CIP Safety Embracing IEC 61784-3 Edition 4

Xiaobo Peng[1], Steven Seidlitz[2], David Crane[3], Arun Guru[4].

[1] Xiaobo Peng, Sr. Safety Architect, Rockwell Automation, Shanghai, China.
[2] Steven Seidlitz, Sr. Project Engineer, Rockwell Automation, Milwaukee, USA
[3]David Crane, Sr. Staff Engineer, ODVA, Michigan, USA
[4] Arun Guru, Principal Engineer, Rockwell Automation, Mequon, USA

Presented at the ODVA
2022 Industry Conference & 21st Annual Meeting
March 9, 2022
San Diego, California, USA

Abstract

Edition 4 of IEC 61784-3 (Functional safety fieldbus - General rules and profile definitions) was recently published in February of 2021. It added several significant enhancements to address Timeliness errors, Authenticity errors and Masquerade errors to accompany the previously considered Data Integrity errors. A new requirement was also included that requires proof of the effectiveness of the safety CRC polynomials used by a protocol, based on its residual error probability calculation. The ODVA SIG for CIP Safety™ has been actively monitoring the development and normalization of the IEC 61784-3 standard. To accommodate these new requirements raised in IEC 61784-3 Edition 4, several updates were made to the CIP Safety specification from Edition 2.20 to Edition 2.22. Most notably, the Base format has been deprecated and the Max_Fault_Number has been adjusted. This white paper will first explain what major changes were introduced in IEC 61784-3 Edition 4 and why. Second, the paper will discuss how the CIP Safety specification has adapted and evolved to maintain compliance. To call the attention of CIP Safety technology adopters, this paper will not only discuss the rationale behind these changes but that going forward how all products applying for an ODVA Declaration of Conformity will be expected to apply these protocol changes to their implementations, and when the required compliance became effective.

Keywords

IEC 61784-3, Base Format, Extended Format, Residual Error Probability, Binary Symmetric Channel, Residual Error Rates, TADI Model, Errors, Maximum Fault Number

Definition of terms

CIP:	Common Industrial Protocol
CPF:	Communication Profile Family
CRC:	Cyclic Redundancy Check
DOC:	Declaration of Conformity
E/E/PE:	Electrical/Electronic/Programmable Electronic
EPI:	Expected Packet Interval
FSCP:	Functional Safety Communication Profile
IACS:	Industrial Automation Control Systems
IEC:	International Electrotechnical Commission
IEC SC65C WG12:	IEC technical subcommittee 65C Working Group 12
NTE:	Network Time Expectation

NTE multiplier:	The maximum number of 128 Microsecond increments that a consumer should allow the age of the safety data to reach.
PFD_{avg}:	Average probability of dangerous Failure on Demand
PFH:	Average frequency of dangerous failure per hour
SIL:	Safety Integrity Level
SPDU:	Safety Protocol Data Unit

IEC 61784-3 overview

As the adoption of digital communication technologies in industrial automation control systems (IACS) evolves, IEC 61158 series together with IEC 61784-1, and IEC 61784-2 define a set of communication protocols which can be used to enable distributed control in industrial applications. Functional safety needs require a reduction in the level of risk in a device or system [1] and hence it is a critical aspect or property that mission critical industrial automation applications should ensure. The most famous IEC 61508 specification series provides functional safety standards for the lifecycle of electrical, electronic, or programmable electronic (E/E/PE) systems and products. To fulfil requirements for applying industrial communication technologies in distributed functional safety applications as shown in Figure 1, functional safety fieldbuses have been invented to ensure that data can be safely transmitted from one networked node to another networked node. Safety communication is treated as part of a safety function and thus contributes to the total safety function PFH or PFD_{avg}. To simplify system safety calculations, it is recommended that any logical connection between safety communication elements of a safety function does not consume more than 1% of the maximum PFH or PFD_{avg} of the target SIL level [2].

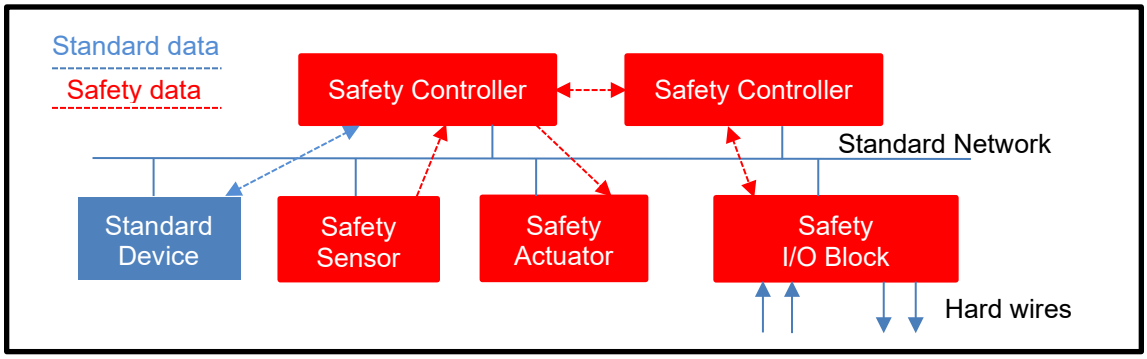

Figure 1: Networked Safety System.

The IEC 61784-3 standard named “Industrial communication networks – Profiles – Part 3: Functional safety fieldbuses – General rules and profile definitions” [2] explains common principles that can be used in exchange of safety-relevant messages between participants within a distributed network in accordance with the requirements of IEC 61508 for functional safety. These principles are based on the *Black Channel* principle, i.e., a communication system containing one or more elements without evidence of design or validation according to IEC 61508. These common principles can be used in various industrial applications such as process control, manufacturing automation and machinery [2]. Under the umbrella of the IEC 61784-3 generic part, a dozen, or so functional safety communication profiles (FSCP) are defined as IEC 61784-3-x to implement safety communication layer (SCL) specifications on top of Data Link Layer or Application Layer of different fieldbuses. These FSCPs are also sometimes mentioned as technology-specific parts of IEC 61784-3. As an example, IEC 61784-3-2 (FSCP 2/1, known as CIP Safety) implements safety communication layer specifications on top of Communication Profile Family 2 (CPF2, known as CIP™) and Family 16 (CPF 16, known as SERCOS™).
IEC 61784-3 categorizes communication errors into corruption, unintended repetition, incorrect sequence, loss, unacceptable delay, insertion, masquerade and addressing errors. Masquerade error happens when a message from non-safety related source is interpreted as if it originated from a valid safety related source. This standard further recommends deterministic remedial measures to these communication errors, including the use of a sequence number, time stamp, time expectation, connection authentication, feedback message, data integrity assurance, redundancy with cross checking and different data integrity

assurance systems. A table of the overview of the effectiveness of the various measures on possible errors is given in IEC 61784-3, as guidance for functional safety communication protocol designers. Importantly IEC 61784-3 also defines requirements and/or models for estimation of total residual error rate, which need to be considered and fulfilled by all FSCPs. Besides models, this standard specifies other requirements for functional safety communications such as installation guidelines, safety manual, safety policy, structure of technology-specific parts, and assessment guideline etc.

Major Changes in IEC 61784-3 Edition 4

IEC SC65C WG12 [3] is the IEC working group creating and maintaining the IEC 61784-3 standard. The first edition of IEC 61784-3 was published in December 2007 and the second edition was published in June 2010. In the first two editions, only data integrity considerations were given to calculation of the residual error rate. As WG12 continued to improve the models in the standard, the third edition was published in May 2016 which introduced “informative” extended models for estimation of the total residual error rate, considering all listed communication errors. The latest one, fourth edition was published in February 2021 which enforces the extended models as “normative”.

Extended models (TADI models)

To introduce the extended models, all listed communication errors in the standard were further grouped into Timeliness error, Authenticity error, Data Integrity error and Masquerade error. As an example, unacceptable delay, unintended repetition, and insertion errors are all treated as kind of Timeliness error. Since Masquerade errors are more likely to be detected with all the timeliness, authenticity and data integrity measures, the extended models are sometimes simply referred as TADI models.
By turning the extended models to be normative, IEC 61784-3 Edition 4 requests that supplier of FSCP should provide proof of a sufficient overall residual error rate considering all these errors [2].
IEC 61784-3 Edition 4 also gives example equations for the calculation of residual error rates for explicit FSCP category, with contribution of residual error rates of data integrity errors (RR_I), authenticity errors (RR_A), timeliness errors (RR_T) and masquerade errors (RR_M). The total residual error rate can be based on the summation of the four residual error rates RR_I, RR_A, RR_T and RR_M, or it can be based on other quantitative proofs.

Effectiveness of CRC polynomials

IEC 61784-3 explains safety communication channel model using CRC-based error checking, which is based on binary symmetric channel (BSC). The residual error probability which is based on the detection using a CRC-mechanism for BSC can be calculated as

$$R_{CRC}(P_e) = \sum_{i=1}^{n} A_i \times P_e^i \times (1 - P_e)^{n-i} \quad \text{…Equation (1)}$$

Where A_i is the distribution factor of the code, n is the number of bits in block and P_e is the bit error probability [2].
As indicated by the Equation (1), exploration of A_i is critical for R_{CRC} calculation. Other residual error probability calculation algorithms such as dual code method are also applicable. Based on the observation of no conservative approximation formulas, IEC 61784-3 Edition 4 has an additional requirement to explicitly calculate R_{CRC} for the selected generator polynomial over all values of n in use and all relevant values of P_e.
To verify the correctness of CRC calculation algorithms for R_{CRC}, Edition 4 also includes tables providing reference data for verification of CRC calculation algorithms. The CRC calculation algorithm to be verified can calculate the residual error probabilities of CRC polynomial 0x14eab or 0x1f1922815 with given data lengths and bit error probabilities and then compare the calculation results with the reference data given in the standard.

Transition to new editions

Because product's assessment process could have a lengthy duration, FSCPs published prior to or concurrently with new edition of the generic part can only be assessed using the methods from previous editions [2]. A transition strategy was hence provided in Edition 4. Basically, the philosophy is that critical new requirements come to the standard as informative first and then become normative in next edition. FSCPs are given a caution and a time window (roughly 4 to 5 years) to be prepared for the upcoming new normative requirements in next edition. Such FSCP preparations happen during an IEC 61784-3 validity period, which is the time duration between publication of one edition to next edition. The TADI models were informative in generic part Edition 3 so FSCPs were expected to update specifications based on TADI models during Edition 3 validity period. Upon the publication of Edition 4 of the generic part, FSCPs should be assessed using the methods from Edition 4.
As the communications technologies keep evolving (e.g., Gigabit Ethernet, wireless communication, cyber security), more error patterns are being considered and evaluated by IEC SC65C WG12 [3]. It is foreseeable that additional new requirements such as consideration of uniformly distributed segment errors will be added to IEC 61784-3 future editions, and similar transition pattern will be needed for FSCPs to comply with upcoming new editions.

Description of Changes to CIP Safety

There have been a few changes to the CIP Safety specification to accommodate the recent updates to the IEC 61784-3 standard as discussed in the previous section. In summary, Changes include deprecation of the Base Format and setting the Maximum Fault Number (Max_Fault_Number) of the Extended Format to a fixed value of 2. This section will describe the rationale behind these changes to the CIP Safety Specification.

First change to the CIP Safety Specification is that the Base Format option can no longer be used and is no longer supported within the ODVA suite of conformance tests. The driver for this change is the inclusion of Timeliness errors within the industrial safety communication error model as newly added to Edition 4 of IEC 61784-3. The PFH for an industrial safety communication protocol must be no larger than 10^{-9} failures per hour, which is 1% of the SIL3 budget, or 1% of 10^{-7} failures per hour. However, when Timeliness errors are included in the error model, the resultant PFH for the Base Format of CIP Safety as shown in Figure 2, both short and long form, is much larger than this threshold. Therefore, this makes the Base Format no longer compliant with Edition 4, and hence requiring it to be deprecated. Note: This is not the case for the Extended Format which uses more bits to represent a Time Stamp.

To understand why this is the case for the Base Format, let's examine the three most dominating factors in the Residual Error Rate (RR) of Timeliness of the data for the Base Format data section of the CIP Safety protocol. Please note that not all the detection mechanisms for CIP Safety are illustrated in the equation (Equation 2) below but just the two most dominating detection mechanisms. Other terms related to the check of the Run_Idle bit and the check of the Ping Count are excluded here for the purposes of brevity as they are not dominating factors.

$$RR(Timeliness\_DataSection) \approx R(T)_{Data} \cdot \left[1 + RP\left(CRC_{Timestamp}\right)\right] \cdot RP(NTE_{check}) \qquad \text{... Equation (2)}$$

where:

RR = Residual Error Rate (faults per hour)
$R(T)_{Data}$ = Timeliness Error Rate (per hour)
RP = Residual Error Probability
NTE = Network Time Expectation

This equation is not all that intuitive; so, let's unpack it one component at a time. Starting with the first term $\boldsymbol{R(T)_{Data}}$, this is the error rate of Timeliness errors in units of faults per hour. This is how often Timeliness error can occur in a system, regardless of whether they can be detected or not. Practically speaking, Timeliness errors represent communication messages that are old and stale such that their corresponding safety data is potentially not fresh and if applied to a safety function could cause indeterminate and potentially unsafe behavior for a system. This can realistically happen in a Black Channel network device that employs store-and-forward behavior such that an old packet may be cached

and potentially transmitted again onto the network and sent to a safety consuming device. Once a Timeliness error occurs, then the next terms determine the likelihood that they will not be detected. For this reason, the three terms are multiplicative in the Equation 2 above.

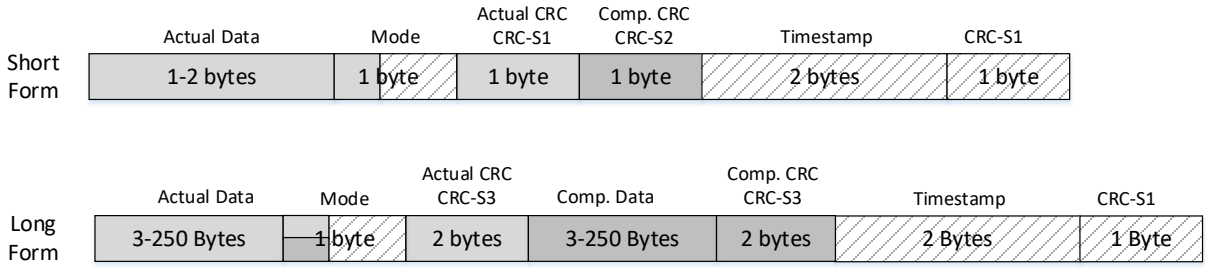

Figure 2: Base Format Messages.

However before delving deep into the two RP terms in the Equation 2 above, it is necessary to first review the Base Format protocol as that plays an integral part in understanding its limitations with respect to Edition 4. The Figure 2 above illustrates the layout of the safety protocol data unit, SPDU, for each form of the Base Format. Both Short Form and Long Form employ an 8-bit CRC across the Timestamp field. The second term $\boldsymbol{RP}(\boldsymbol{CRC_{Timestamp}})$ makes use of that CRC field. It represents the probability that data integrity errors within the Time Stamp field will not be detected by the corresponding 8-bit CRC across it. So, if a Timeliness error occurs at an error rate of $\boldsymbol{R}(\boldsymbol{T})_{\boldsymbol{Data}}$, this measure becomes the first line of defense in detecting those errors. Its effectiveness is directly related to the strength and properness of the corresponding CRC polynomial and the number of bits of that CRC. Mathematically, this is represented by the value of 2^{-r}, where r is the number of bits of the CRC. For the Base Format in both forms, r has a length of 8 bits. Therefore, the probability that the CRC will not detect a fault in the Time Stamp field is 2^{-8}.

From a practical standpoint, if a message containing stale or inaccurate safety data had reached a safety consumer in an end device, this represents the probability that:

a) The Time Stamp may have been corrupted such that its value yet falls within the proper time window when otherwise it would not have

   -and-

b) The CRC check had failed to detect this changed Time Stamp value up to a small probability of $2^{-8} = 3.9 \times 10^{-3}$.

In this scenario, since the CRC didn't detect the corruption up to said probability and since the data age check of the Time Stamp would not be able to detect the message being old and stale because the Time Stamp happened to fall within the valid data age window, then this leads to a potentially unsafe condition. Old and stale safety data would be applied within the safety function of a system, thereby, potentially causing indeterminate behavior in the system and safety actuator(s).

The third term is the Network Time Expectation check, $\boldsymbol{RP}(\boldsymbol{NTE_{check}})$**.** This is the second layer of defense in detecting Timeliness errors after the Time Stamp CRC check. As it turns out, this is the primary factor for why the Base Format cannot achieve a resultant PFH value of under 10^{-9}. In short, this is due to the size of the Time Stamp field and not being able to detect Time Stamp rollovers. This term represents the probability that an invalid or stale packet will pass the Network Time Expectation check. The CIP Safety specification defines the Network Time Expectation as **"**the worst-case time from a safety related event occurring as input to a Safety Producer or as a fault within the Safety Producer until the output of the Safety Consumer is put into the safety state**".** There is a small chance that a Time Stamp of a stale message could be wrongly considered valid, if the time value just happens to fall within the valid window either through data corruption that is undetected by the CRC check or if the CRC check does pass but the

Time Stamp value happens to be so old that the NTE check evaluates the data age associated with the Time Stamp to yet be valid. That's why the second term in Equation 2 contains a "1 +" factor. The Time Stamp could be corrupted, or it may not be but rather just old enough such that the data age check passes because of a rollover. The old age is not attributable to data corruption or from random hardware faults but from systematic anomalies such as retransmission of cached messages. Refer to the Figure 3 below. This can occur if the elapsed time of when an old message is retransmitted is an integral multiple of the rollover time. In simple terms the previous time could be so old that it looks current according to the safety time domain at the safety consumer. This endpoint would not be able to decipher if the Time Stamp is truly within the NTE window or so old that it looks current now because rollover is not considered. This is even further complicated by the fact that the rollover time is not very long when considering contemporary network devices or even a large number of nodes in a network. Therefore, the size of the Time Stamp field is of the utmost importance, and 16-bits is no longer considered sufficient. If more bits were used than 16-bits, the rollover time would be much larger, and the statistical window of overlap will be greatly reduced.

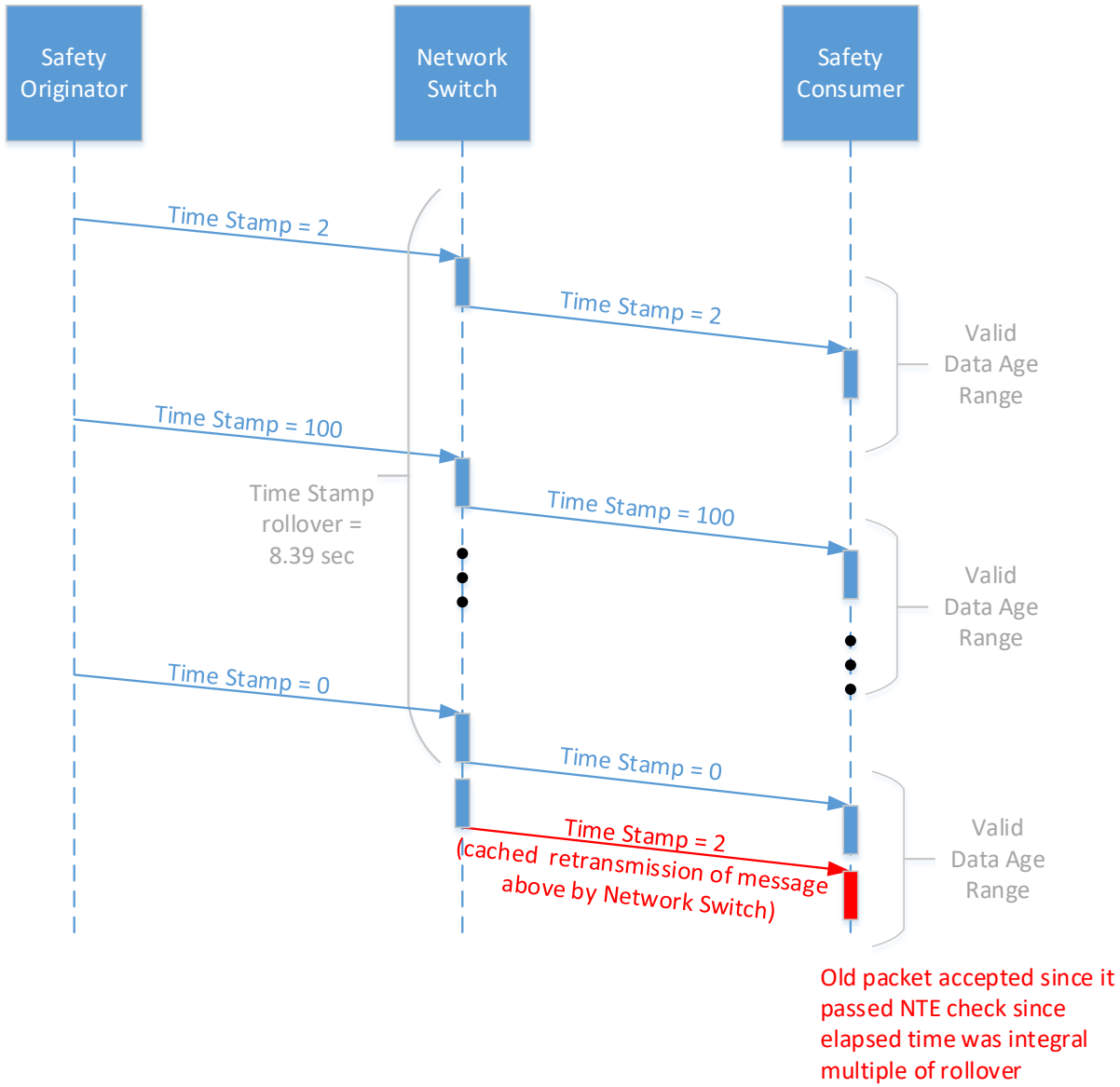

Figure 3: Example sequence diagram of a stale message.

The $RP(NTE_{check})$ term is mathematically defined by the ratio of the NTE value over the Time Stamp rollover time, as shown by the equation below. It is the likelihood that any random number across a 16-bit dataset of the Time Stamp will be contained in the dataset of the NTE range of valid values.

$$RP(NTE_{check}) = \frac{NTE}{Timestamp\ Rollover\ Time}$$

$$RP(NTE_{check}) = \frac{EPI \cdot [NTE_{Multipler} + 2]}{Timestamp\ Rollover\ Time}$$

What this all means is if a large EPI value is selected by an end user, the NTE value will be quite large, especially in comparison to the Time Stamp rollover time. Given the Time Stamp field of the Base Format has a length of 16 bits with no provisions for counting the number of rollovers and each bit count corresponds to a time increment tick of 128us, this translates to a Time Stamp rollover time of (128us * 2^{16}) = 8.39 seconds. That's not a very large rollover time when considering contemporary store-and-forward network routers in the marketplace. The probability that the NTE_{check} does not detect a Timeliness error is based on how big the NTE window is in comparison to the time it takes for the Time Stamp to rollover, which in this case is 8.39 seconds. The higher the NTE window for a safety connection, the higher the residual error probability becomes for CIP Safety. When calculating the final residual error rate for the Base Format, this term is just too high of a probability to overcome to stay under the **1 FIT (10^{-9} per hour) threshold without making format changes, and therefore is not compliant with Edition 4 changes to the IEC 61784-3.**

It is important, though, to keep in mind that when the Base Format was originally developed for CIP Safety, it was targeted for DeviceNet® networks of which store-and-forward routers and switches are not used or needed in its network topology. Delayed packets on the order of a Time Stamp rollover or longer was not a real-world concern that needed to be mitigated by using a larger Time Stamp field. Rather, 16-bits was deemed sufficient at the time. When the Extended Format was introduced in the year 2007, it was designed with EtherNet/IP™ in mind, where such types of network switches and behaviors are commonplace.

To overcome this Timeliness limitation, the Extended Format implements a 16-bit rollover count in addition to the 16-bit Time Stamp. Although the rollover count is not implemented explicitly in the SPDU definition of the Extended Format to form a true 32-bit Time Stamp field, the rollover count is included in the 24-bit CRC calculations as shown in the figure below for both the Short Form and Long Form. This rollover count assists significantly to achieve a resultant residual error rate below the 1 FIT threshold.

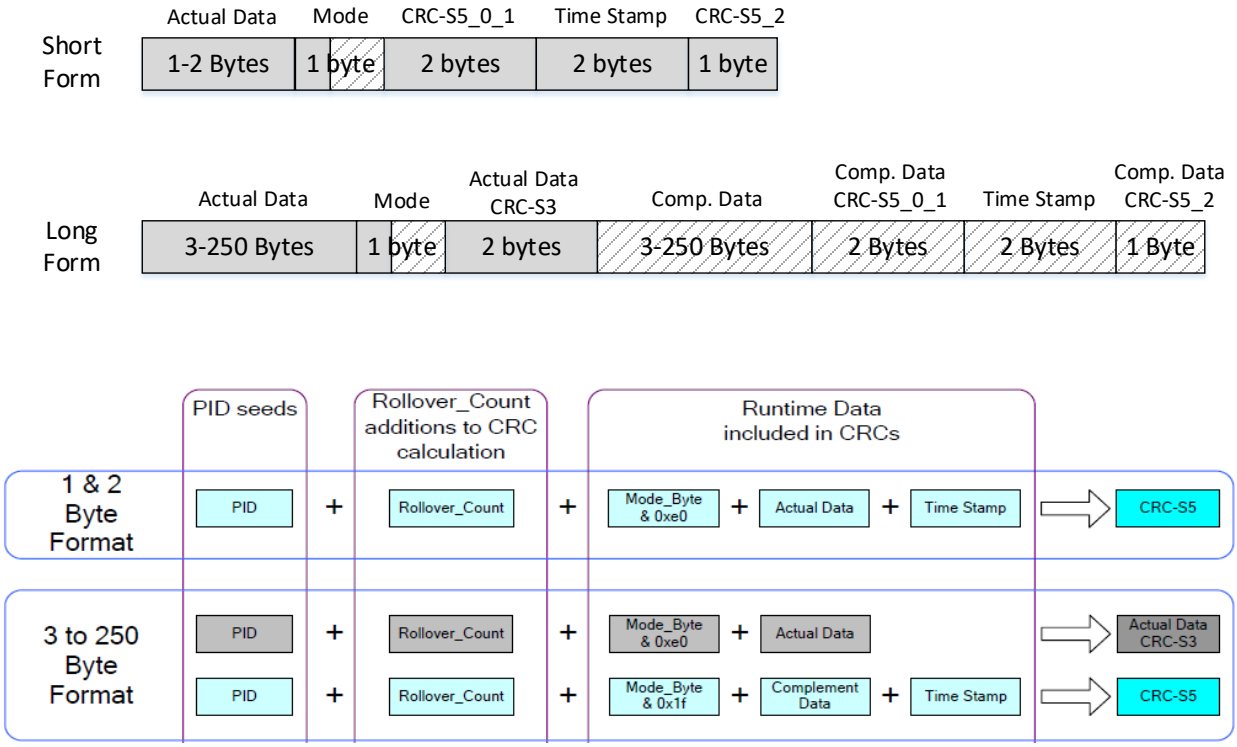

Figure 4: Extended Format Messages.

However, in and of itself, the rollover count is not enough to meet SIL3 for the Extended Format. There was one additional change needed for it to be compliant to Edition 4 of IEC 61784-3. The Extended Format provides for a certain amount of error tolerance through a parameter called the Maximum Fault Number (Max_Fault_Number). It had previously defaulted to a value of 5. This means that the safety protocol and state machine will allow up to 5 detectable safety layer errors per hour per safety connection before a safe-state action is taken such as closing the corresponding safety connection. This parameter was variable and set during the connection establishment process. For example, this parameter is included in the Type 1 Safety Open message. When considering the new changes from Edition 4 of IEC 61784-3 in relation to the Extended Format, Timeliness errors also became the same dominating factor as was the case for the Base Format. Unlike the Base Format, though, the residual error rate for the Extended Format can be made to stay under the threshold of 10^{-9} per hour. The format did not need to be completely deprecated but rather slightly modified. To stay under this probability error rate threshold the Maximum Fault Number (Max_Fault_Number) is now fixed to a constant value of 2 and is no longer variable from this point forward. Prior to the Edition 4, there was a large enough statistical margin with the residual error rate to support this parameter being adjustable. But, when considering Timeliness errors in the error model, this is no longer the case.

Since Maximum Fault Number is a parameter included in a safety open message, originators only need to change their Max_Fault_Number value to this fixed value of 2.. Originators will be checked to verify that they do not send any value other than 2 for Max_Fault_Number parameter in the Safety Open. This check has now become part of the ODVA conformance test. Similarly, a safety open message establishing up a connection that uses the Base Format must also be rejected by a safety target device.

Impact to Conformance

The changes to the specification described above impact the ODVA conformance test in the following ways:

- Devices with safety originator functionality will be checked for correct setting of the Maximum Fault Number (Max_Fault_Number) parameter. The evaluation of this behavior may require review of accompanying documentation and configuration software tools.

- Devices with safety target functionality will be checked for rejection of Base Format connection requests. Eligible target-only devices still supporting base format and seeking to maintain a previously issued DOC (Declaration of Conformity) may use the Amended DOC test process according to ODVA policy in Pub 8 and Pub 261.

The grace period provided for gradual adoption of these requirements ended on January 1, 2022. Since that time, target-only devices which have not deprecated the Base Format cannot obtain a new Declaration of Conformity from ODVA.

Summary

This paper described changes in the IEC 61784-3 Edition 4 which had an impact on the CIP Safety protocol as a FSCP when it became normative in February 2021. It described in detail the impact of changes in the total residual error rate calculations for both Base and Extended Formats of CIP Safety messages while considering those additional errors in communications channels as required by the Edition 4 of IEC61784-3. It explained rationale for deprecating Base Format and change in the Max_Fault_Number parameter value.

Installed legacy products will continue to be covered by their certifications accorded by certifying bodies. However, when significant modifications to an installed safety system are done, this modified safety system would need to comply with the most current version of the standards. Product owners and end users should consult their certifying bodies for specific guidance. Also note that support for Base Format will continue to exist in the originator devices in order to support legacy installations.

These changes along with specification enhancements have been released in Edition 2.22 of CIP Safety Volume 5 [4]. A grace of period of one year till January 2022 was granted to vendors to allow them to make these changes if needed in safety devices. Conformance testing will be used to govern these changes in the CIP Safety specification as new devices undergo conformance certification testing beyond January 2022.

References

[1] Safety and functional safety, https://www.iec.ch/safety
[2] IEC 61784-3, https://webstore.iec.ch/publication/62095
[3] IEC SC65C WG 12,
https://www.iec.ch/dyn/www/f?p=103:14:13271249644174::::FSP_ORG_ID,FSP_LANG_ID:2583,25#
[4] CIP Safety Volume 5, Edition 2.22