 Type 1A, Performance Class P2/P3. Based on using IEDs

with a 2-millisecond operating time, the typical signal transmission time is less than 5 milliseconds. The maximum signal transmission time of 20 milliseconds was selected to satisfy specific protection and automation schemes [3]. The time delta between maximum and typical transmission times (in this case, 15 milliseconds) becomes the absolute maximum network darkness duration. The true maximum network darkness duration must be short enough to allow a GOOSE transmission to satisfy the maximum signal duration. We configured the IEDs to retransmit at 4, 8, and 16 milliseconds after the initial change-of-state GOOSE message. Therefore, the true and absolute maximum network darkness values are both 15 milliseconds.

If IEDs are not capable of this retransmission pattern, the test may fail. For example, if the retransmission pattern is to transmit at 1, 3, 7, and 14 milliseconds after a change of state, the absolute duration of darkness would overlap all the messages, so the actual maximum would be 13 milliseconds. Alternately, retransmitting at 94, 500, and 1,000 milliseconds after a change of state will never meet the maximum signal transmission time if the initial message is dropped. In this case, if a failure occurs to disrupt delivery of the first message and then network darkness ends within 15 milliseconds, the change of state will not be published to subscribers until 94 milliseconds later.

We also tested a different category of PDs that have a slower operating cycle and do not have count-up timers. Therefore, we used a countdown timer to verify that a GOOSE ping-pong completes within twice the acceptable transfer time. These IEDs also have a 4.17-millisecond operating cycle duration in 60 Hz systems. Therefore, both the timer start and the timer stop in PD1 will each have a typical error of 0 + 2.08 milliseconds, for a total duration calculation typical error of 0 + 4.17 milliseconds. For each of these tests, the change of state is controlled automatically for repetitive testing of large numbers of samples or by a front-panel pushbutton on an IED for in-service samples. Typical roundtrip transmission time was measured to be ≤16 milliseconds, averaging to a one-way transmission time of ≤8 milliseconds within these IEDs.

These same transmission, transfer, and transit time duration tests are performed while injecting additional traffic into the Ethernet network. If the traffic does not travel on the specific perimeter ports, any timing differences are the result of the performance of the channel. When additional traffic is allowed on the perimeter ports, it also affects processing in the IEDs.

5 Ethernet network reconfiguration time testing

Accurate testing of network darkness during a failure or restorative event requires the use of measurement techniques that are analogous to the application of interest. In the case of a critical GOOSE multicast message application, a multicast message test must be used. Using a standard unicast message or a specialized message, such as ping (which is used to test IP network address connectivity), is not appropriate. Signaling network tests must be performed using a multicast message with no IP address, which is the format of the GOOSE message. Using a ping-based tester will not give accurate results for the reconfiguration times of the network for GOOSE message signaling.

Data transit time duration that requires multiple messages is validated with an independent surrogate network darkness test (NDT) device, which publishes messages that mimic the critical application messages at a fixed frequency and monitors their reception. Network darkness that causes dropped packets is observed by counting the number of consecutive undelivered packets. The period of darkness is calculated as the number of packets undelivered due to loss or delay multiplied by the time between publications. For this testing, the NDT device was set to publish a message every 0.25 millisecond.

Darkness measurements indicate the impact of each failure and subsequent reconfiguration to a hot-standby Ethernet network path. These times are then used to calculate the total application impact.

The NDT device automatically controls and measures the network failure event and restoration (both bridge and link failures and restorations) so that a statistically significant number of samples necessary to understand the statistical distribution of each network are measured. These large amounts of accurate data on many different network topologies, the measurement locations on those topologies, and the different failure modes provide necessary network design information. These data about network darkness durations enable analysis for every possible failure scenario of each port pair in the network. With this information, it is possible to find locations in certain topologies that will always satisfy the needs of the application with sufficiently short durations of network darkness during reconfiguration events. It is important to note that some applications consist of numerous signals. Each source and destination port pair must be considered.

6 Ethernet network architectures

Even though STA and RSTP algorithmically enable and disable links in a topology to remove physical loops in the network and minimize the distance between any two points (balance the network), they must operate within the physical wiring of the network. The physical wiring of the network has a large impact on the performance characteristics in terms of reconfiguration and network congestion.

We performed testing and comparisons of ring, dual star, and ladder topologies using RSTP for reconfiguration. These topologies are shown in Fig. 2, Fig. 3, and Fig. 4. These designs use fiber gigabit backbone links instead of copper gigabit ports (copper gigabit ports operate more slowly during reconfiguration). During the testing, we found that the actual behavior of the dual star topology was not appropriate for signaling. This behavior was previously unknown and only came to light as a direct result of this testing. The two

remaining topologies include the ring and ladder. The ladder is so named because the rows of switches look like rungs of a ladder. The ladder performs best, and IEDs are easily dual-connected in failover mode between the two switches on each rung.

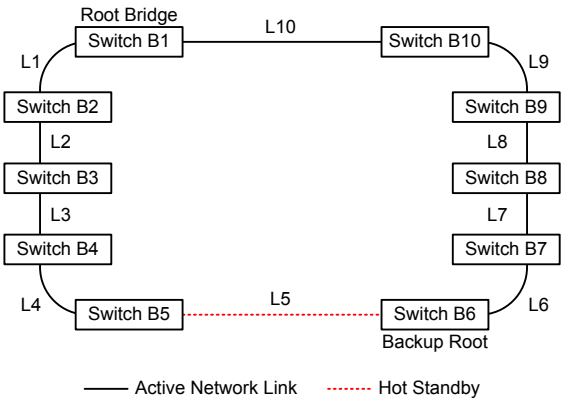

Fig. 2. Ring Ethernet switch topology.

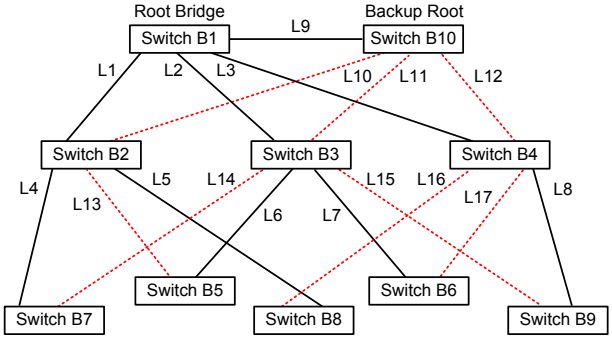

Fig. 3. Dual star topology.

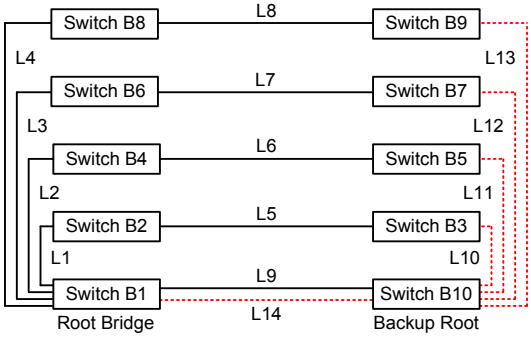

Fig. 4. Ladder Ethernet topology.

Topology	Every Channel Meets <15 ms Maximum Link Loss Recovery Time	Root Bridge Death Typical Reconfiguration Time Is <15 ms	Non-Root Bridge Death Typical Reconfiguration Time Is <15 ms	Network Performance Is Unaffected by Additional Switches	Complexity of Choosing Pair of Perimeter Ports That Will Provide Acceptable Signaling Between IEDs
Ladder	Yes	No	Yes	Yes	Port selection does not matter; all pairs are acceptable
Dual star	No	No	No	Yes	Cannot know behavior in advance; we must test each choice
Ring	No	No	No	No	Cannot know behavior in advance; we must test each choice

Table 2: Results of Ethernet network reconfiguration tests.

As mentioned previously, we selected the network maximum duration of darkness during reconfiguration as 15 milliseconds. Other specific applications need to be tested based on their individual transmission time criteria. Root bridge death is a very troublesome failure because it disrupts the switch commanding the STA and causes extended darkness. Root bridge death was measured separately and, as mentioned previously, should be managed by choosing a very reliable switch to keep the probability of failure to a minimum. Results were gathered for each topology and every failure scenario, and the results are summarized in Table 2.

As mentioned, the results verified unexpected excursions from acceptance criteria in the dual star topology, and therefore, it is not recommended for signaling. This paper presents analysis of Ethernet switches designed and built specifically for GOOSE signaling, and these results do not apply to other switches. Every application has its own failure condition requirements. Thousands of data samples gathered during automated testing revealed that the ladder topology satisfied our criteria of 15-millisecond maximum darkness duration and that the ring topology fell short. If the network darkness requirement was not as restrictive, then the ring topology could be a viable solution as well as other switches less optimized for GOOSE signaling.

Testing confirmed that the ladder topology could guarantee acceptable performance (less than a 15-millisecond reconfiguration), regardless of which non-root pair of switches is selected. This guaranteed performance greatly simplifies the task of cabling among IEDs and switches. Values for link failure ranged from 12.8 to 13.8 milliseconds, and non-root bridge loss was always less than 10 milliseconds. Root bridge death was occasionally measured up to 18 milliseconds. This shows that when using these switches in a ladder configuration, failover is fast enough to always satisfy the most stringent signaling requirements. Redundancy methods like Parallel Redundancy Protocol (PRP) would not increase the reliability of these switches in a ladder topology but may increase the reliability in other designs.

There are many other benefits to using the ladder topology, including the segregation of network traffic, which reduces latency and saturation concerns. The ladder is simply expanded by adding rungs that will never become part of the original and hot-standby paths of the established channels and will therefore not affect channel performance if they experience failure. This cannot be said for other topologies, such as the ring and dual star. Light travels through fiber at about 186,282/1.467 = 124,188 miles per second. Therefore, latency due to message transit through fiber is negligible for cable lengths within a substation. This means that switches in the field can be configured in the ladder topology regardless of their proximity. Because every non-root switch pair is satisfactory, IEDs can be connected to any perimeter port. This strength and others of both the ring and ladder topologies are listed in Table 3. The dual star topology results were so poor, and characteristics so undesirable, that we chose not to continue considering it for networks performing signaling.

Ring Topology	Ladder Topology
Is simple to build.	Is very robust and can handle many failures.
Requires shorter cable runs, which are less expensive.	Has consistent latency in failure conditions.
Has maximum IED-to-switch ratio.	Has consistently small latency.
Only requires two backbone links per switch.	Has very localized network darkness during failure.
	Can scale without affecting performance.
	Has localized traffic on network segments.
	Requires minimum settings changes even for a large network.
	Has very consistent reconfiguration times.
	Provides guaranteed locations on network with good reconfiguration times.

Table 3: Comparisons of strengths of different Ethernet network topologies

When using a ten-node ring topology, there are some switch pair combinations that have adequate performance of less than a 15-millisecond reconfiguration, but many others take minutes. Also, it is not possible to know which switch pair will always experience a less than 15-millisecond darkness duration, so testing is required to confirm channel performance. Once known, appropriate channels are relegated to certain switch combinations in relation to the root bridge. Therefore, this requires that IEDs be connected to specific switches, regardless of their actual physical proximity in the field. Also, as the ring size increases, the network reconfiguration times continue to increase.

7 Conclusion

Simple tools, application and test IEDs, and very specific network test devices play an important role in Ethernet network performance testing. IED features should be deployed for acceptance testing and ongoing monitoring of application behavior, as mentioned in [2]. However, Ethernet network reconfiguration testing requires new special-purpose test devices to verify configuration and performance. These devices must be configurable to use enough resolution and accuracy to measure true performance and automatically trigger link loss and bridge failure to collect statistically meaningful results. Also, they must use appropriate technology to verify network behavior for the specific signal message types, such as multicast GOOSE messages.

Application tests confirmed typical times for an error-free network to be 14-millisecond application, 4-millisecond transmission, and 2-millisecond transfer times. SLIs time-stamp changes and measure these times with an accuracy of + 0 to 0.5 operating cycle duration time. Protective relays time-stamp changes with an accuracy of + 0 to 0.5 operating cycle duration and measure transmission duration with an accuracy of + 0 to 1 operating cycle duration time. These times meet IEC 61850 Type 1A, Performance Class P2/P3.

Reconfiguration tests confirmed that the chosen Ethernet switches, designed specifically for PCM applications, routinely deliver packets with a transit time typically well under 1 millisecond. Network reconfiguration behavior and worst-case transit time depend greatly on the network topology, switch settings, and the design of the switches. Any one of these characteristics can easily mean the difference between meeting the application requirements for critical messaging and failing to do so.

References

[1] S. Chelluri, D. Dolezilek, J. Dearien, and A. Kalra, "Design and Validation Practices for Ethernet Networks to Support Automation and Control Applications," proceedings of the Power and Energy Automation Conference, Spokane, WA, March 2014.

[2] IEC/TR 61850-90-4, Communication Networks and Systems for Power Utility Automation—Part 90-4: Network Engineering Guidelines, Technical Report, August 2013. Available: http://webstore.iec.ch/.

[3] D. Bekker, T. Tibbals, and D. Dolezilek, "Defining and Designing Communications Determinism for Substation Applications," proceedings of the 40th Annual Western Protective Relay Conference, Spokane, WA, October 2013.

© 2015 by Schweitzer Engineering Laboratories, Inc.
All rights reserved.
20151117 • TP6739