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 algorítmo 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. Ela surgiu quando Robert Metcalfe escreveu um memorando aos seus chefes descrevendo o potencial dessa tecnologia para redes locais, logo após saiu da empresa onde trabalhava (Xerox), porém conseguiu convencer a DEC (Digital Equipment Corporation), Intel e Xerox a trabalharem juntas para promoverem a Ethernet como um padrão.

Esse protocolo é atualmente padronizado pelo IEEE 802.3, um grupo de estudo pertencente ao IEEE (Institute of Eletrical 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) 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, para 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 cada norma tem um reconciliador específico, na 100 Gigabit Ethernet é recomendado o CGMII. 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 é determinada 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 implementação dessa camada, na 100 Gigabit Ethernet, é dada através do 100GBASE-R ou 100GBASE-X, que é a família de dispositivos que trabalha na velocidade de 100 Gb/s, sendo que este usa o PCS com codificação (64B/66B), enquanto aquele usa codificação External Sourced (4B/5B, 8B/10B). As demais familias da 100GE (100GBASE-KR4, 100GBASE-CR4, 100GBASE-SR10 e etc.) são baseadas nessas duas codificações, a 100GBASE-R e 100GBASE-X.

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 (Foward Error Correction) age com o objetivo de evitar a perca de dados através da redundância no envio de bits, onde ele faz a mesma adicionando bits ao streaming de dados pelo algorítimo Reed-Salomon, 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 Meidium Attachment), fornece o serviço de intermediamento 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 transforma-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. Como o PMA possui faixas de transmissão para prover tais serviços, há também uma padronização para tais faixas, onde o 802.3 define que o número de faixas do PMA sempre é divisor do número de faixas do PCS, então para suportar a 100GBASE-R é necessário que o PMA tenha 20 PCSLs. Ainda na terceira camada, quando há a comunicação entre dois PMAs, pode-se usar a instanciação CAIU-10 ou CAIU-4, onde a CAIU-10 é o uso de dez faixas a 10.3125 GBd e a CAIU-4 é o uso de 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 stramings 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. Há maneiras de ser implementado o PMA e alguns dos padronizados pelo 803.2 são os modos: 100GBASE-KR4 que é o PHY a 100 Gb/s, codificação de 64B/66B, RS-FEC e amplitude da modulação de 2 pulsos em 4 faixas de backplane elétrico; o 100GBASE-CR4 com PHY a 100 GB/s, codificação de 64B/66B, RS-FEC e 4 faixas de cabo de cobre balanceado e distância de 5m; o 100GBASE-SR10 com PHY a 100Gb/s, codificação de 64B/66B e 10 Faixas

de fibra em multimodo e distância de 100m. Há mais modos padronizados no IEEE Standards for Ethernet (2018) na Tabela 80-1.

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, sendo que na 100GBASE-SR10 é apresentada três opções de implementação e a recomendada é a que há dez faixas para transmissão e outras dez para recepção (IEEE Standards for Ethernet (2018) - Figura 86-7).

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 constituda 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 e, essa semantica especificada, refere-se a premissa MA_DATA.request. Quando há a implementação do controle de MAC, há também a recepção de status, que é usado para informar ao cliente MAC a vinda de um frame, e esta se trata da premissa MA_DATA.indication.

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 fomarto 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 / 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. Se o primeiro bit for 0, significa que se refere a um endereço (unicast), se for 1 segnifica que que é mais de um endereço (multicast ou broadcast), enquando o segundo bit (do endereço destino) irá dizer se o endereço é administrado localmente ou globalmente. Para o broadcast, todos os bits do elemento devem ser 1. 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 trasmitidos 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 eleminado 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 ultimo campo (Extensão) é usado para slotTime e serve apenas para que o frame tenha o tamanho mínimo exigido em implementações de baixa velocidade, não sendo necesseário 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 incodizente com o especificado no elemento de tamanho/tipo; se o frame não possuir a quantidade de bits multipla 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 algoritimo CSMA/CD (Carrier Sense Multiple Access with Collision Detection). Tal algorimo não é util na 100GE pois ela opera semente 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. Tal funcionalidade é exemplicada no serviço provido pelos equipamentos chamados Switches. Essa tecnologia é normalizada pelo grupo 802.2 da IEEE e não está voltada somente para a Ethernet, ela pode e é implementada em outros padrões de rede como FDDI, Token Ring e WLAN.

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, quando não implementado, deve haver conformidade na implementação para que a mesma funcione como se o MII estivesse implementado, pois o PCS é especificado para o MII (nomeado 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 composto de transmissão e recebimento de dados, divididos em 64 sinais através de 8 faixas com 8 sinais cada e esses dados são tramitados atrvés de um frame interno cuja a estrutura é quase identica ao padrão de frame do MAC.