

100 GigaBit Ethernet

A Ethernet é um conjunto de normas e padrões de rede que define regras numa LAN (Local Internet Network) para a transmissão de dados, implementando o algoritmo CSMA/CD (Carrier Sense Multiple Access with Collision Detection) para acesso a dados com detecção de colisão e o MAC (Medium Access control) para controle de acesso ao meio.

Esse protocolo é atualmente padronizado pelo IEEE 802.3, um grupo de estudo pertencente ao IEEE (Institute of Electrical and Electronics Engineers), cuja a responsabilidade é estudar e padronizar esse modelo de rede, tal qual atua na camada física e de enlace de dados no modelo OSI (Open Systems Interconnection). Os padrões são especificados por velocidade, ou seja, para cada velocidade há uma normalização. Dentro da camada física do Modelo OSI, a ethernet define padrões de cabeamento, dispositivos (switches e patch panels), faixas de envio de dados e estruturas para que a velocidade desejada seja atingida. Já na camada de enlace, é usado um controlador de link lógico para destinar os dados de forma mais eficiente e também o MAC, que define frames de dados e garante que cada dispositivo conectado a rede tenha um endereço único, evitando o envio e processamento desnecessário de informações. Para interligar essas duas camadas foi desenvolvido o reconciliador e o MII (Media Independent Interface). Nesse âmbito, a 100 Gigabit, ou 100GE, é um conjunto de normas e tecnologias de rede para transmissão de dados numa velocidade de 100 Gb/s (IEEE Computer Society (2018)).

Nesse padrão, inicialmente são determinadas as especificações da camada física (PHY - Physical Layer Device) para a transmissão desses dados, tal qual é dividida em subcamadas, são elas: Physical Coding Sublayer (PCS), Forward Error Correction (FEC), Physical Medium Attachment (PMA), Physical Medium Dependent (PMD) e o Medium.

A primeira subcamada física (PCS) provê o serviço de codificação/decodificação dos dados em blocos de 66bit (64B/66B), é responsável por distribuir os dados em diferentes faixas, realizar o *scramble* dos blocos de bits, compensação de diferença de taxas entre o reconciliador e o PMA, determinar quando uma conexão foi estabelecida informando então ao gerenciador quando o dispositivo está pronto para uso.

Já na segunda subcamada física o FEC (Forward Error Correction) age com o objetivo de evitar a perda de dados através da redundância no envio de bits, onde ele faz a mesma adicionando bits ao streaming de dados pelo algoritmo Reed-Solomon, sendo então nomeado como RS-FEC (Reed Solomon Forward Error Correction). Em cada especificação o RS-FEC trabalha de uma forma e, em sua implementação na 100GE, é necessário exatamente quatro faixas de envio e outras quatro para recebimento, sendo

indispensável o mapeamento 10:4 quando trabalha com o PMA possuindo 10 faixas, pois tal PMA opera com 10 faixas para envio e outras 10 para recebimento.

A terceira subcamada, o PMA (Physical Medium Attachment), fornece o serviço de intermediação entre um PMA e um cliente, podendo esse cliente ser um PCS, FEC ou outro próprio PMA. Entre esses serviços têm-se a adaptação dos sinais das faixas dos PCS para o número de faixas físicas ou abstratas do cliente, ou seja, ele pode receber 10 faixas de stream de dados e transformá-lá em 4 faixas de stream de dados. Também provê especificações de tempo para transmissão dos dados entre as faixas assim como gerenciamento dos mesmos. O PMA faz o direcionamento de bits de dados para que todos os bits de uma stream vão e voltem pela mesma faixa. Ainda na terceira camada, quando há a comunicação entre dois PMAs, pode-se usar a instanciização CAUI-10 (100 Gb/s ten-lane Attachment Unit Interface) ou CAUI-4 (Chip-to-chip/Chip-to-module 100 Gb/s four-lane Attachment Unit Interface), onde a CAUI-10 é a especificação elétrica de módulos plugáveis com dez faixas a 10.3125 GBd e a CAUI-4 é a especificação elétrica, de módulos plugáveis e pontos combinados, com quatro faixas a 25.78125 GBd.

A quarta subcamada (PMD) provê o serviço de intermédio entre o PMA e o MDI controlando o envio e recebimento dos dados entre os mesmos, traduzindo o código recebido do PMA de streamings de bit para streamings elétricas ou de streamings de bits para streamings de sinais óticos e o contrário também, onde o PMA trabalha com bits e o MDI com sinais elétricos e/ou óticos. Também na implementação do PMD é decidido qual modo de comunicação/conexão usar, exemplo: Fibra ótica em Single-Mode, MultiMode ou também cabos de cobre.

Relacionado ao PMD, tem-se ainda o MDI (Medium Dependent Interface), que é a interface de comunicação entre o dispositivo PMD e o Medium, podendo o Medium ser entendido como meio de comunicação (fibra ótica, cabo de cobre, backplane). Essa interface pode ser compreendida de outro modo como o receptor e/ou transmissor acoplado ao dispositivo PMD, e varia conforme a normativa.

Já na camada de enlace, tem-se também as divisões de especificações e como principais entidades há o LLC (Logical Link Control), o MAC (Media Access Control) e também o MAC Control, que na implementação da 100GE não é necessário.

Entre as entidades, inicialmente há o MAC, que provê o serviço de transferência de dados entre MACs, onde sua semântica de transferência é constituída de: endereço de destino (que pode ser um MAC ou um grupo), endereço de origem, unidade de serviço de dados MAC e sequência de checagem de frame. Tais semânticas trabalham através de frames e pacotes e, durante os anos, foram adicionados mais capacidades ao encapsulamento desses frames e como consequência há mais de um tipo de frame, todos utilizando o mesmo formato de frame Ethernet e o 802.3 padroniza três deles: o básico, o Q-tagged e o envelope. Tal frame é encapsulado num pacote pelo MAC e

cada elemento é especificado conforme a tabela abaixo:

		Quantidade de Bytes	Campo
Pacote		7 Bytes	Preâmbulo
		1 Byte	SDF
	Frame	6 Bytes	Endereço de Destino
		6 Bytes	Endereço de Origem
		2 Bytes	Tamanho / Tipo
		46 a (1500 ou 1504 ou 1982) Bytes	Dados Cliente MAC (PayLoad) / Pad (Opcional)
		4 Bytes	Sequência de checagem de frame
		4 Bytes	Extensão

Formato de Frame e Pacote Ethernet

O primeiro elemento (preâmbulo), ajuda na sincronização do PLS com o tempo do pacote e serve para avisar que um frame está a caminho. O SFD é a sequência de dados fixada (10101011) que antecede o frame, ou seja, depois dela o receptor saberá que será os bits do frame. Os campos de endereço possuem 48 bits cada, e o endereço de destino pode ser um MAC unico, um grupo ou todos os endereços da LAN. O campo de Tamanho / Tipo possui dois significados, se for menor ou igual a 1500 indica o número de bytes dentro do próximo campo (Dados Cliente MAC), se estiver entre 1501 e 1536 então indica o Ethertype do protocolo do cliente MAC.

No campo de dados do cliente MAC, há os dados a serem transmitidos e a implementação deve suportar no mínimo os tamanhos de dados de 1500 Bytes (frame básico), 1504 Bytes (frame q-tagged) e 1982 Bytes (frame envelope). Já o elemento Pad é utilizado quando o campo de dados não atingem o número mínimo de 48 bytes, ou seja, ele é a adição de dados ao campo para que o frame não seja eliminado no futuro como um frame com quantidade de dados insuficiente. A sequência de checagem de frame (FCS) é utilizada para validação do frame e é gerada a partir de dos dados do mesmo para que haja detecção de erro no recebimento, ou seja, se o calculo da sequência no recebimento for diferente do FCS recebido, significa que o frame está errado. O último campo (Extensão) não é necessário a partir da 10GE e também não entra no calculo do FCS.

Depois de encapsulado, o frame é enviado e na recepção é considerado inválido quando: seu tamanho é incondizente com o especificado no elemento de tamanho/tipo; se o frame não possuir a quantidade de bits múltipla de 8, pois deve ser uma cadeia de bytes; ou o FCS calculado não coincidir com o valor FEC recebido.

O MAC Control com CSMA/CD não se faz necessário na 100GE pois essa funcionalidade usa o algoritmo CSMA/CD (Carrier Sense Multiple Access with Collision Detection). Tal algoritmo não é útil na 100GE visto que ela opera somente em modo *full duplex*, logo não risco de colisão de dados.

Ainda na camada de enlace, porém acima do MAC, tem-se o LLC (Logical Link Controller) que facilita, através de mecanismos de multiplexação e demultiplexação, o transito e coexistência de vários pacotes num meio de rede com vários pontos. Isso é possível pois ele guarda o endereço de cada MAC dentro da rede e faz todos se enxergarem como um, ou seja, enquanto o MAC guarda a informação dos dados e dispositivos para mostrar a origem e destino do pacote, o LLC mostra o melhor caminho a ser percorrido para esse pacote chegar ao objetivo.

Esses conceitos tecnológicos (PHY, MAC e LLD) se referem as duas primeiras camadas físicas do modelo OSI e para interligar as duas o 802.3 também padroniza o reconciliador (RS). Opcionalmente o 802.3 também padroniza as MII (Media Independent Interface), que provê a interconexão lógica entre o MAC e o PHY, atuando então embaixo do RS. O MII foi desenvolvido para que a camada de enlace de dados e o meio físico trabalhem de forma independente e é especificado na 100GE como CGMII.

Em suma, o RS converte a stream de dados dada pelo MAC para dados (sinais) paralelos do CGMII e também o mapeamento dos sinais providos pelo CGMII para as primitivas do MAC, já CGMII é o facilitador de transmissão e recebimento de dados entre o RS e o PHY.

Todas essas definições são padronizadas pela IEEE para a 100GE e vários fatores foram essenciais para o alcance de tal velocidade, isso fica claro ao comparar-lo com outros padrões como 10GE, 25GE e 400GE, sendo eles conjuntos de normas para a velocidade, respectivamente, de 10 Gb/s, 25 Gb/s e 400 Gb/s, todos eles definidos pelo grupo 802.3.

	10GE	25GE	100GE	400GE
Bloco de dados no RS (bits)	32	32	64	64
Faixas	1	1	4 ou 10	16
Velocidade por faixa (Gb/s)	10	25	25 para 4 faixas ou 10 para 10 faixas	25

Especificações de Normas 802.3

O primeiro dado se refere aos blocos de bits transmitidos através do RS, a qual

se observa um aumento para o dobro do tamanho, a importância desse item é visto quando calcula-se a velocidade de transmissão com 10 faixas transmitindo a 156,25 Mhz: $10(faixas) \times 64(bits) \times 156,25 = 100(Gb/s)$.