40GbE MTTFPA when using transcoding

Steve Trowbridge, Alcatel-Lucent
Osamu Ishida, NTT

IEEE P802.3ba Task Force
Orlando, FL
March 2008

Purpose of This Presentation

- Potential issue with appropriate support for OTN
  - 512b/513b transcoding could degrade MTTFPA
    - to the order of BER^1 (< lifetime of the universe)
    - far below the order of BER^3 in native 64b/66b
- Proposed improvements for MTTFPA
  - Alternative transcodings increase MTTFPA
    - to the order of BER^2 (> lifetime of the universe)
  - ITU-T experts will meet in June to consider proposals
    - IEEE input welcome
- Identify minimal MTTFPA related OTN support requirements to IEEE802.3ba

Potential “OTN compatibility” Issue

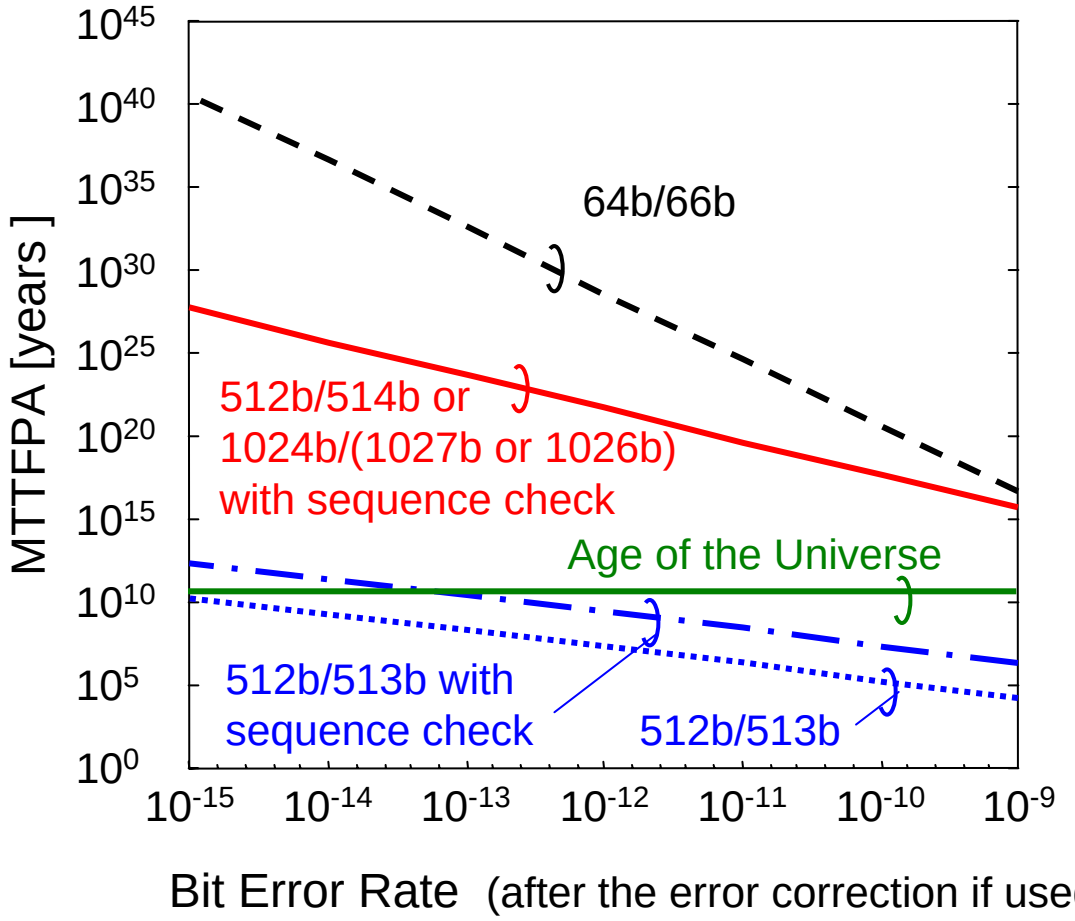
- Native 64b/66b provides excellent MTTFPA
  - of the order of BER^3
  - more than the age of the universe
- While transcoding is normally used into an FEC protected environment of OTN, MTTFPA could be degraded given uncorrected or miscorrected errors in OTN
  - to the order of BER^1 when using 512b/513b
  - less than the age of the universe

Remedy for MTTFPA

- During ITU-T Q11/SG15 February meeting, alternative transcoding proposals appear that improve MTTFPA
  - to the order of BER^2
  - bellow performance of native 64b/66b slightly, but
  - well more than the age of the universe (See next slide)

- Alternative transcoding proposals (See Appendix)
  - 512b/514b (Remedy 1) and 1024b/1027b (Remedy 2)
    - NTT, “40 GbE MTTFPA when using 512B/513B transcoding ”, C786, ITU-T Q11/15, Feb 2008, Geneva
  - 1024b/1026b (Remedy 3)
    - Alcatel Lucent, “Mapping of 40 GbE into OTN and MTTFPA”, C892, ITU-T Q11/15, Feb 2008, Geneva

Estimated MTTFPA



NTT, “40 GbE MTTFPA when using 512B/513B transcoding ”, C786, ITU-T Q11/15, Feb 2008, Geneva

ITU-T Will Discuss Pros and Cons in June

Transcoding	512B/513B	512B/514B	1024B/1027B	1024B/1026B
Proposal *in ITU-T Q11/SG15*	Alcatel-Lucent *WD12,Q11/15 Oct 2007*	NTT *COM 15 – C 786 – E*	NTT *COM 15 – C 786 – E*	Alcatel-Lucent *COM 15 – C 892 – E*
Features				
MTTFPA at link BER of 10^-12	NO 10^7.4 Years	YES 10^22 Years	YES 10^22 Years	YES 10^22 Years
Self-synchronization	NO	YES	YES	YES
Coding delay	512 bits	512 bits	1024 bits	1024 bits
OTN mapping				
Bit rate for 40GbE	40.078 Gb/s	40.156 Gb/s	40.117 Gb/s	40.078 Gb/s
40GbE mapping into OPU3	YES	YES*1	YES	YES
Full Ethernet symbol transmission	YES	NO	YES	YES
Client clock transmission	YES	YES*2	YES	YES
Comments		*1: IDLE shrinkage of 100ppm and using ODU3 NJO (1byte) *2: Constant rate IDLE shrinkage and insertion		

MTTFPA OTN support Requirements

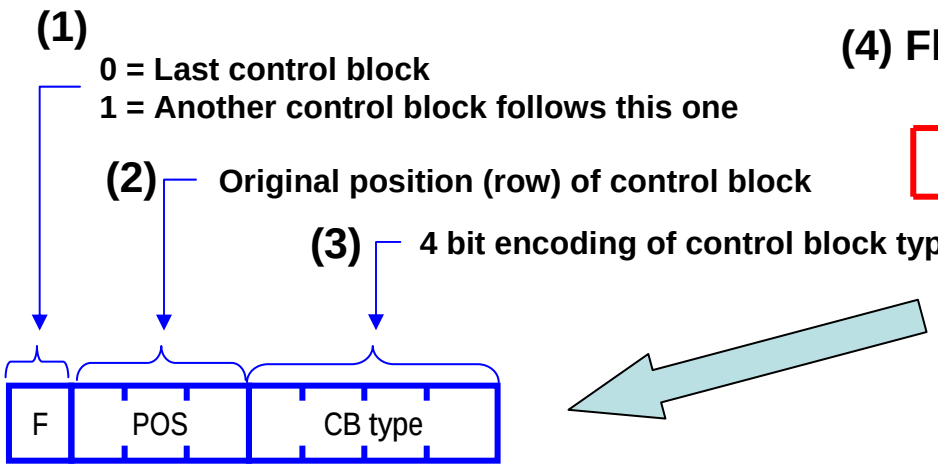
- Permit MTTFPA of the order of BER^2
  - Below performance of native 64b/66b, but
  - still greater than lifetime of universe
- If with 512b/514b (Remedy 1)
  - No extension of 64b/66b control code
  - Operate at lower end of +/- 100ppm clock tolerance range
- If with 1024b/1027b (Remedy 2)
  - No extension of 64b/66b control code
- If with 1024b/1026b (Remedy 3)
  - No extension of 64b/66b control code
  - Only start packets and ordered sets on 8B boundary

Appendix

Alternative Transcoding Details

- 512b/513b and Its MTTFPA Issues
- What happens if a bit error hits ….?
  - (1) Flag Bit for 512b Payload
    - requires 512b/514b (Remedy 1), or
    - 1024b/1027b (Remedy 2), or
    - 1024b/1026b (Remedy 3)
  - (2) 1-Bit Control-Block Continuation Flag
    - requires Block Sequence Check
  - (3) 3-bit original position of control block
    - (not a problem)
  - (4) 4-bit representation of 66B control block type
    - requires re-mapping the control block type

512b/513b and Its MTTFPA Issues



Control block header byte

"Proposed mapping of 40 Gb Ethernet into standard ODU3 ", (Alcatel Lucent) wd12, ITU-T Q11/15, Oct 2007, Shenzhen

(4) Flag bit 512 bit payload (8 x 66b block)

1	1aaa TTTT	7 byte (per format of 64B/66B control block type)						
	1bbb TTTT	7 byte (per format of 64B/66B control block type)						
	0ccc TTTT	7 byte (per format of 64B/66B control block type)						
D0	D1	D2	D3	D4	D5	D6	D7	
D0	D1	D2	D3	D4	D5	D6	D7	
D0	D1	D2	D3	D4	D5	D6	D7	
D0	D1	D2	D3	D4	D5	D6	D7	
D0	D1	D2	D3	D4	D5	D6	D7	

- What happens if a bit error hits the following bits?
  - A flag bit for 512b payload **(1)**
  - A control-block header byte
    - 1-bit continuation flag (F) **(2)**
    - 3-bit original position of control block (POS) **(3)**
    - 4-bit representation of 66B control block type (CB) **(4)**

(1) Flag Bit for 512b Payload

- If a bit error hits this 1-bit header of the 512b block that includes 66B control blocks, all the 66B blocks are misinterpreted as data
- This is sometimes critical, since the minimum inter-frame spacing is far less than 512b, and hence two consecutive MAC frames may be mis-integrated and still pass the CRC check.
- Our recommendations
  - Remedy 1: Add one more bit for the flag; i.e. 512b/514b
    - Like 64b/66b, two-bit header also provides capabilities for block synchronization and bit-error monitoring
  - Remedy 2: Concatenate two 512b/513b blocks and add one more bit for the flag; i.e. 1024b/1027b
    - Allow ODU3 mapping without any IDLE shrinkage
  - Remedy 3: Increase the transcoding compression to 1024b/1025b per malpass_03_0907, then add one more bit for the flag; i.e. 1024b/1026b

Remedy 1: 512b/514b

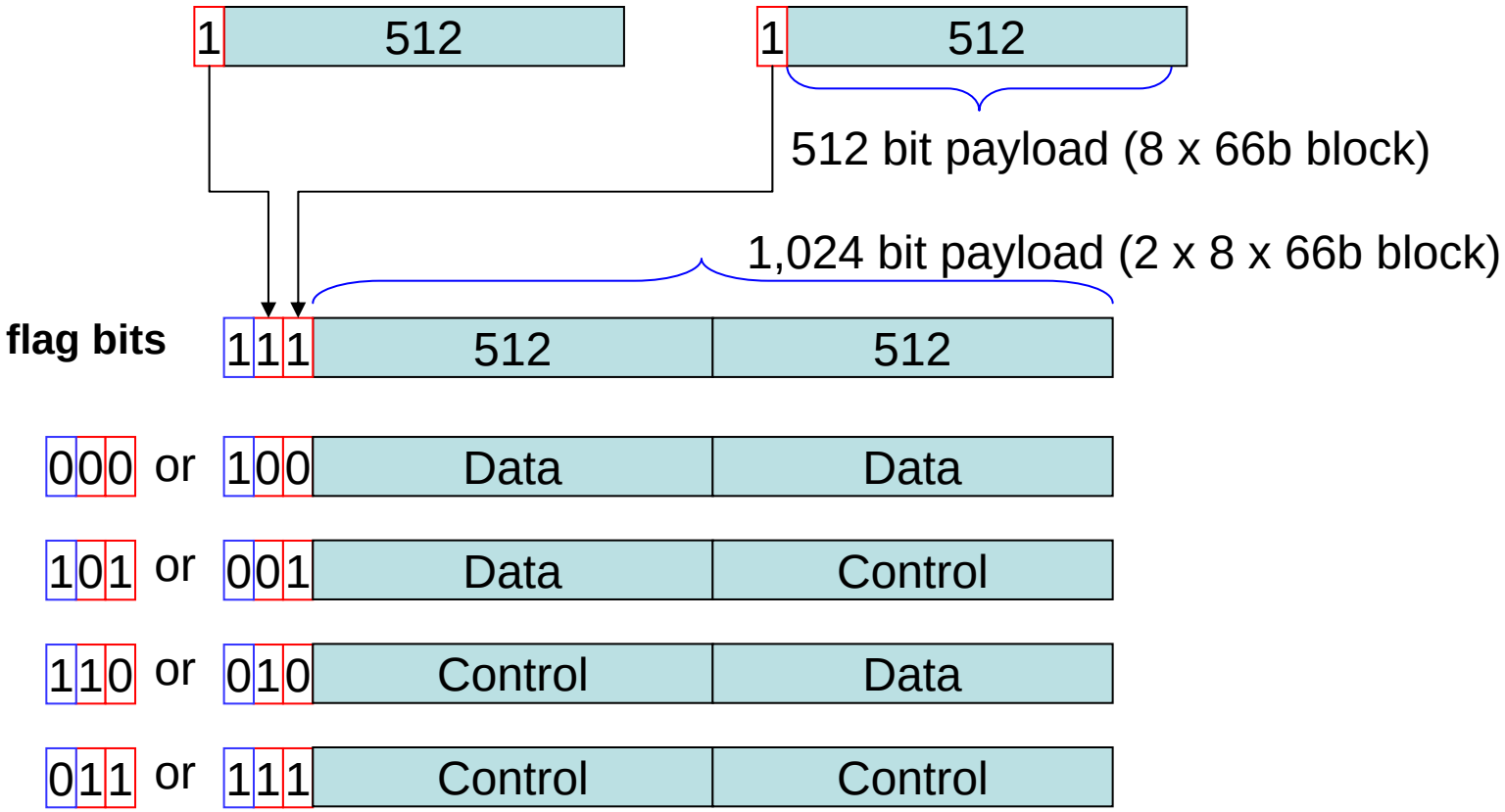
- Nominal rate of 512b/514b is slightly higher than that of OPU3 payload area, but can be accommodated within 100ppm clock tolerance margin of Ethernet
  - by removing some of inter frame idles

40 Gigabit Ethernet	LAN PMD	41.25 Gb/s (4x10.3125 Gb/s)
	Effective rate	40 Gb/s
512B/514B transcoding	Nominal rate	40.15625 Gb/s
	Removal allowance clock tolerance of -100 ppm	**40.15223438 Gb/s**
OPU3 payload area	Nominal payload rate	40.150519322 Gb/s
	Maximum payload rate with NJO byte	40.15315525 Gb/s
	Maximum payload rate under worst case clock tolerance -20 ppm	**40.15235219 Gb/s**

40.15235219 Gb/s > 40.15223438 Gb/s

Remedy 2: 1024b/1027b

- Concatenate two 512b/513b blocks and add one more flag bit
  - Provide two-bit Hamming distance between 3-bit flags



Remedy 2: 1024b/1027b on ODU3 ?

- Yes, nominal rate of 1024b/1027b is lower than that of OPU3 payload area!

40 Gigabit Ethernet	LAN PMD	41.25 Gb/s (4x10.3125 Gb/s)
	Effective rate	40 Gb/s
1024B/1027B transcoding	Nominal rate	40.1171875 Gb/s
	Maximum rate under worst case clock tolerance of + 100 ppm	**40.1211992187 5 Gb/s**
OPU3 payload area	Nominal payload rate	40.150519322 Gb/s
	Minimum payload rate under worst case clock tolerance of -20ppm	**40.1497163116135 6 Gb/s**

40.1211992187 5 Gb/s < 40.1497163116135 6 Gb/s

Remedy 3: Additional compression to allow more robust code to fit standard ODU3

- Huawei proposed generalization of 512B/513B to (64*N)B/(64*N+1)B to enable more compact coding (malpass_03_0907.pdf)
- More compact code than 512B/513B not needed just to get 40 GbE to fit standard ODU3, and not generally desirable as larger block sizes for transcoding increase latency
- “One notch” more compact coding to apply MTTFPA improvements and still fit standard ODU3.
  - Doubling of Flag bit to two bits (01 or 10) improves MTTFPA and provides additional benefit that it is “self framing”
  - Use 1024B/1026B code (sixteen 66B input blocks per 1026B block), which has same efficiency as 512B/513B
  - Four bit position indicator (instead of 3) for control block position
  - Use variable length control block type, taking advantage of unused bits in certain control block formats and 66B control block types not needed given 8-byte packet start and ordered set position rules with 2-bit Hamming distance between control block types with packet terminate /T/
- **Since 1024B/1026B is the same ratio as 512B/513B, it fits ODU3**

Possible control block reductions for 40/100 GbE
From MLD proposal

Ordered sets can't start in 5th lane

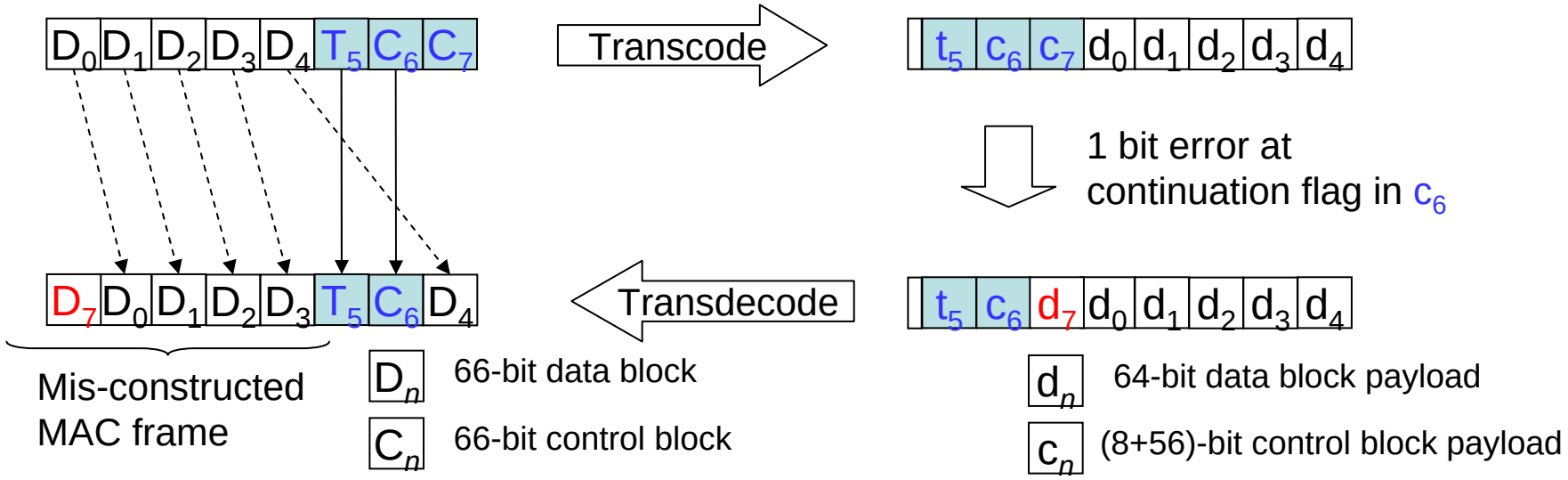
Packets can't start in 5th lane

Input Data	Sync	Block Payload							
Bit Position: / Data Block Format:	0 1	2							65
$D_0\,D_1\,D_2\,D_3/D_4\,D_5\,D_6\,D_7$	01	D_0	D_1	D_2	D_3	D_4	D_5	D_6	D_7
Control Block Formats:		Block Type Field							
$C_0\,C_1\,C_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0x1e	C_0	C_1	C_2	C_3	C_4	C_5	C_6	C_7
$C_0\,C_1\,C_2\,C_3/O_4\,D_5\,D_6\,D_7$	10	0x2d	C_0	C_1	C_2	C_3	O_4	D_5	D_6	D_7
$C_0\,C_1\,C_2\,C_3/S_4\,D_5\,D_6\,D_7$	10	0x33	C_0	C_1	C_2	C_3		D_5	D_6	D_7
$O_0\,D_1\,D_2\,D_3/S_4\,D_5\,D_6\,D_7$	10	0x66	D_1	D_2	D_3	O_0		D_5	D_6	D_7
$O_0\,D_1\,D_2\,D_3/O_4\,D_5\,D_6\,D_7$	10	0x55	D_1	D_2	D_3	O_0	O_4	D_5	D_6	D_7
$S_0\,D_1\,D_2\,D_3/D_4\,D_5\,D_6\,D_7$	10	0x78	D_1	D_2	D_3	D_4		D_5	D_6	D_7
$O_0\,D_1\,D_2\,D_3/C_4\,C_5\,C_6\,C_7$	10	0x4b	D_1	D_2	D_3	O_0	C_4	C_5	C_6	C_7
$T_0\,C_1\,C_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0x87		C_1	C_2	C_3	C_4	C_5	C_6	C_7
$D_0\,T_1\,C_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0x99	D_0		C_2	C_3	C_4	C_5	C_6	C_7
$D_0\,D_1\,T_2\,C_3/C_4\,C_5\,C_6\,C_7$	10	0xaa	D_0	D_1		C_3	C_4	C_5	C_6	C_7
$D_0\,D_1\,D_2\,T_3/C_4\,C_5\,C_6\,C_7$	10	0xb4	D_0	D_1	D_2		C_4	C_5	C_6	C_7
$D_0\,D_1\,D_2\,D_3/T_4\,C_5\,C_6\,C_7$	10	0xcc	D_0	D_1	D_2	D_3		C_5	C_6	C_7
$D_0\,D_1\,D_2\,D_3/D_4\,T_5\,C_6\,C_7$	10	0xd2	D_0	D_1	D_2	D_3	D_4		C_6	C_7
$D_0\,D_1\,D_2\,D_3/D_4\,D_5\,T_6\,C_7$	10	0xe1	D_0	D_1	D_2	D_3	D_4	D_5		C_7
$D_0\,D_1\,D_2\,D_3/D_4\,D_5\,D_6\,T_7$	10	0xff	D_0	D_1	D_2	D_3	D_4	D_5	D_6	

Figure 49–7—64B/66B block formats

(2) 1-Bit Control-Block Continuation Flag

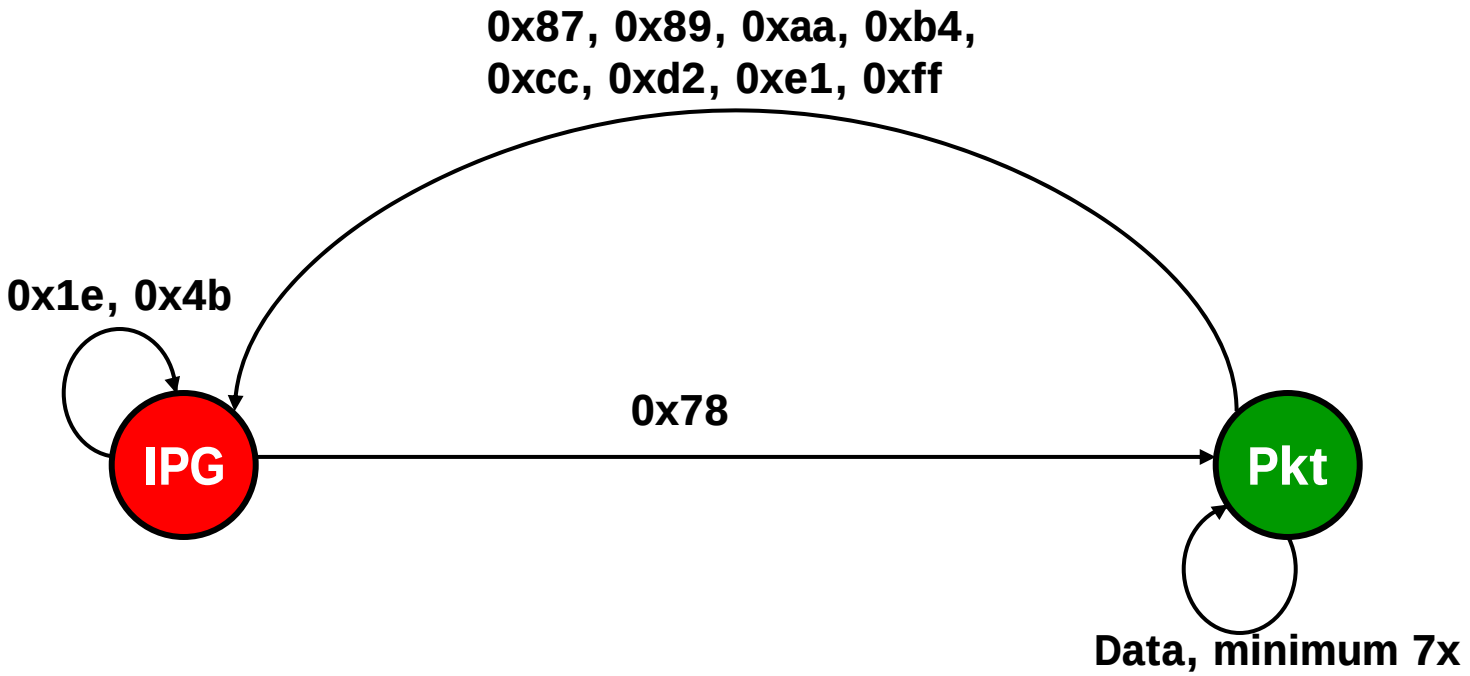
- If a bit error hits this flag, the following 66B control blocks are all misinterpreted as data blocks
- This is critical, since some of 66B control blocks may be misinterpreted as data, and a MAC frame may be mis-constructed and still pass the CRC check.



- Our recommendation
  - Check 66B block sequence at the decoder
  - Replace the 512-bit payload with eight 66B Error Control Blocks if the sequence violation is detected

Block Sequence Check

Check #1 – Block sequence needs to make sense



Check #2 – Control blocks must be in ascending and non-overlapping order of position

Probability negligible that a start or end of packet moves and doesn't appear out of sequence with other blocks

(3) 3-bit Original Position of Control Block

- If a bit error hits these bits, the 66B control block is misplaced in the 512-bit payload, yielding misplaced 66B data blocks
- This is not critical, since such wrong 66B block sequence will be detected at 64/ 66 PCS

- Not a Problem

(4) 4-bit representation of control block type

- If a bit error hits these bits, the 66B control block is misinterpreted
- This is critical, especially if the frame delimiter block T or S is misinterpreted, MAC frame may be stretched or shrunk by several bytes and still pass the CRC check
  - 64B/66B avoids this issue by selecting type codes with Hamming distance of 4
    - 512b/513b can not fully utilize this since we steal 4-bit from this type field, but ….
- Our recommendation
  - Re-map the 4-bit control block (CB) type to keep the Hamming distance of 2 among CB-types for T and S
    - Rearrangement of code values for flag bit Solution 1 or 2
    - Take advantage of reduced set of control block types for 40 GbE and 100 GbE for flag bit solution 3

Control Block (CB) Type Update Example

Flag bit for Remedy 1 or 2

#	64b/66b Block: Control Block Formats	64b/66b Block: Type Field	CB Type
1	$C_0\ C_1\ C_2\ C_3 / C_4\ C_5\ C_6\ C_7$	0x1e	0001
2	$C_0\ C_1\ C_2\ C_3 / O_4\ D_5\ D_6\ D_7$	0x2d	0010
3	$C_0\ C_1\ C_2\ C_3 / S_4\ D_5\ D_6\ D_7$	0x33	0111
4	$O_0\ D_1\ D_2\ D_3 / S_4\ D_5\ D_6\ D_7$	0x66	1011
5	$O_0\ D_1\ D_2\ D_3 / O_4\ D_5\ D_6\ D_7$	0x55	1101
6	$S_0\ D_1\ D_2\ D_3 / D_4\ D_5\ D_6\ D_7$	0x78	1110
7	$O_0\ D_1\ D_2\ D_3 / C_4\ C_5\ C_6\ C_7$	0x4b	1000
8	$T_0\ C_1\ C_2\ C_3 / C_4\ C_5\ C_6\ C_7$	0x87	0011
9	$D_0\ T_1\ C_2\ C_3 / C_4\ C_5\ C_6\ C_7$	0x99	0101
10	$D_0\ D_1\ T_2\ C_3 / C_4\ C_5\ C_6\ C_7$	0xaa	1001
11	$D_0\ D_1\ D_2\ T_3 / C_4\ C_5\ C_6\ C_7$	0xb4	1010
12	$D_0\ D_1\ D_2\ D_3 / T_4\ C_5\ C_6\ C_7$	0xcc	1100
13	$D_0\ D_1\ D_2\ D_3 / D_4\ T_5\ C_6\ C_7$	0xd2	0110
14	$D_0\ D_1\ D_2\ D_3 / D_4\ D_5\ T_6\ C_7$	0xe1	0000
15	$D_0\ D_1\ D_2\ D_3 / D_4\ D_5\ D_6\ T_7$	0xff	1111

Hamming distance greater than two

Hamming distance greater than two

Cf. original 512b/513b (Table 1 of [4])

0001
0010
0011
0100
0101
0110
0111
1000
1001
1010
1011
1100
1101
1110
1111

2-bit Hamming distance for control block types with additional compression (flag bit for Remedy 3)

- Flag = 01 – no 66B control blocks in 1026-bit block
- Flag = 10 – One or more 66B control blocks in 1026-bit block
- Flag bits plus sixteen transcoded 66B codewords combined (as below) to form each 1026-bit block
- All 66B control blocks are sorted to front of 1026-bit block with f indicating if there are more control blocks, 4-bit POS indicating original position (row) of this control block
- Minimum 2-bit Hamming distance between coding for block types with terminate /T/ in different positions

Input Data	**Block Payload**																
Bit Position: / Data Block Format:	0																63
$D_0D_1D_2D_3/D_4D_5D_6D_7$	D_0			D_1			D_2			D_3			D_4		D_5		D_6	D_7
	Block Type Field																
$C_0C_1C_2C_3/C_4C_5C_6C_7$	f	pos	001	C_0			C_1			C_2			C_3		C_4	C_5	C_6	C_7
$S_0D_1D_2D_3/D_4D_5D_6D_7$	f	pos	010	D_1			D_2			D_3			D_4		D_5	D_6		D_7
$O_0D_1D_2D_3/C_4C_5C_6C_7$	f	pos	100	D_1			D_2			D_3			O_0		C_4	C_5	C_6	C_7
$T_0C_1C_2C_3/C_4C_5C_6C_7$	f	pos	1100101				C_1			C_2			C_3		C_4	C_5	C_6	C_7
$D_0T_1C_2C_3/C_4C_5C_6C_7$	f	pos	1100110	D_0						C_2			C_3		C_4	C_5	C_6	C_7
$D_0D_1T_2C_3/C_4C_5C_6C_7$	f	pos	1101001	D_0			D_1						C_3		C_4	C_5	C_6	C_7
$D_0D_1D_2T_3/C_4C_5C_6C_7$	f	pos	1101010	D_0			D_1			D_2					C_4	C_5	C_6	C_7
$D_0D_1D_2D_3/T_4C_5C_6C_7$	f	pos	10101	D_0			D_1			D_2			D_3			C_5	C_6	C_7
$D_0D_1D_2D_3/D_4T_5C_6C_7$	f	pos	10110	D_0			D_1			D_2			D_3		D_4		C_6	C_7
$D_0D_1D_2D_3/D_4D_5T_6C_7$	f	pos	000	D_0			D_1			D_2			D_3		D_4	D_5		C_7
$D_0D_1D_2D_3/D_4D_5D_6T_7$	f	pos	011	D_0			D_1			D_2			D_3		D_4	D_5	D_6	

Minimum Hamming distance 2

Probability of False Frame Generation

Location of Bit error \ Coding	Original 512b/513b	512b/514b 1024b/1027b 1024b/1026b	Improved by
Flag for 512b payload	$1 \times 10^{-15} \times BER$	$1 \times 10^{-17} \times BER^2$	One more header bit
Control block continuation flag (F)	$8 \times 10^{-15} \times BER$	None	Sequence check
Original position of control block (POS)	Negligible	$(9 \times 10^{-20} \times BER^3)$	(Sequence check)
4 bit encoding of control block type (CB type)	$2 \times 10^{-14} \times BER$	$2 \times 10^{-16} \times BER^2$	Re-mapping CB type

Summary of Appendix

- MTTFPA estimation has revealed that 512b/513b has issues against a bit error that hits
  - A flag bit for 512b payload
  - A control block header byte
    - 1 bit continuation flag (F)
    - 4 bit representation of 66B control block type (CB)
- MTTFPA can be improved by the following updates
  - Increase header bits
    - Add one more header bit, i.e. 512b/514b, or
    - Concatenate two 512b/513b blocks and add one more header bit, i.e. 1024b/1027b
    - Use additional compression and add one header bit; i.e. 1024b/1026b
  - Check 66B block sequence at the decoder
  - Re-map 4-bit representation of the control block type
- Potential "OTN compatibility" issue
  - MTTFPA can be improved to order of BER^2 at if transcoding is used in an error prone environment
  - Below performance of native 64b/66b, but still greater than lifetime of universe