Software Defined Accelerated Networking @ Microsoft

James Hurd

Who Am I?

- Software Engineer @ Microsoft
  - Azure Host Networking
- KU CS ‘25
  - ACM Co-President (2x)
  - HackKU Director
  - 168/268 Lecture Assitant
- Prev. Los Alamos National Lab
- Prev. Zeroeyes
- Prev. Pinnacle Technology

Lawrence, KS

Seattle,WA

Prerequisites 📚

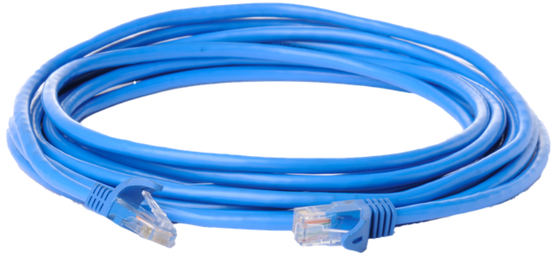
Chapter 1:
Networking Crash Course

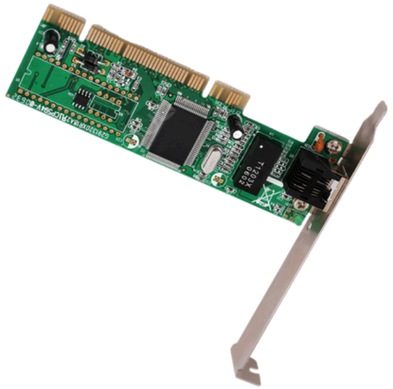
“If you knew how complicated the internet was, you’d be surpised it works at all”~ Mike McClure

A **network** is a group of computers that can share data with each other.

(e.g. JAYHAWK)

World’s Smallest Computer Network!





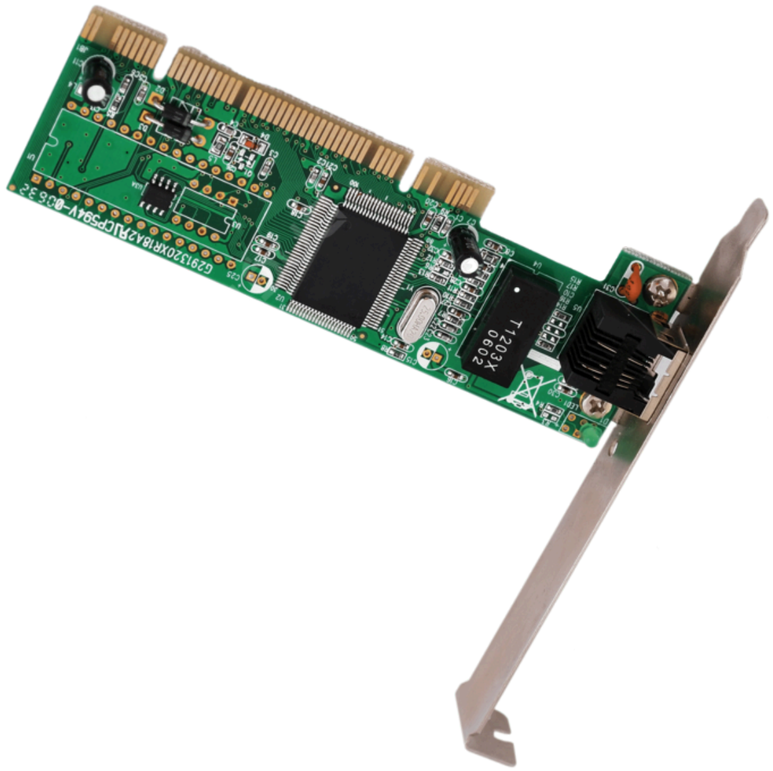
01101110 01100111 00100000 01110100
01110010 01101001 01100101 01100100
00100000 01110100 01101111 00100000
01100100 01100101 01100011 01101111
01100100 01100101 00100000 01110100
01101000 01101001 01110011 00100000
00101101 00101101 00100000 01101110
01100101 01110010 01100100 00100001
00100000

This doesn’t scale…

Switches!!!

How does the switch know where to send data?

We’ve unlocked our first header…. the ethernet frame!



00:00:0A:BB:28:FC

Media Access Control (MAC) Address

Everyone knows everyone's MAC attached to the switch, and the switch knows how to forward traffic

Simplified Ethernet Frame

Example

MAC A

MAC D

MAC B

MAC C

Example

MAC A

MAC C	Data

MAC D

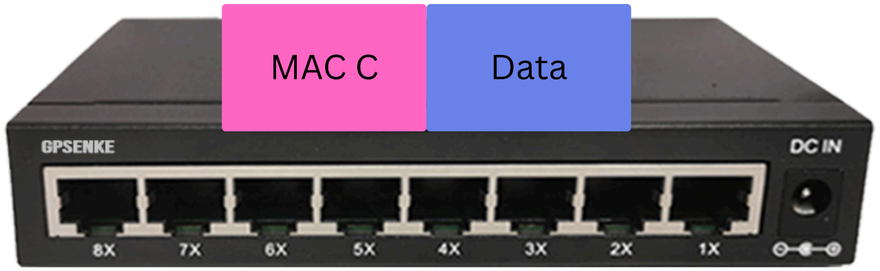
MAC B

MAC C

Example

MAC A

MAC D



MAC B

MAC C

Example

MAC A

MAC D

MAC B

MAC C

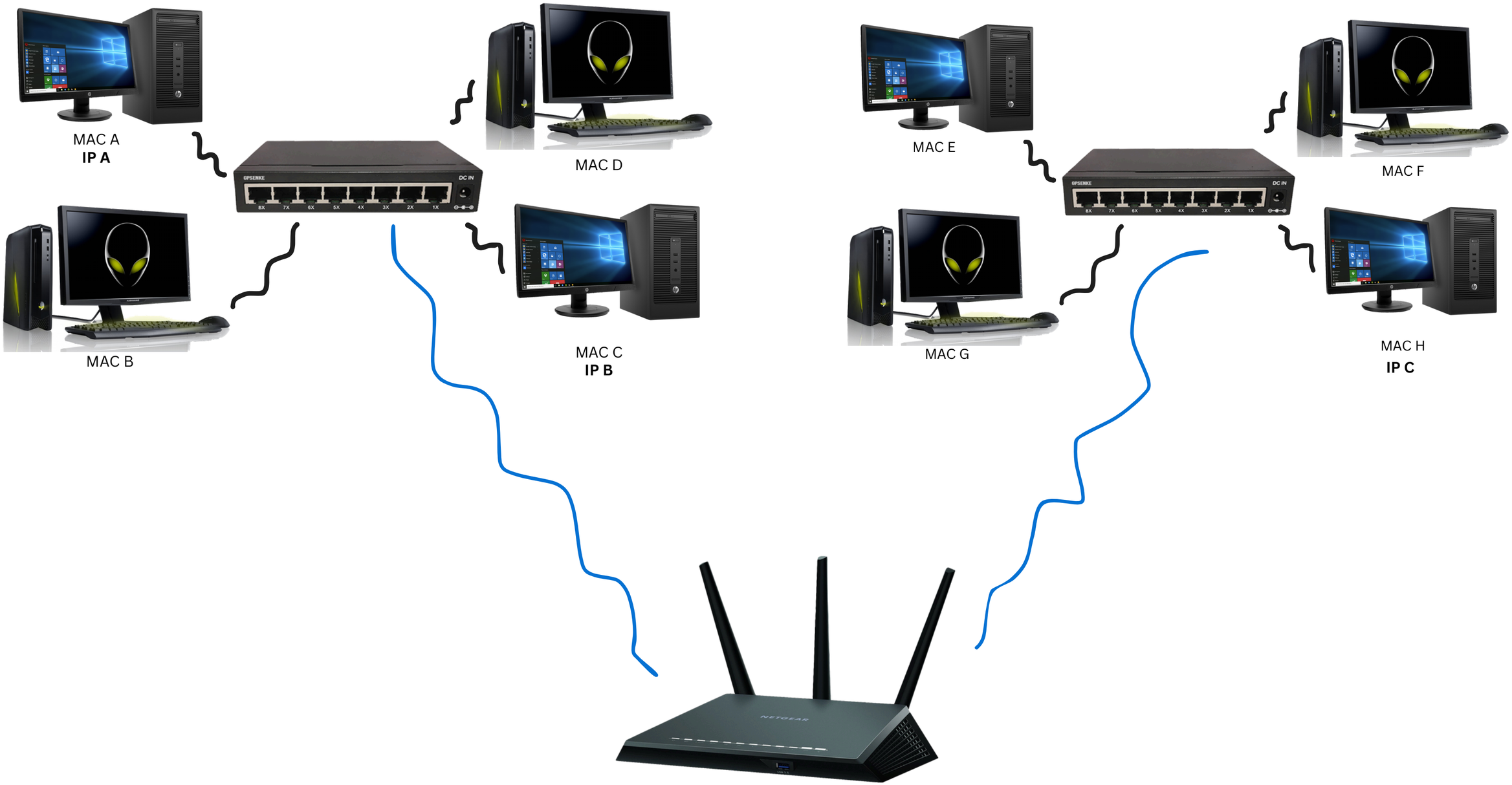
MAC C	Data

What happens if I want to send traffic to somewhere not connected to the switch…..

I’m glad you asked!! 😄

enter …. the router!

Example

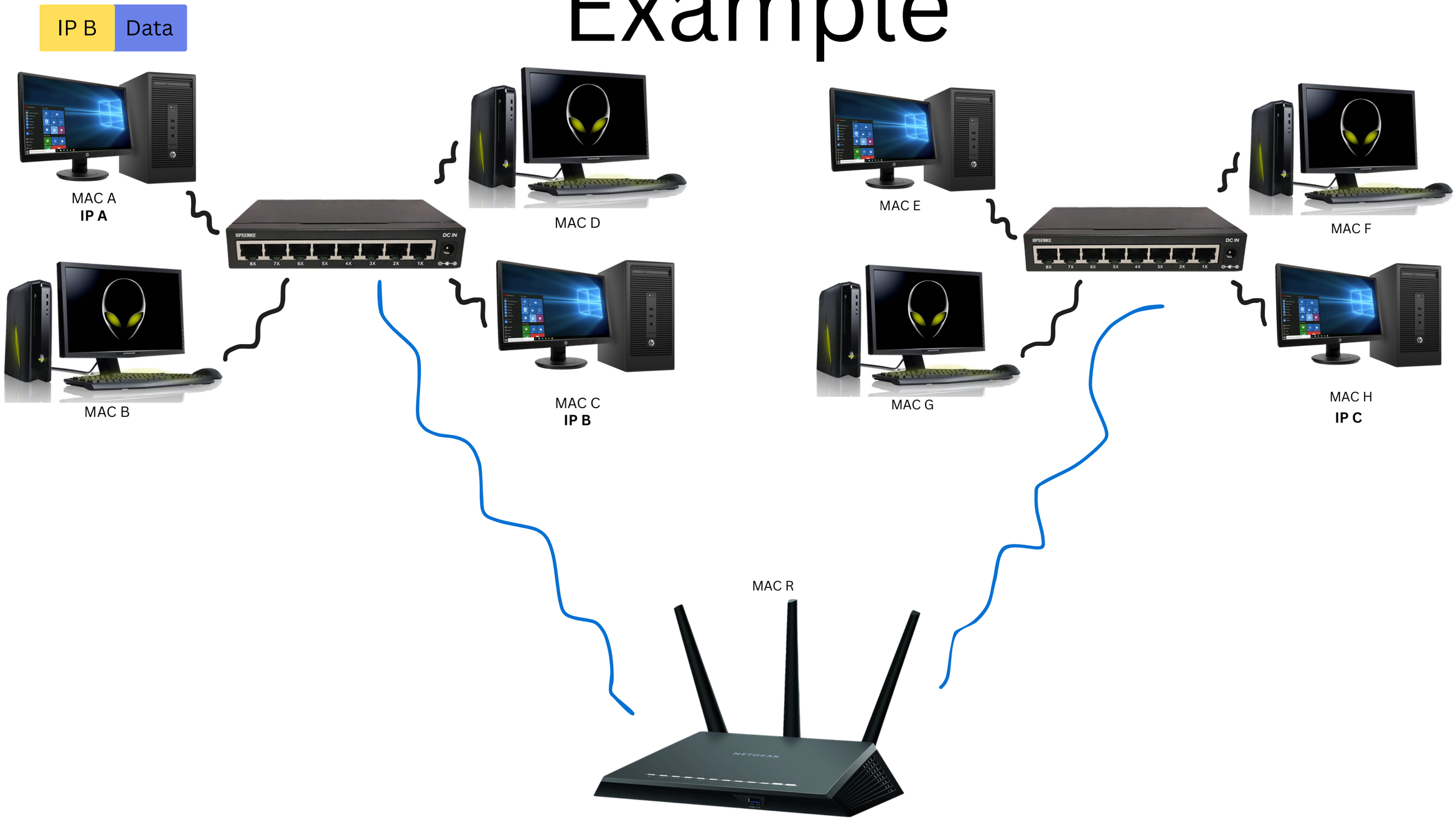


Switches only know how to get to things directly connected to them... how do we identify the destination across switches?

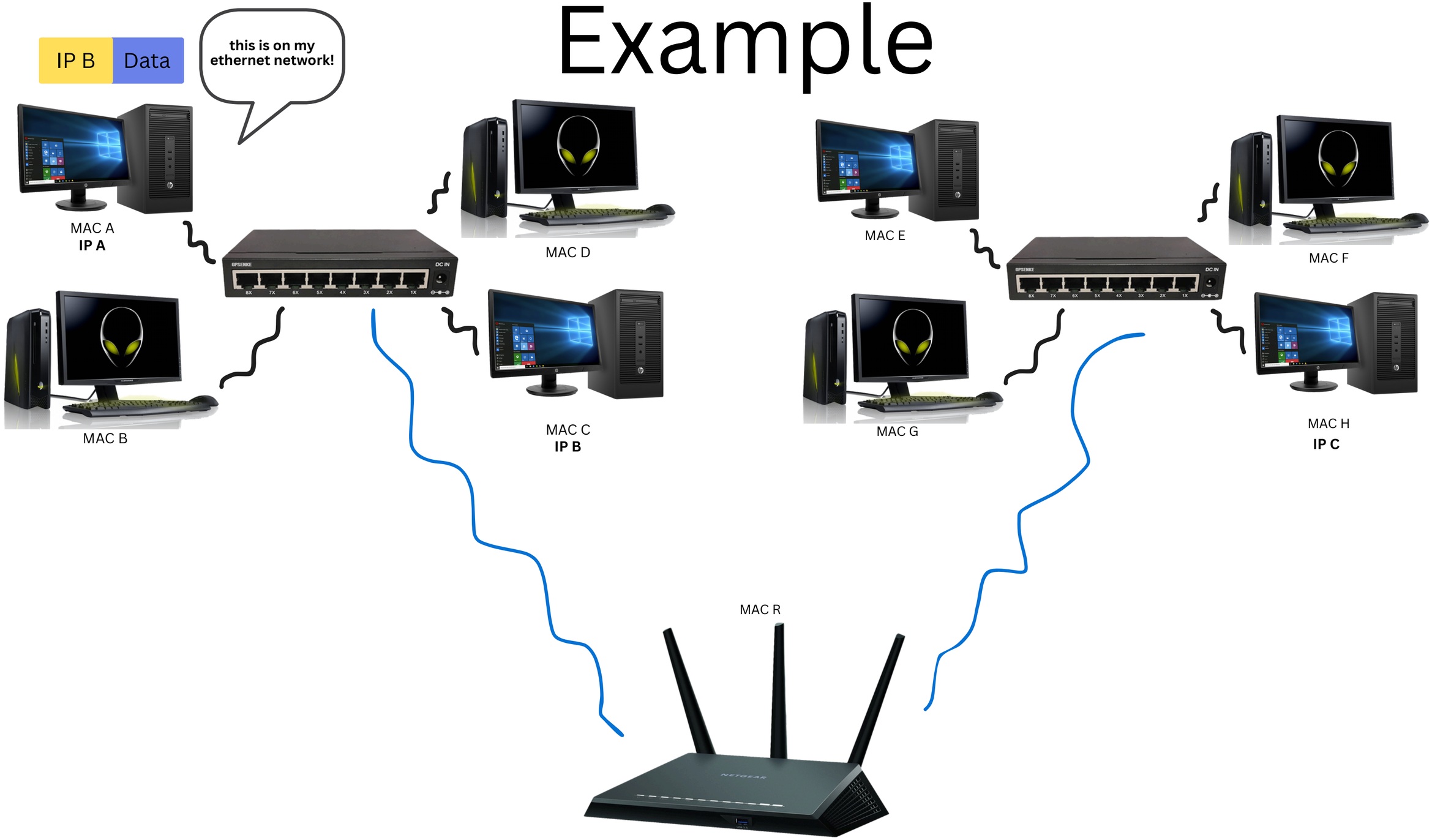
IP Adresses!

Simplified IP Packet

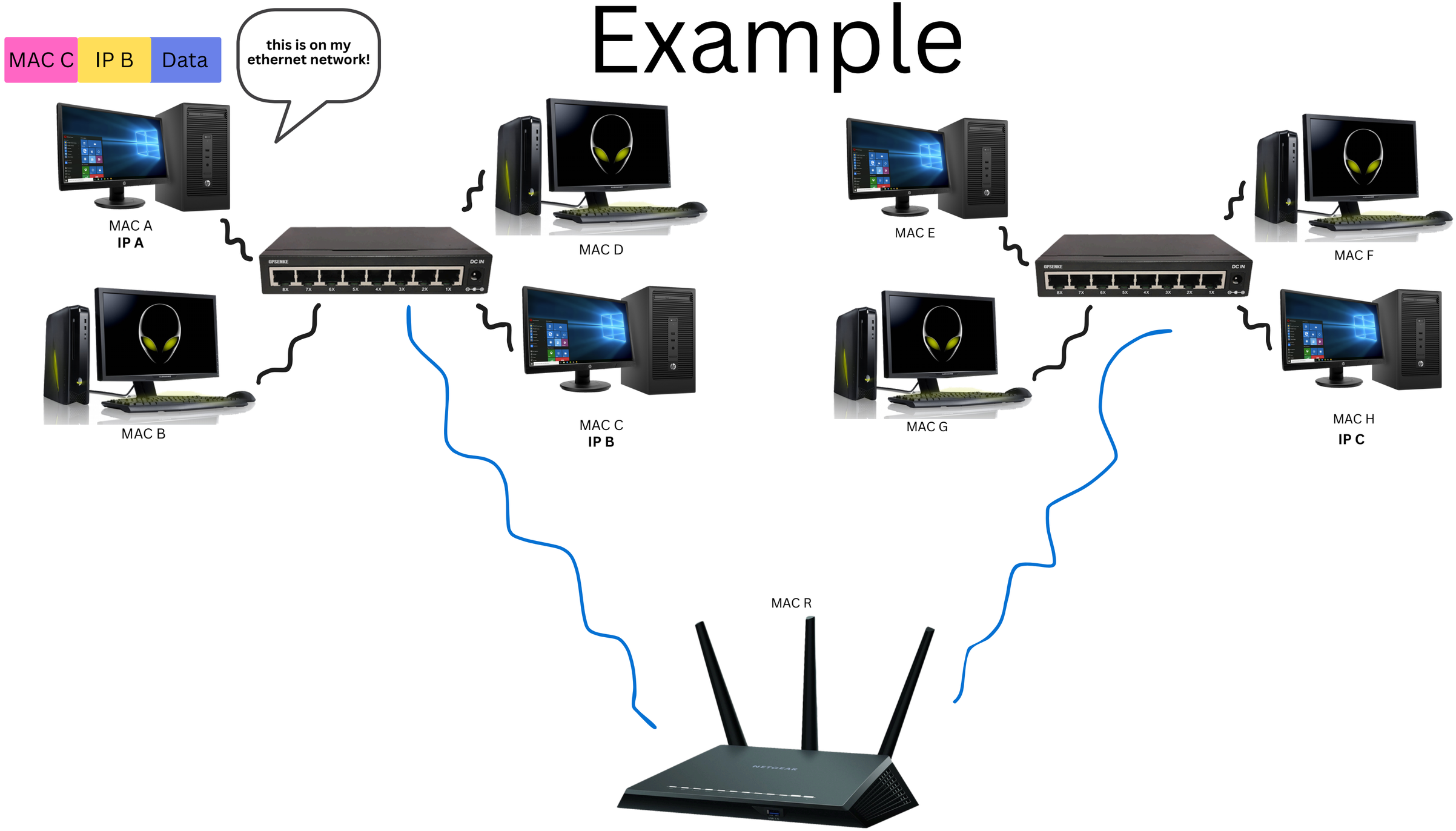
Example


IP B
Data
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

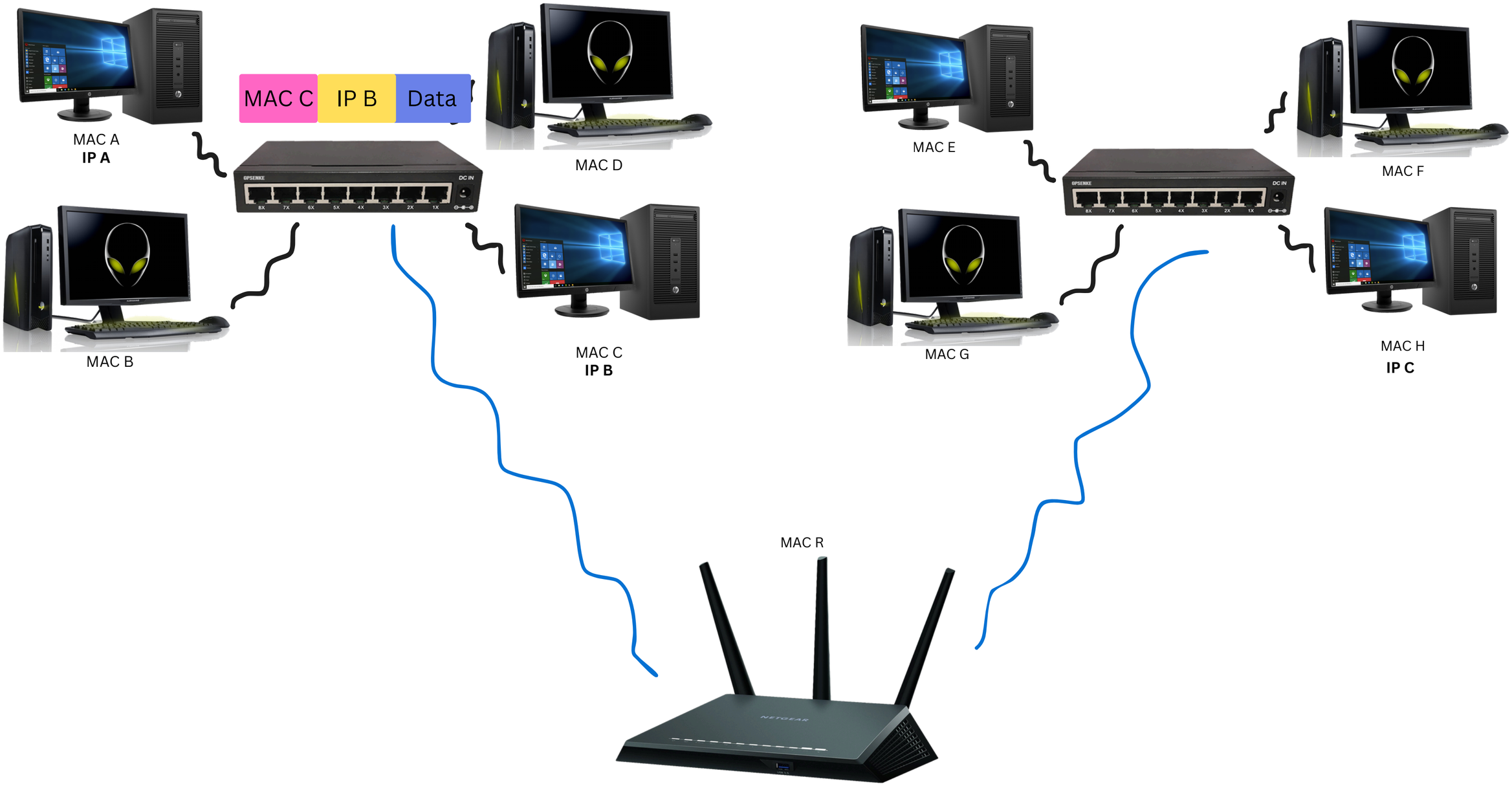
Example


IP B
Data
this is on my ethernet network!
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

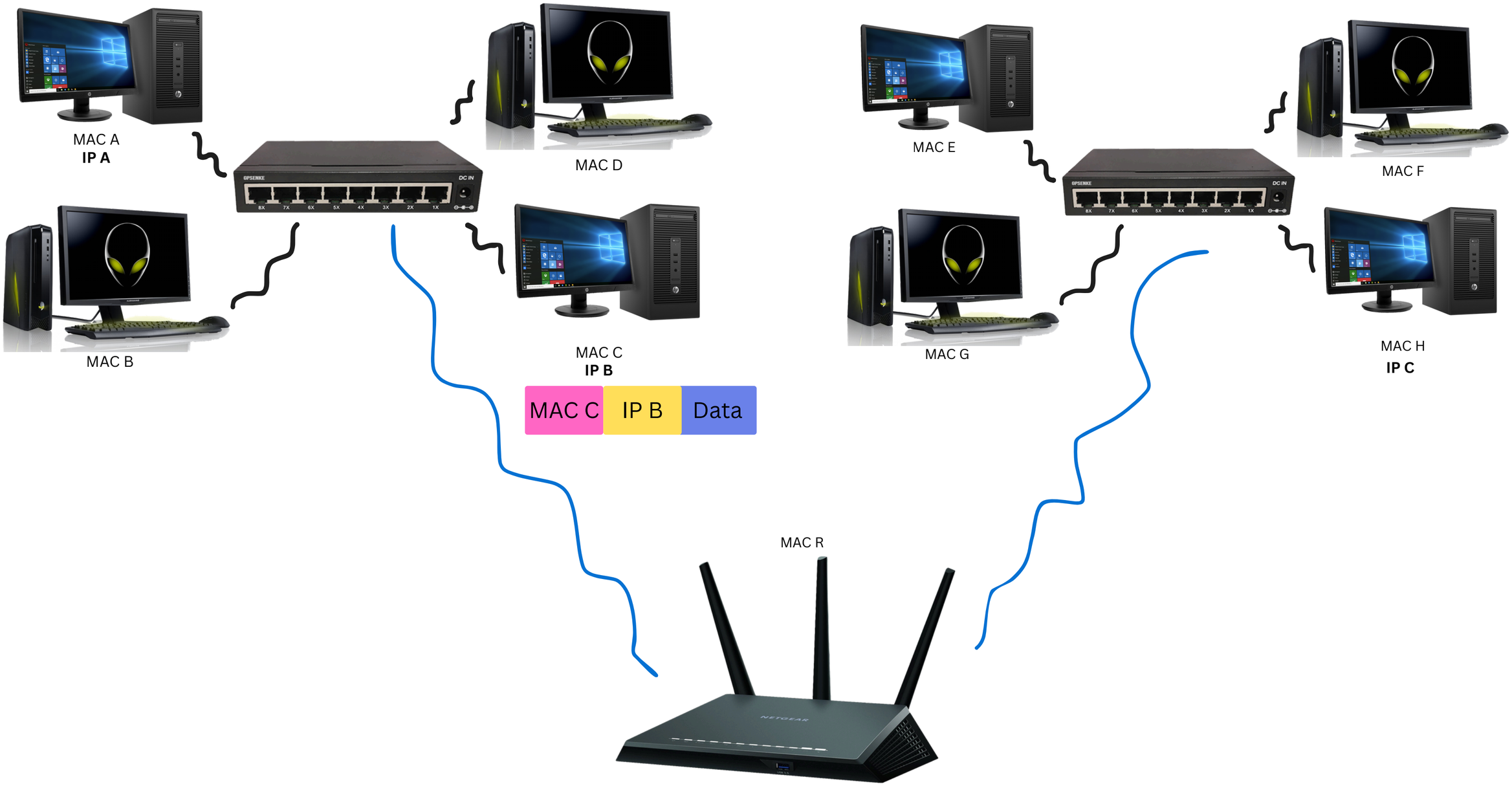
Example


MAC C
IP B
Data
this is on my ethernet network!
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

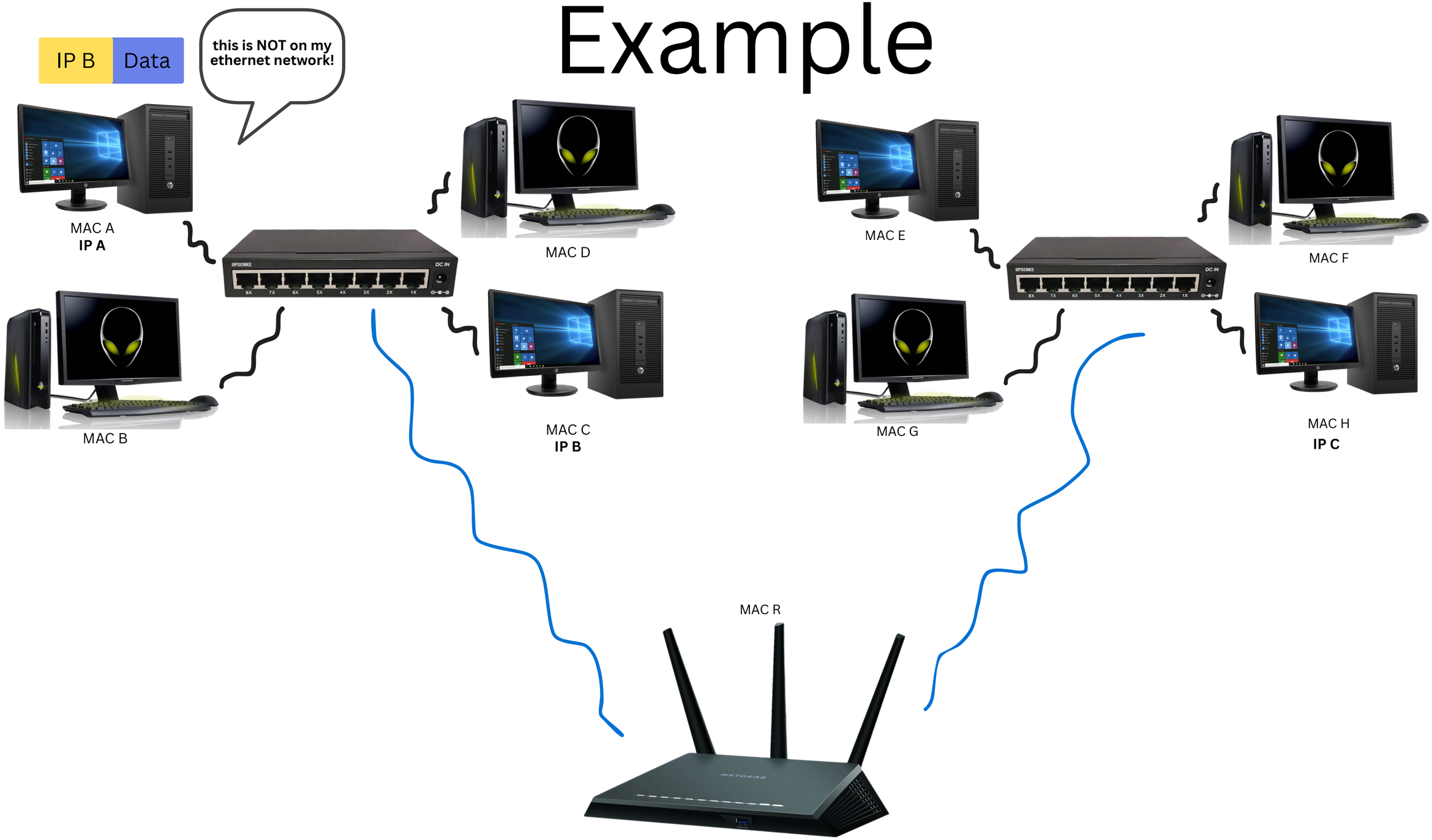
Example


MAC C
IP B
Data
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

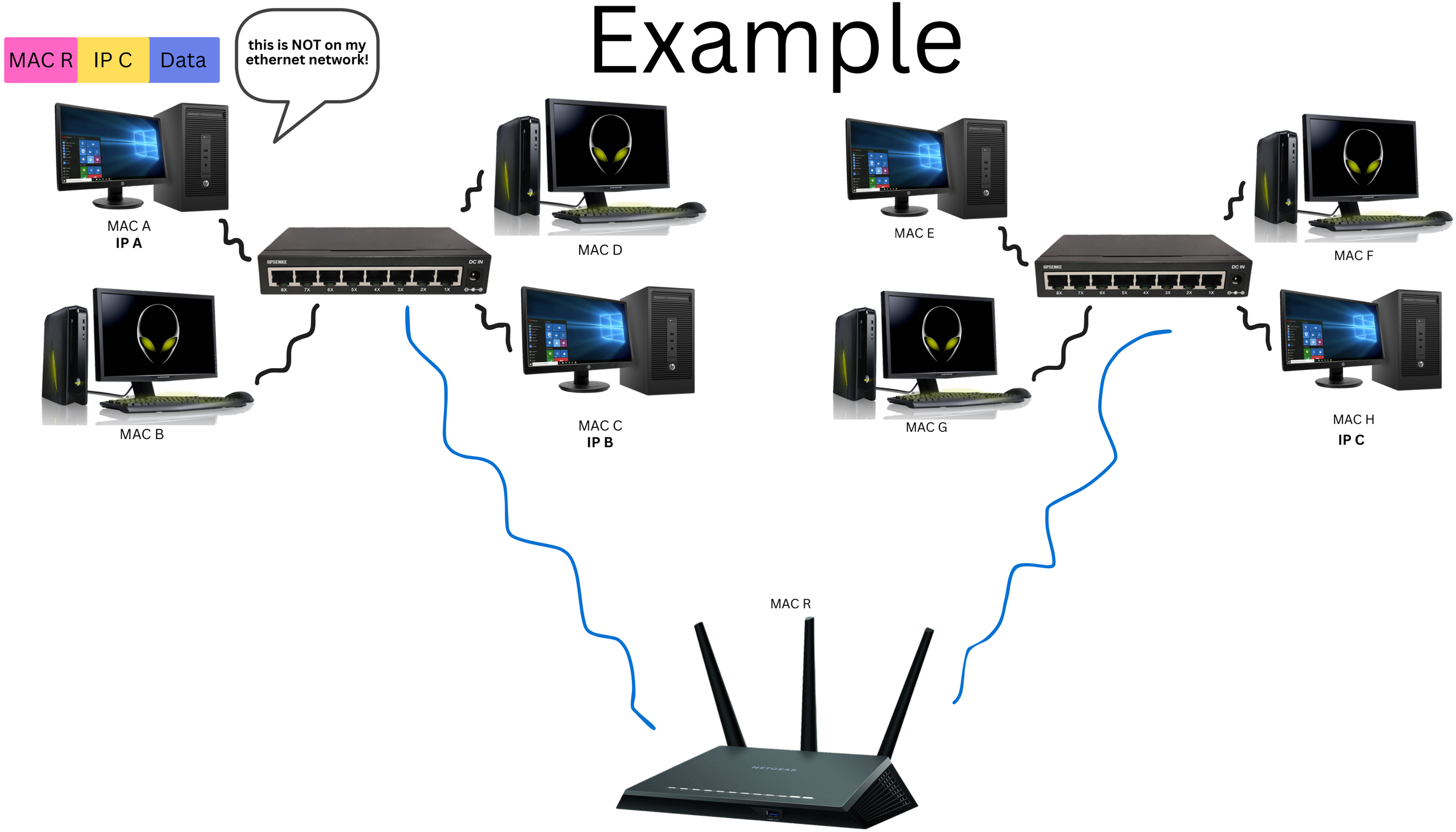
Example


MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC C
IP B
Data
MAC R

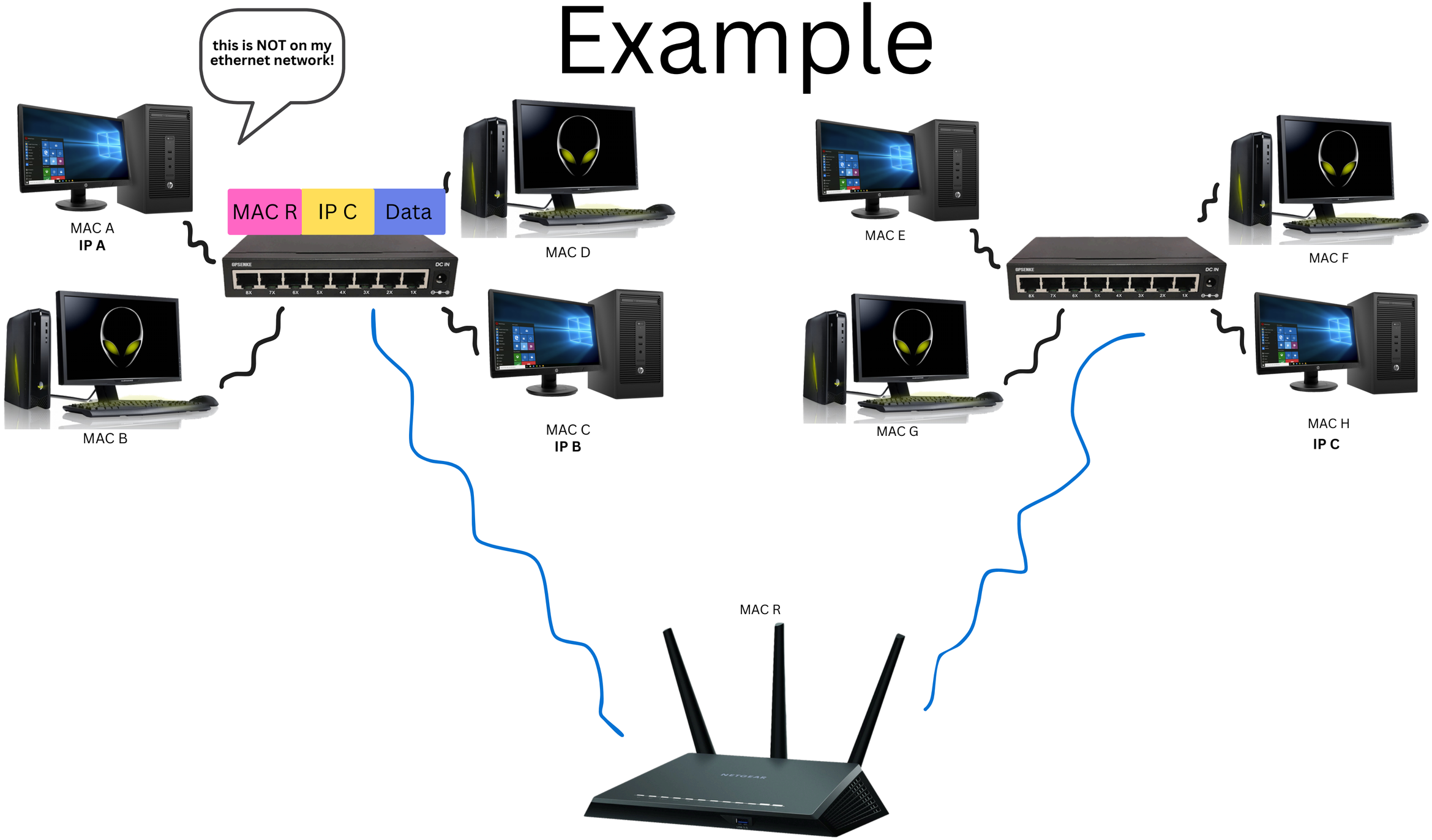
Example


IP B
Data
this is NOT on my ethernet network!
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

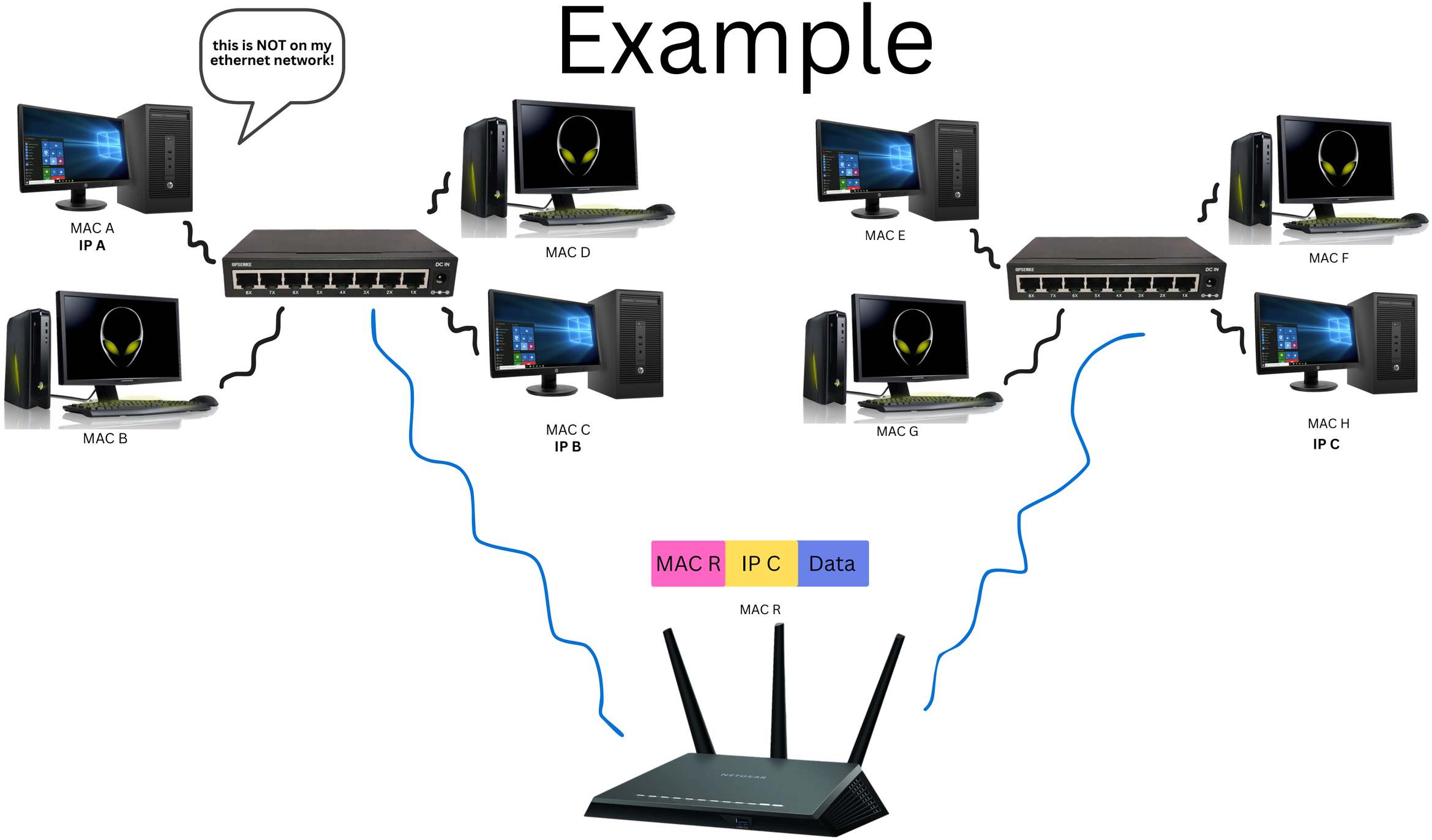
Example


MAC R
IP C
Data
this is NOT on my ethernet network!
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

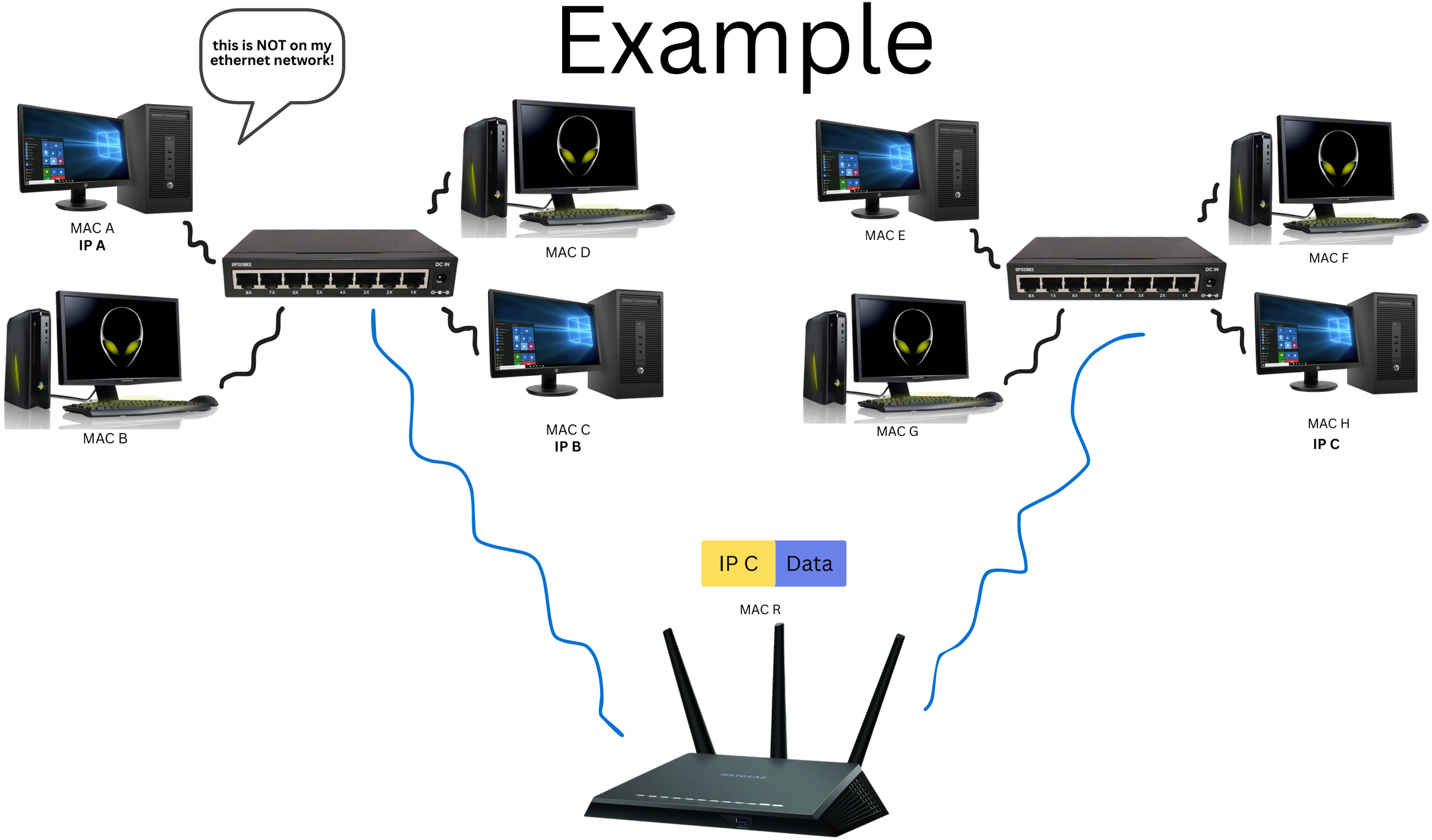
Example


this is NOT on my ethernet network!
MAC R
IP C
Data
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

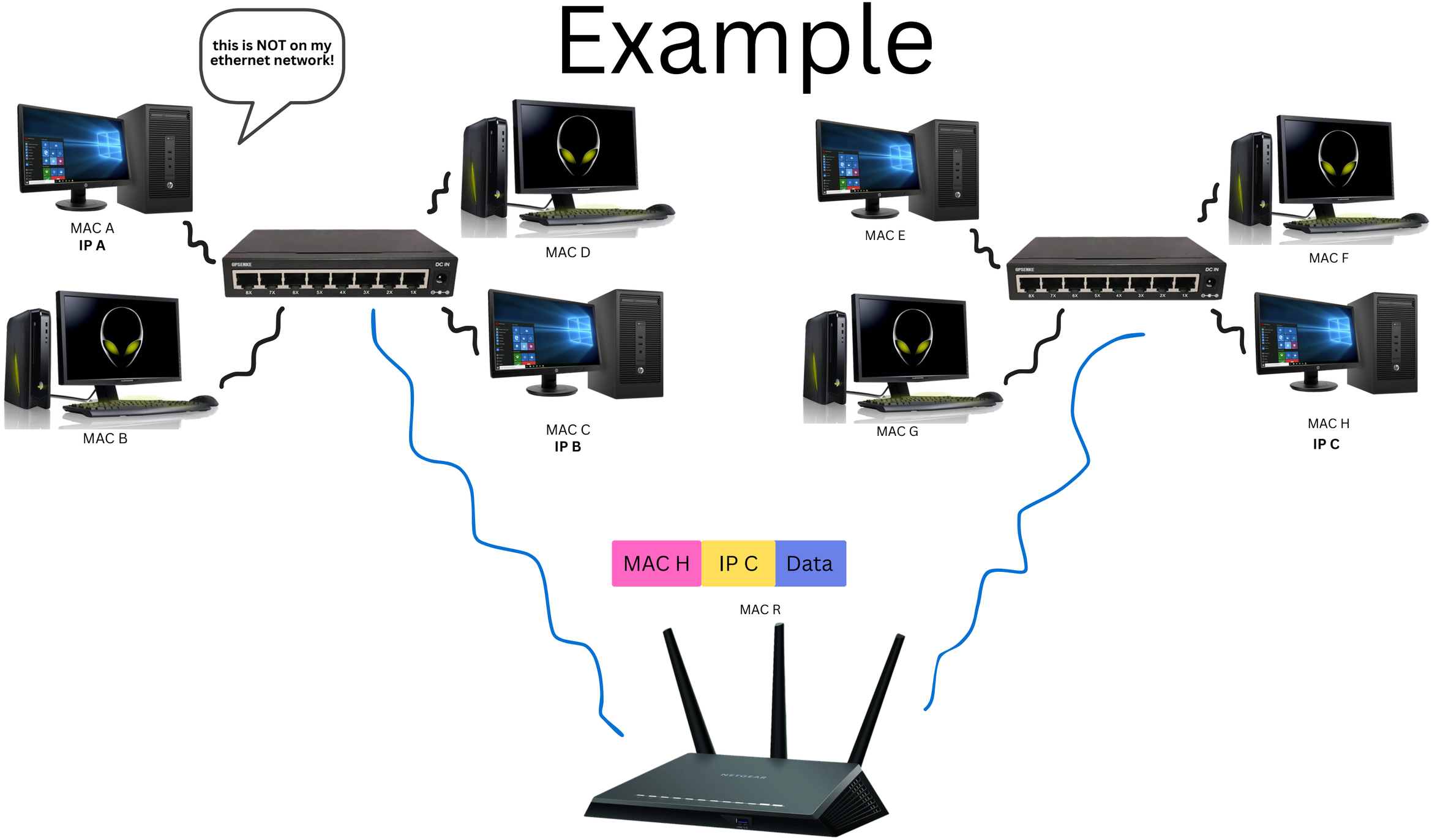
Example


this is NOT on my ethernet network!
MAC A
IP A
MAC B
MAC D
MAC C
IP B
MAC E
MAC F
MAC G
MAC H
IP C
MAC R
IP C
Data
MAC R

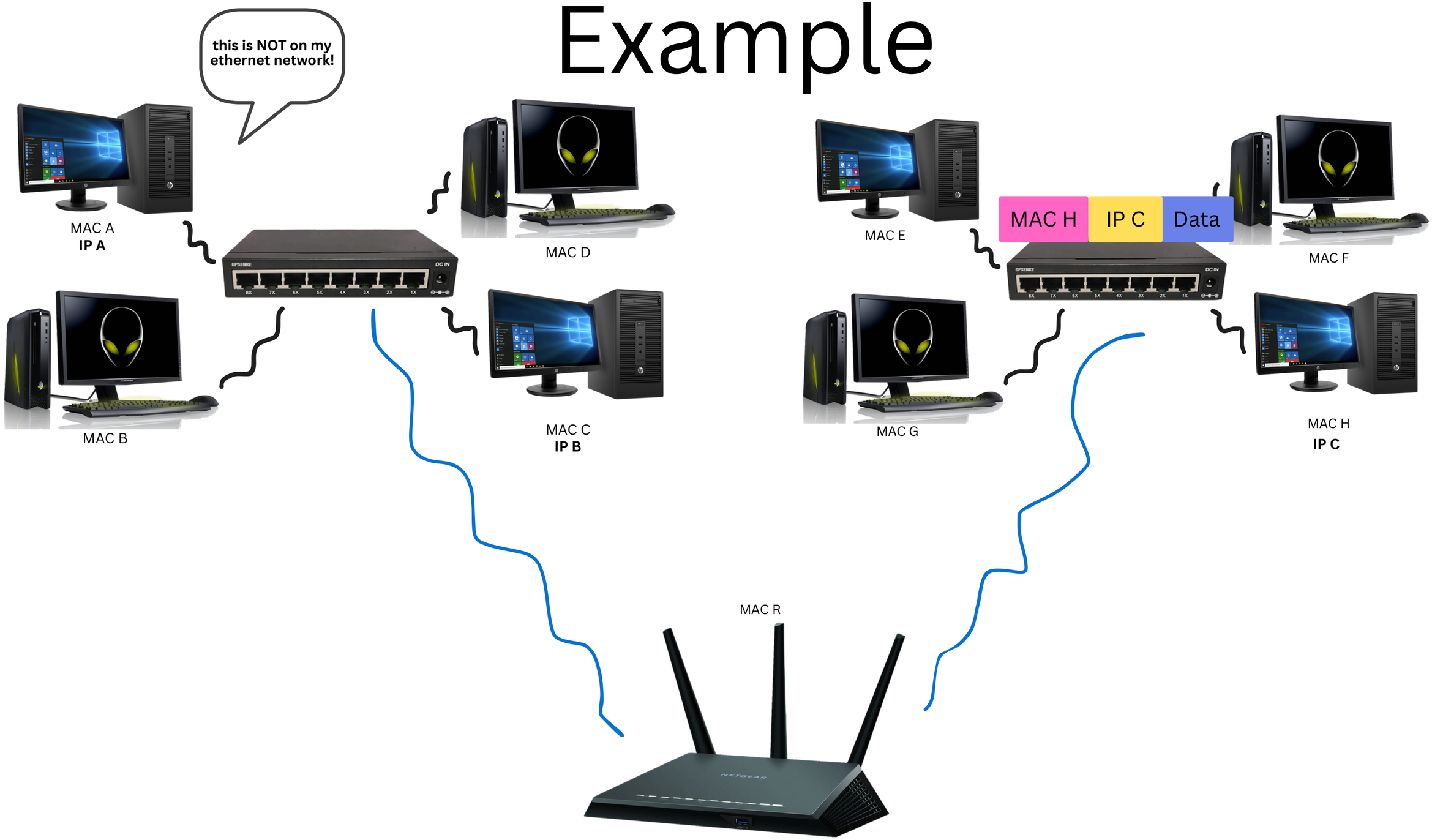
Example


this is NOT on my ethernet network!
MAC A
IP A
MAC B
MAC D
MAC C
IP B
MAC E
MAC F
MAC G
MAC H
IP C
IP C
Data
MAC R

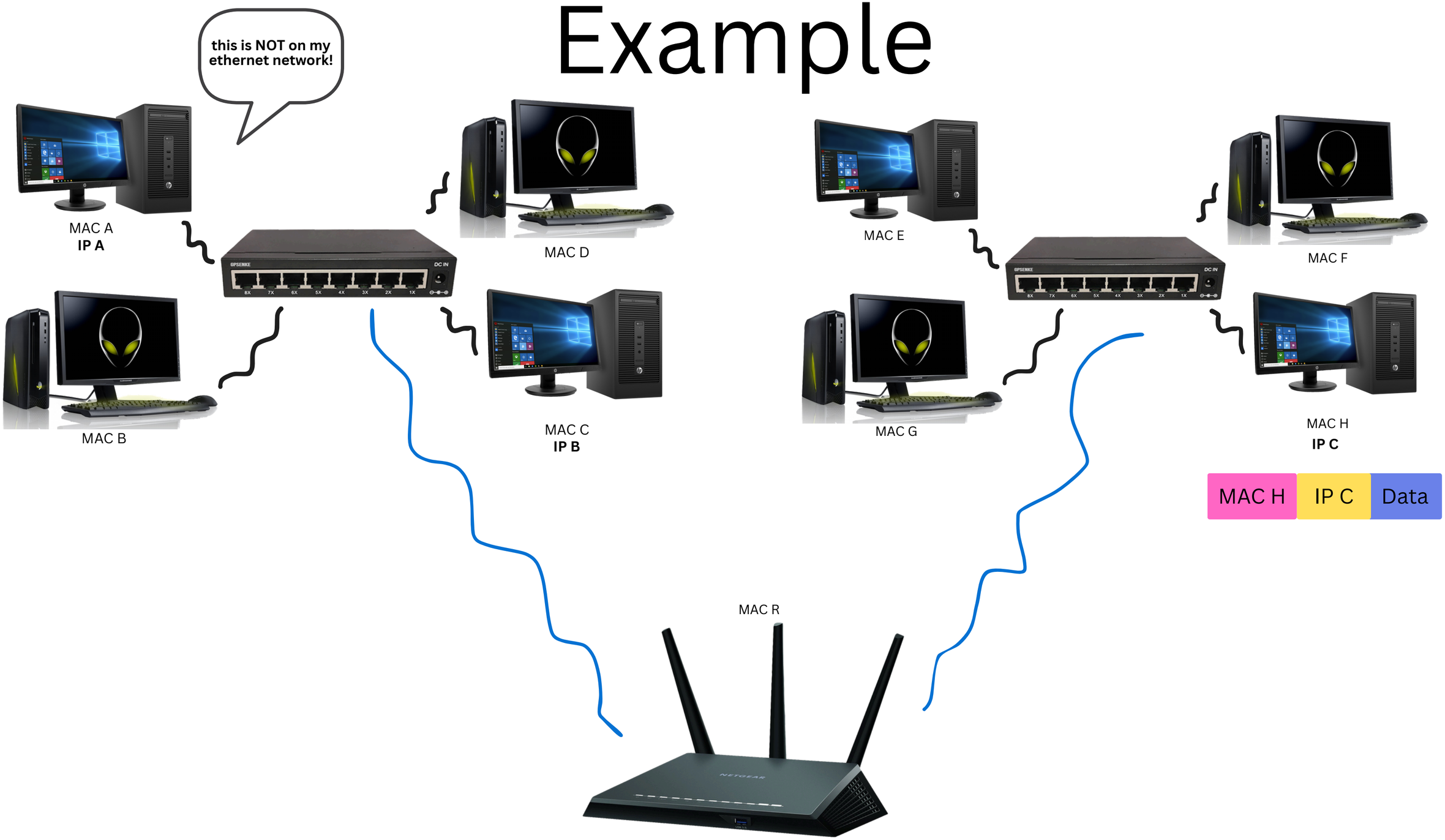
Example


this is NOT on my ethernet network!
MAC A
IP A
MAC D
MAC E
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC H
IP C
Data
MAC R

Example


this is NOT on my ethernet network!
MAC A
IP A
MAC D
MAC E
MAC H
IP C
Data
MAC F
MAC B
MAC C
IP B
MAC G
MAC H
IP C
MAC R

Example

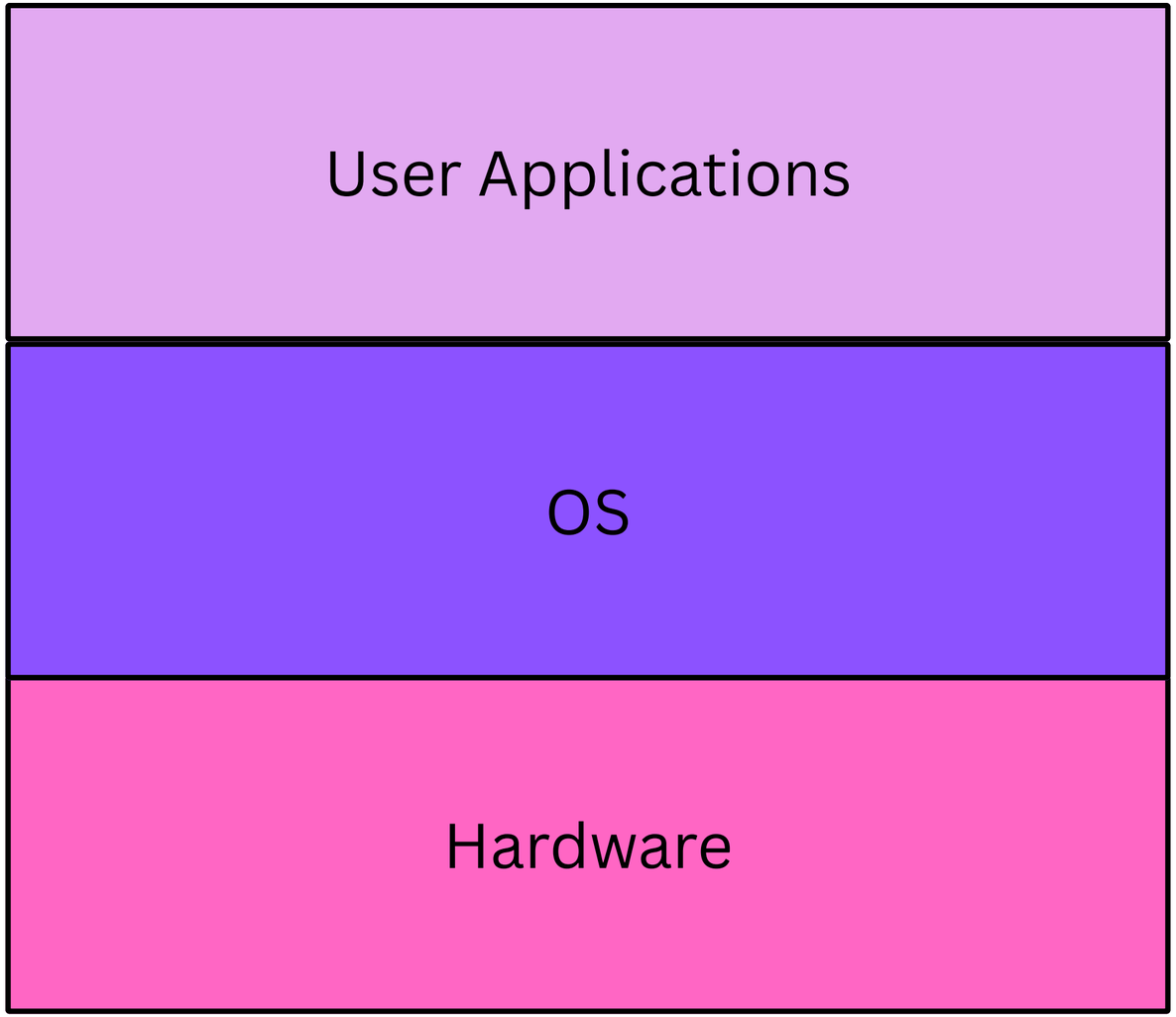

this is NOT on my ethernet network!
MAC A
IP A
MAC B
MAC C
IP B
MAC D
MAC E
MAC F
MAC G
MAC H
IP C
MAC H
IP C
Data
MAC R

There’s WAY more to networking…

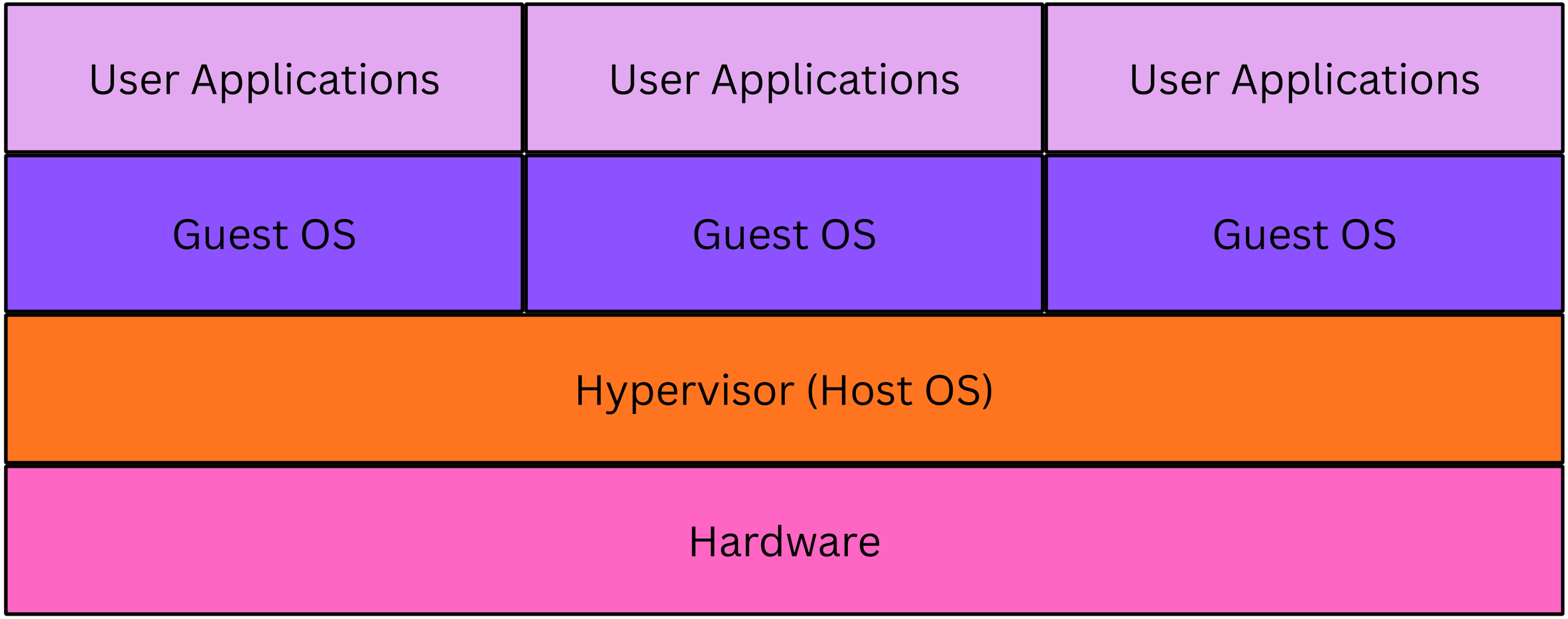
Take EECS 563!

Chapter 2:
What’s Virtualization?

Virtualization is running a “computer” inside a computer.



Standard Enviornment



Virtualized Environment

Now, onto the meat and potatoes! 🤩

What do I do?

“Azure”

“The cloud is just somebody else’s computer”

MILLIONS of machines

TEMU

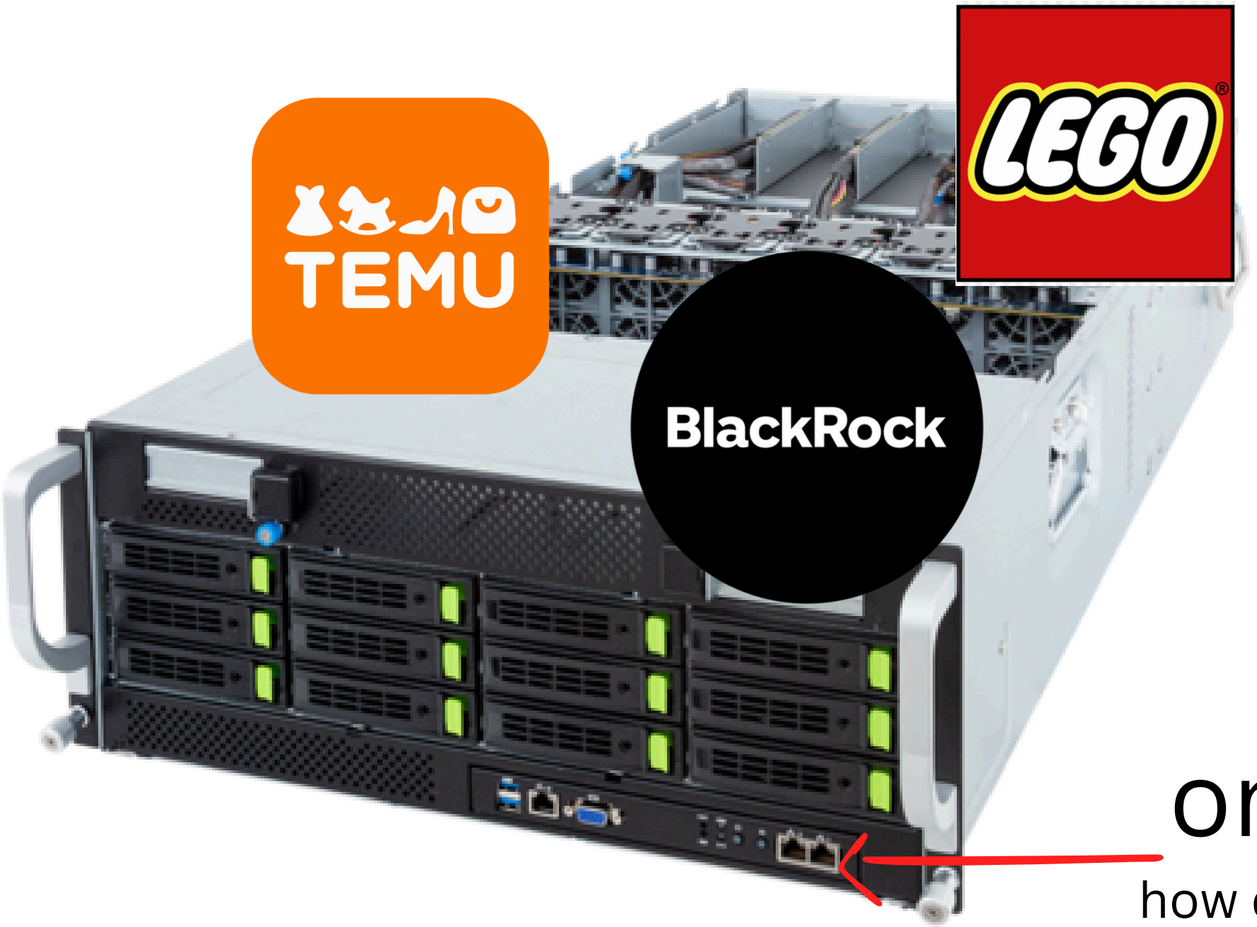
BlackRock

LEGO

Walmart
Save money. Live better.

SAP

Epic

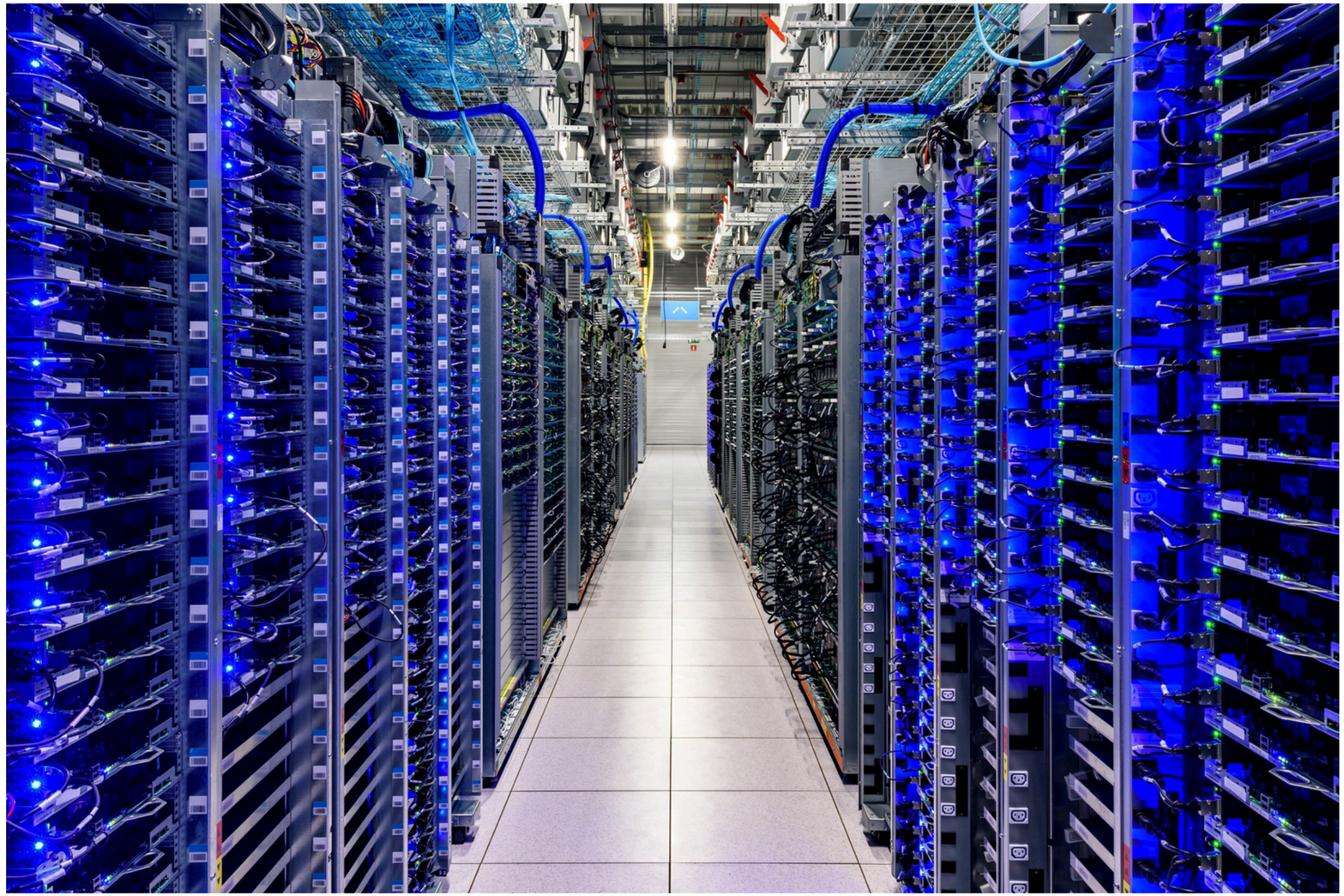


only 1* NIC!

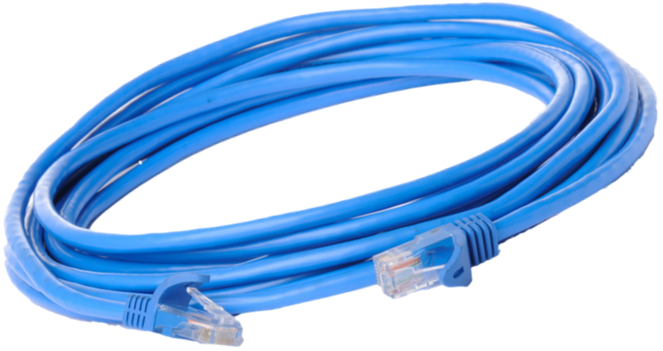
how do we do networking....

Using *virtualization*, many companies can share the same server

“Azure Core”

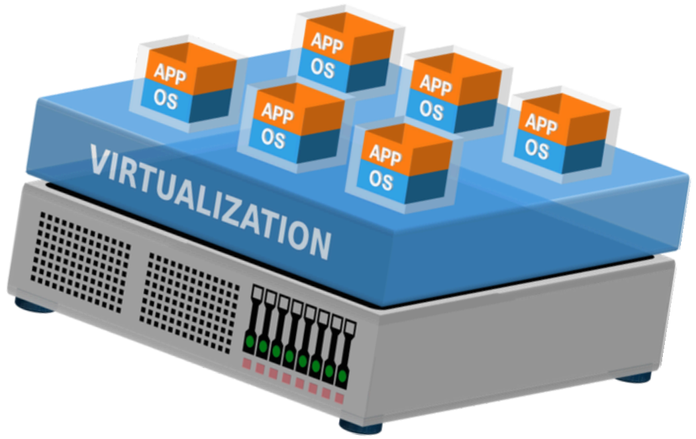


“Azure Core”



+ Anything else that makes the cloud function

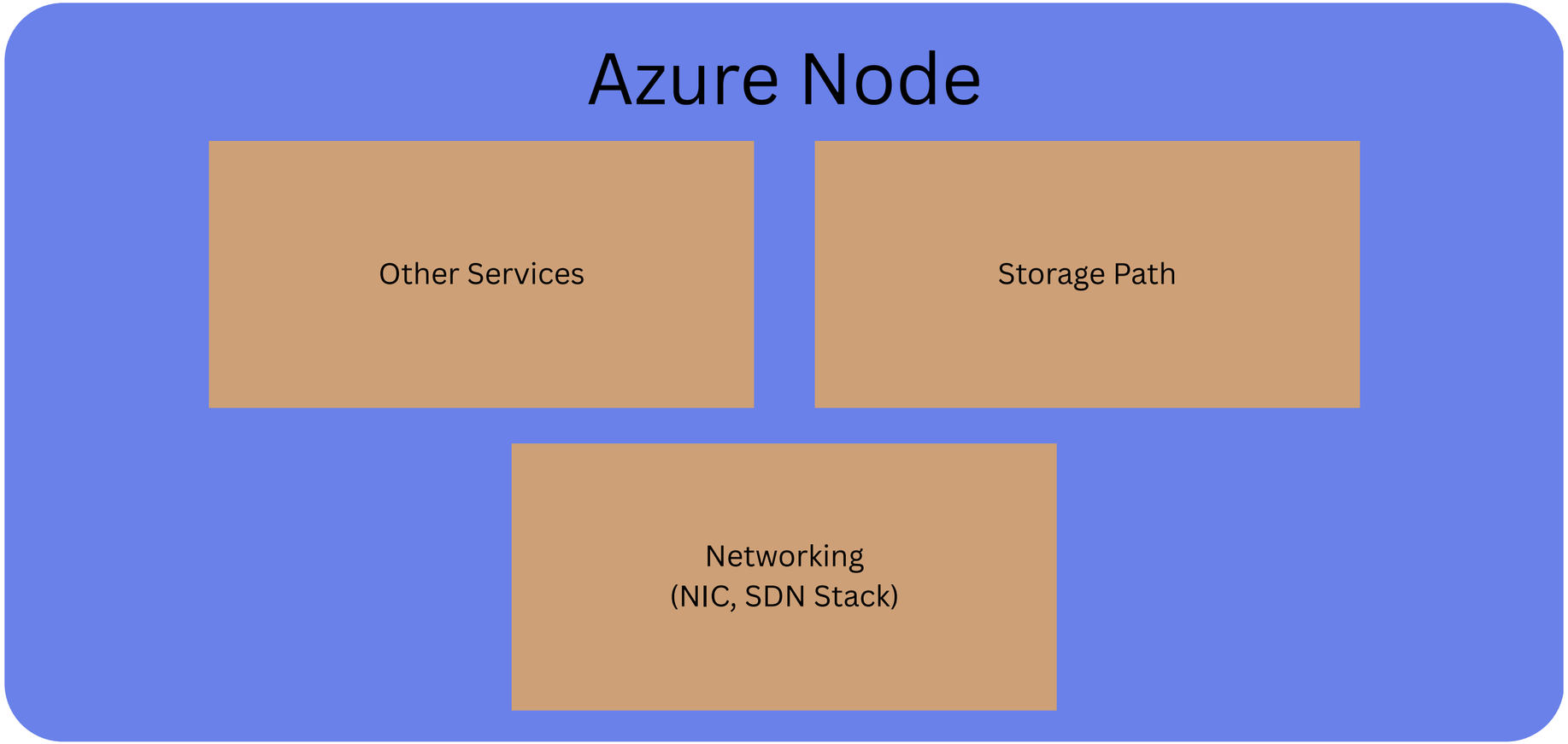
“Azure Compute”



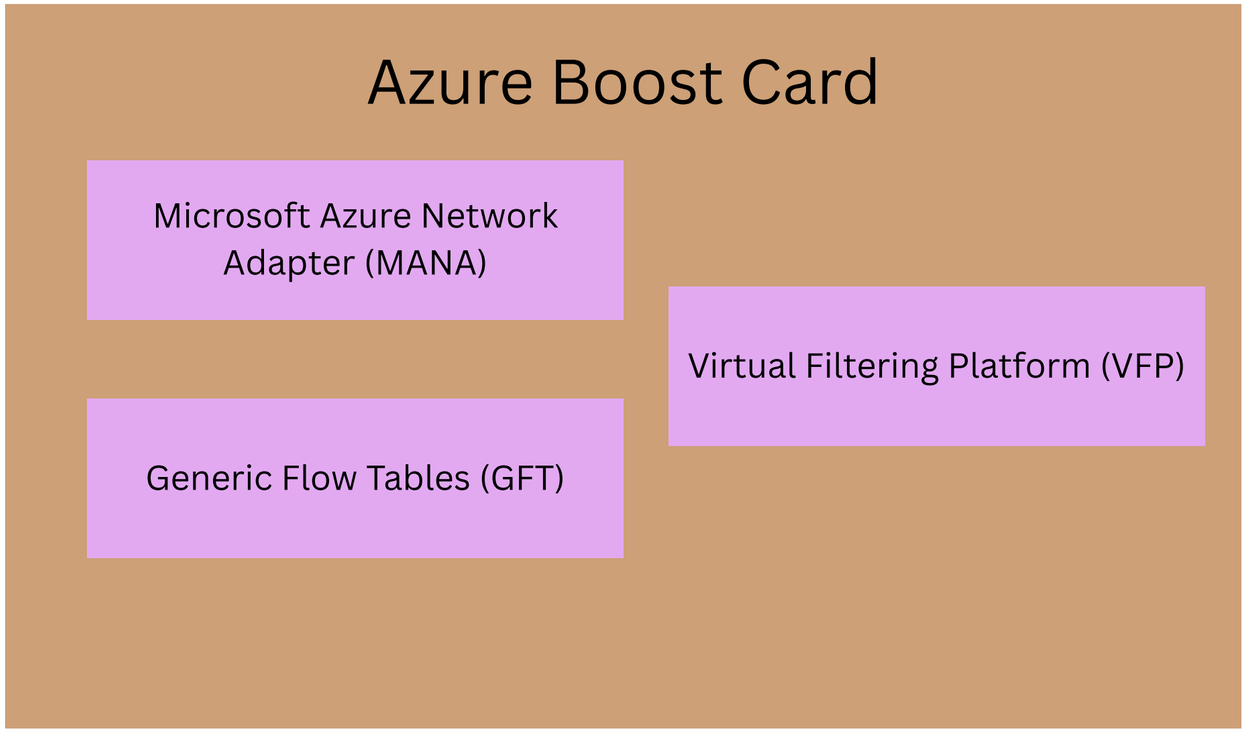
Microsoft
Hyper-V

+ Everything
that makes
Azure work at a
blade-level

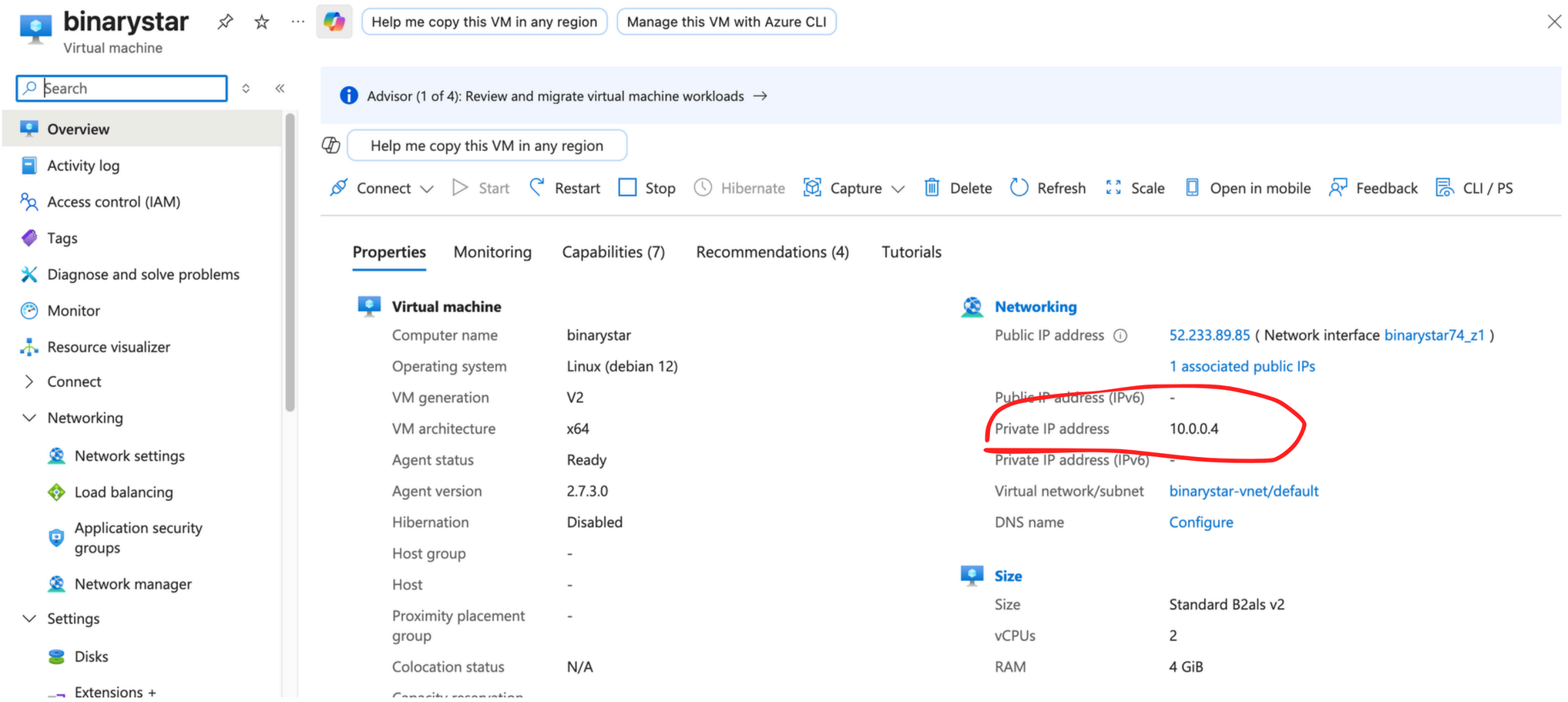
“Compute Node + Host”



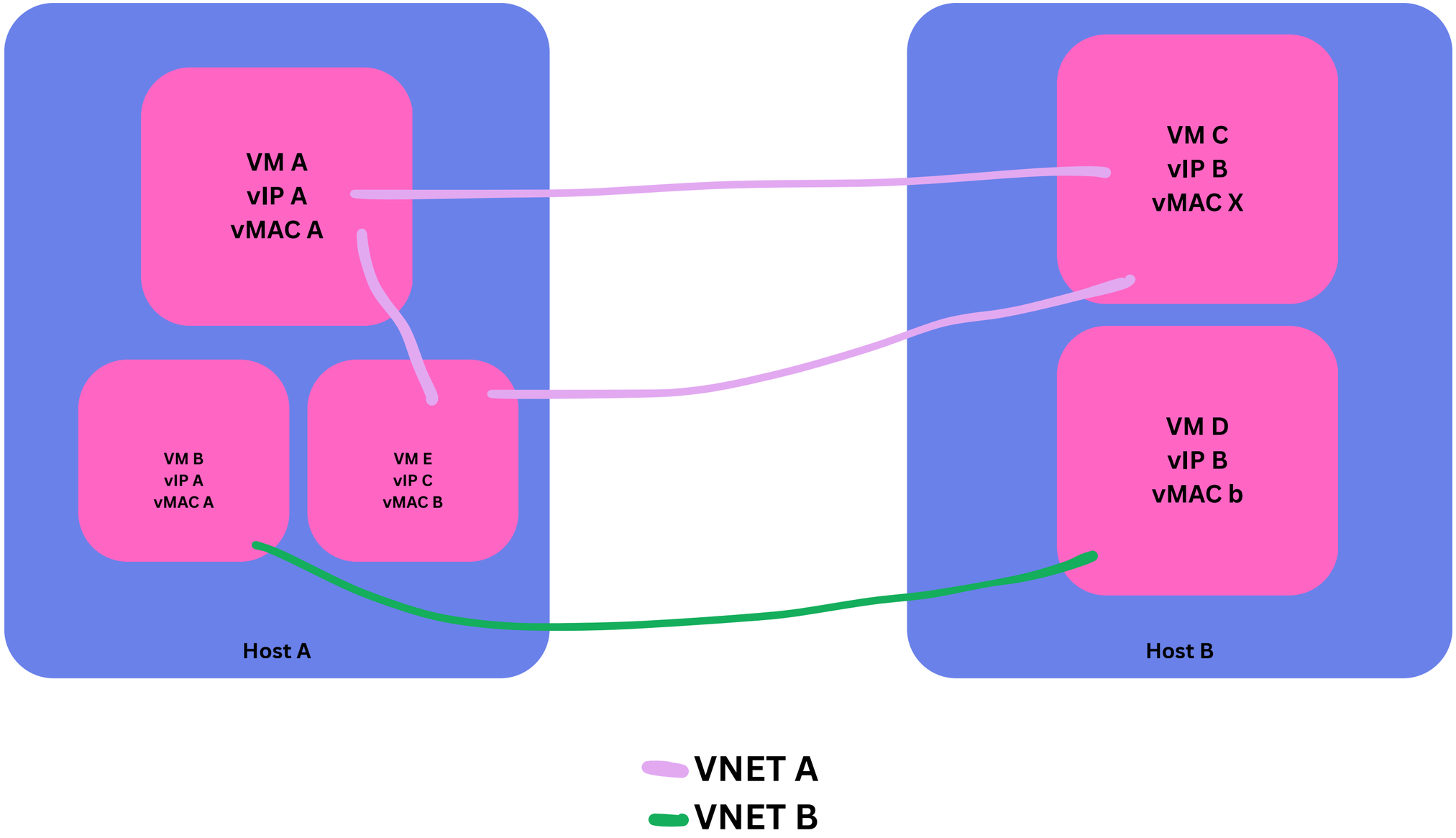
“Host Networking”



We can virtualize networks the same way we virtualize compute hardware.

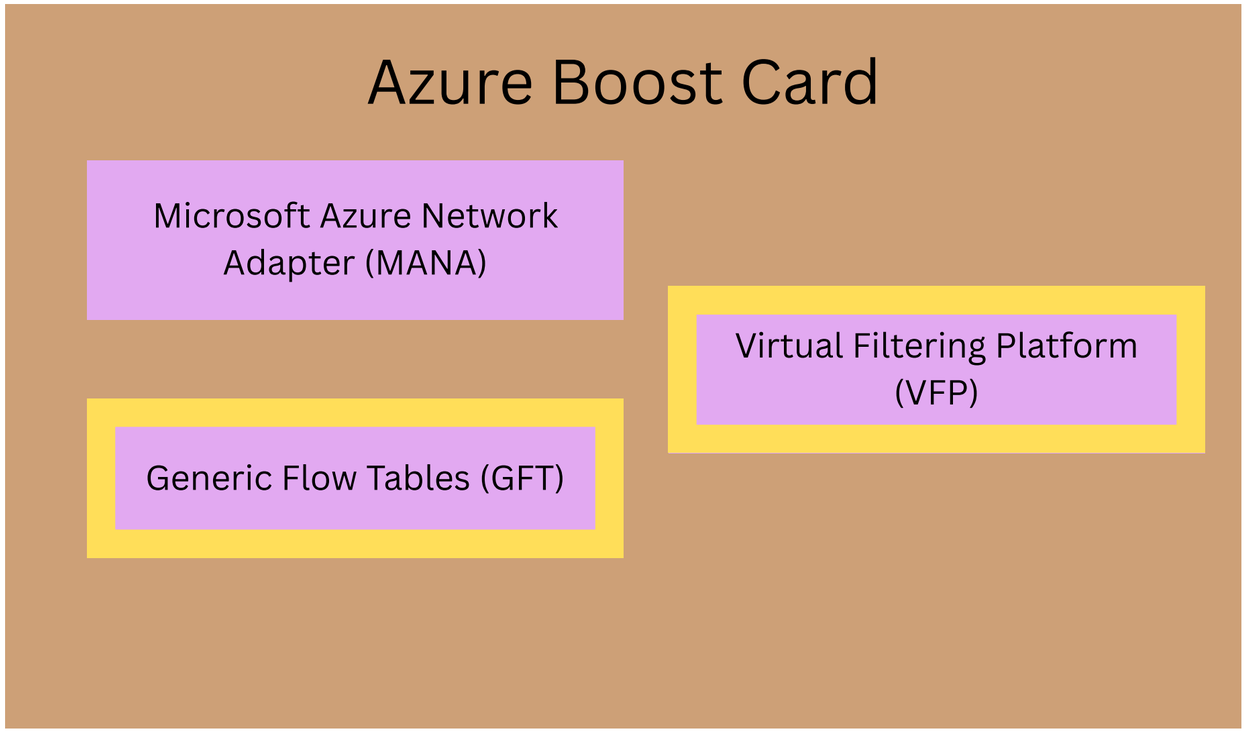


An Example


VM A
vIP A
vMAC A
VM B
vIP A
vMAC A
VM E
vIP C
vMAC B
Host A
VM C
vIP B
vMAC X
VM D
vIP B
vMAC b
Host B
VNET A
VNET B

How do we do this? 🧐

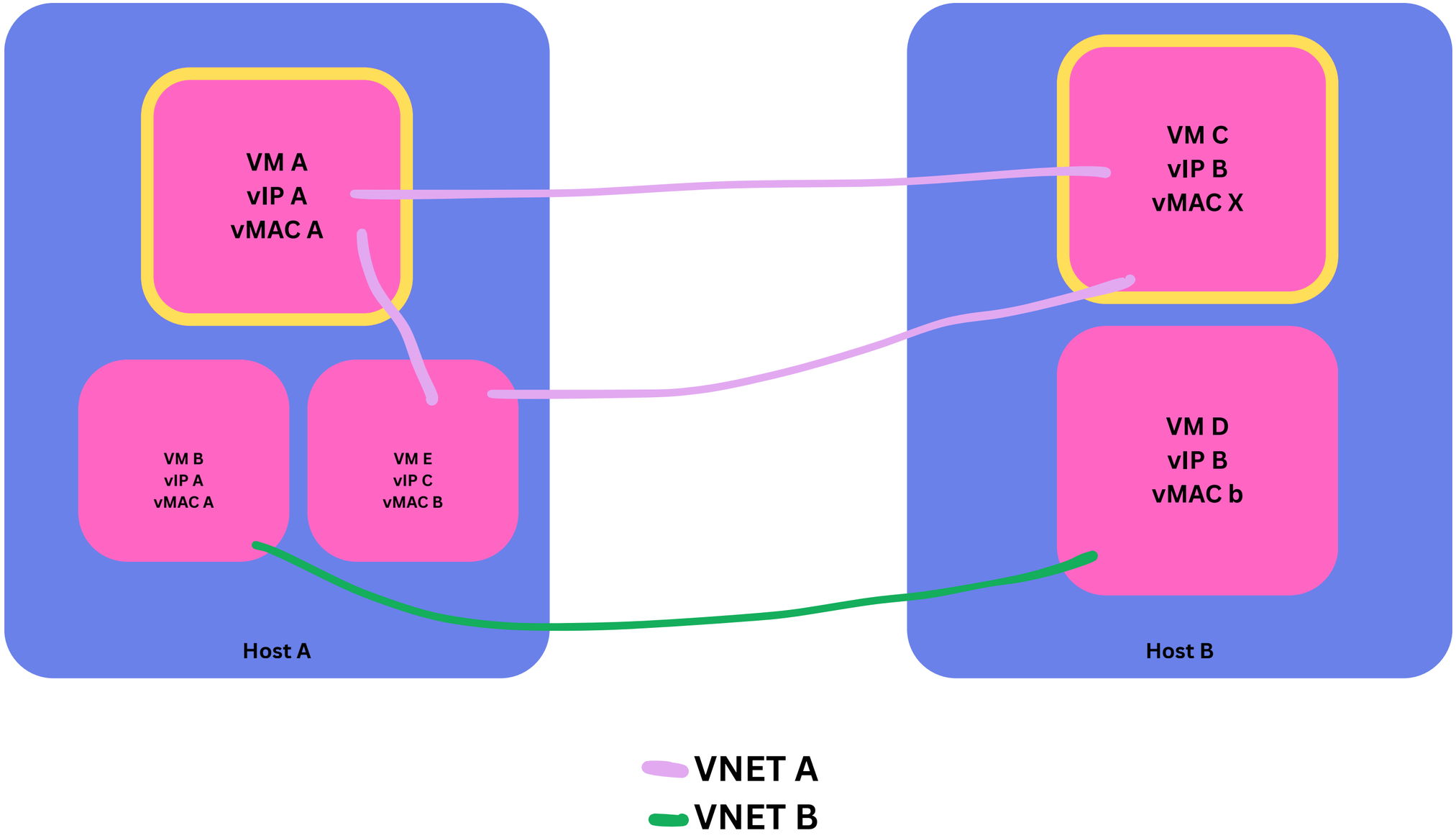
Software-**D**efined **A**ccelerated **N**etworking (SDAN)



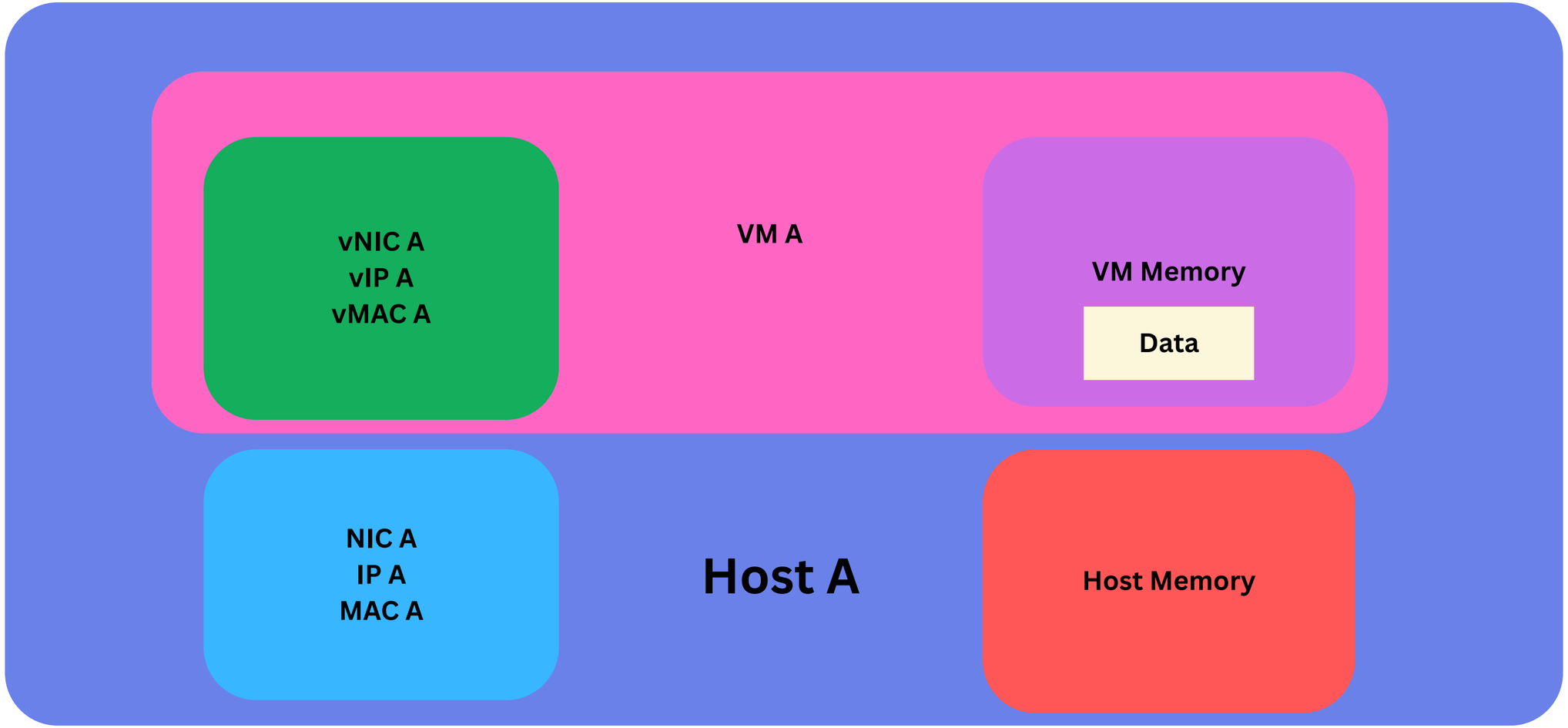
Is anything stopping us from just… adding more headers?

ABSOLUTELY NOT!!! LETS GO CRAZY!!! 💃

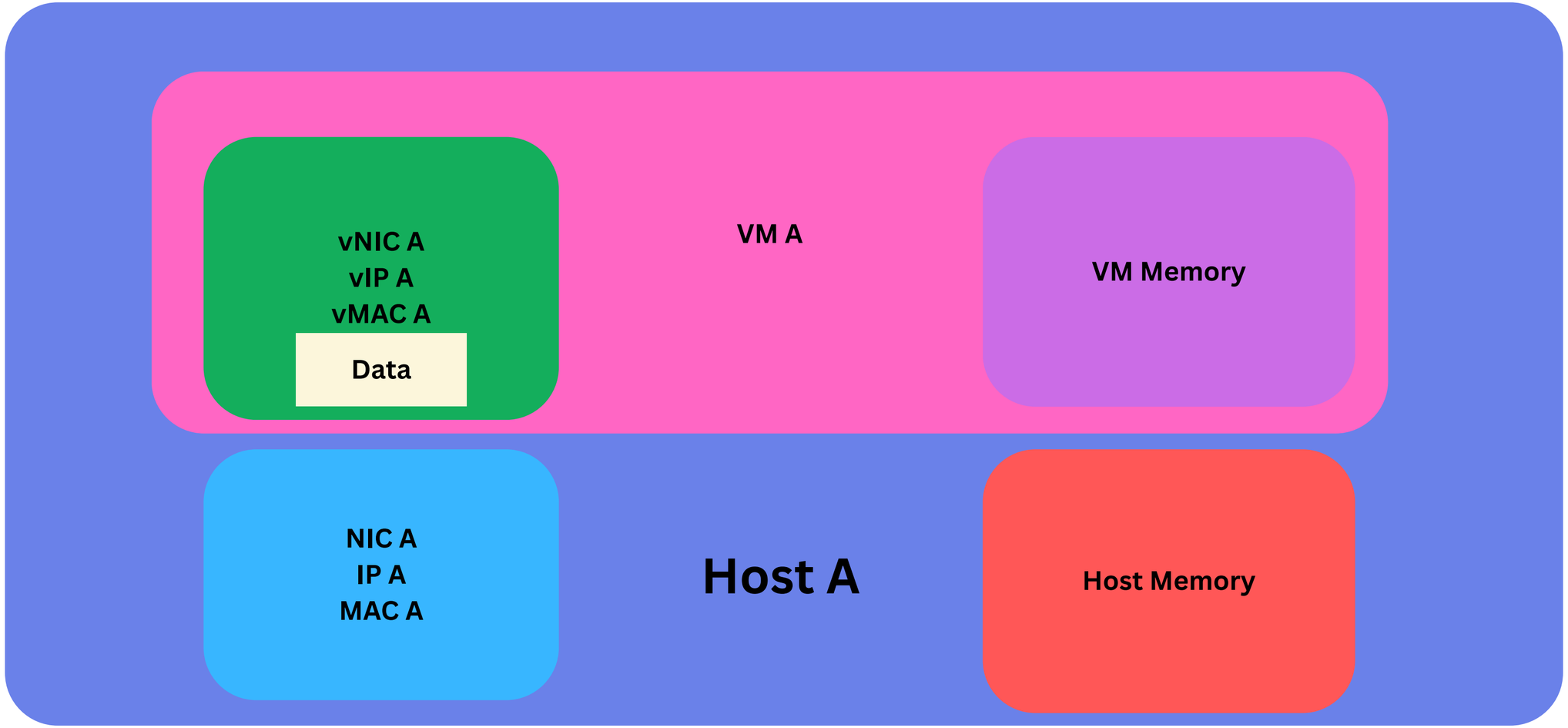
An Example


VM A
vIP A
vMAC A
VM B
vIP A
vMAC A
VM E
vIP C
vMAC B
Host A
VM C
vIP B
vMAC X
VM D
vIP B
vMAC b
Host B
VNET A
VNET B

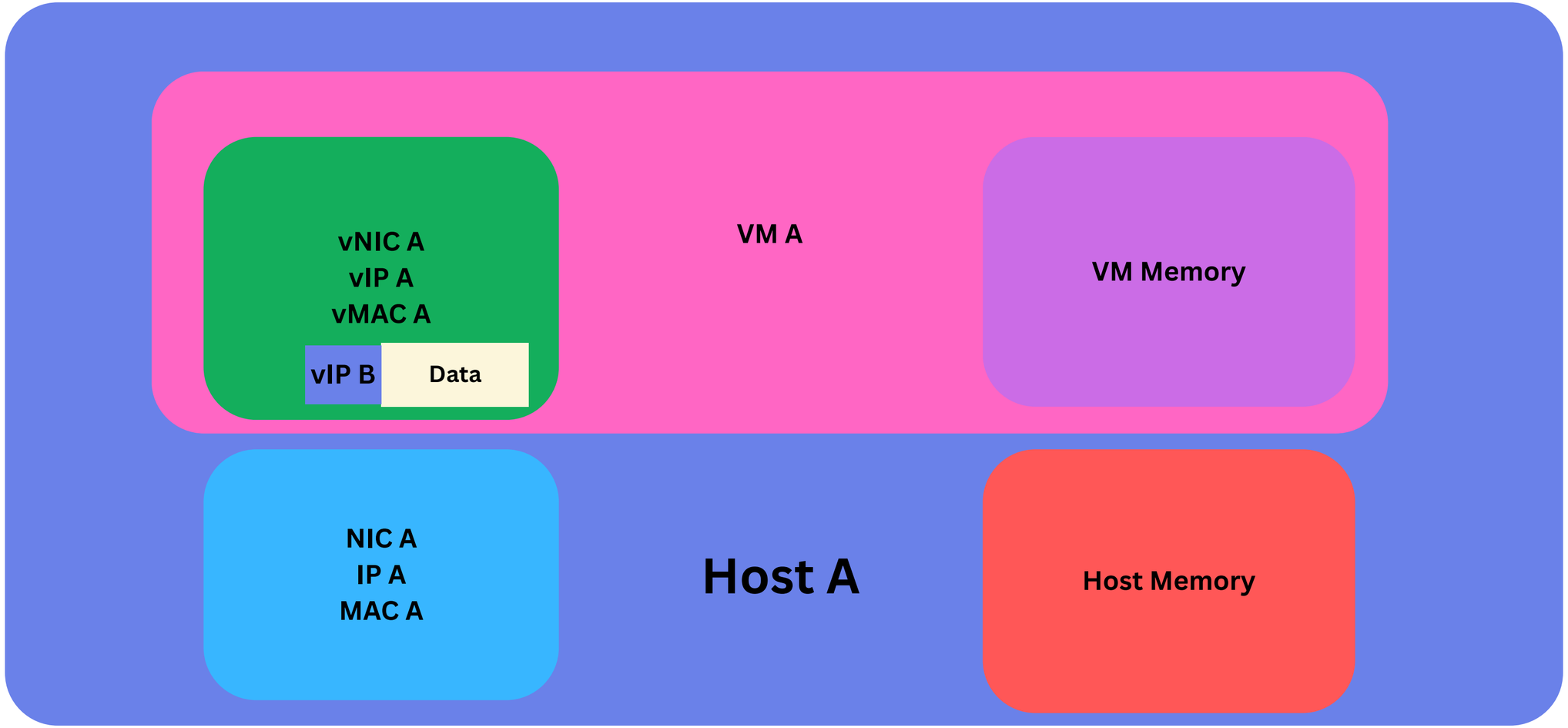
An Example


vNIC A
vIP A
vMAC A
VM A
VM Memory
Data
NIC A
IP A
MAC A
Host A
Host Memory

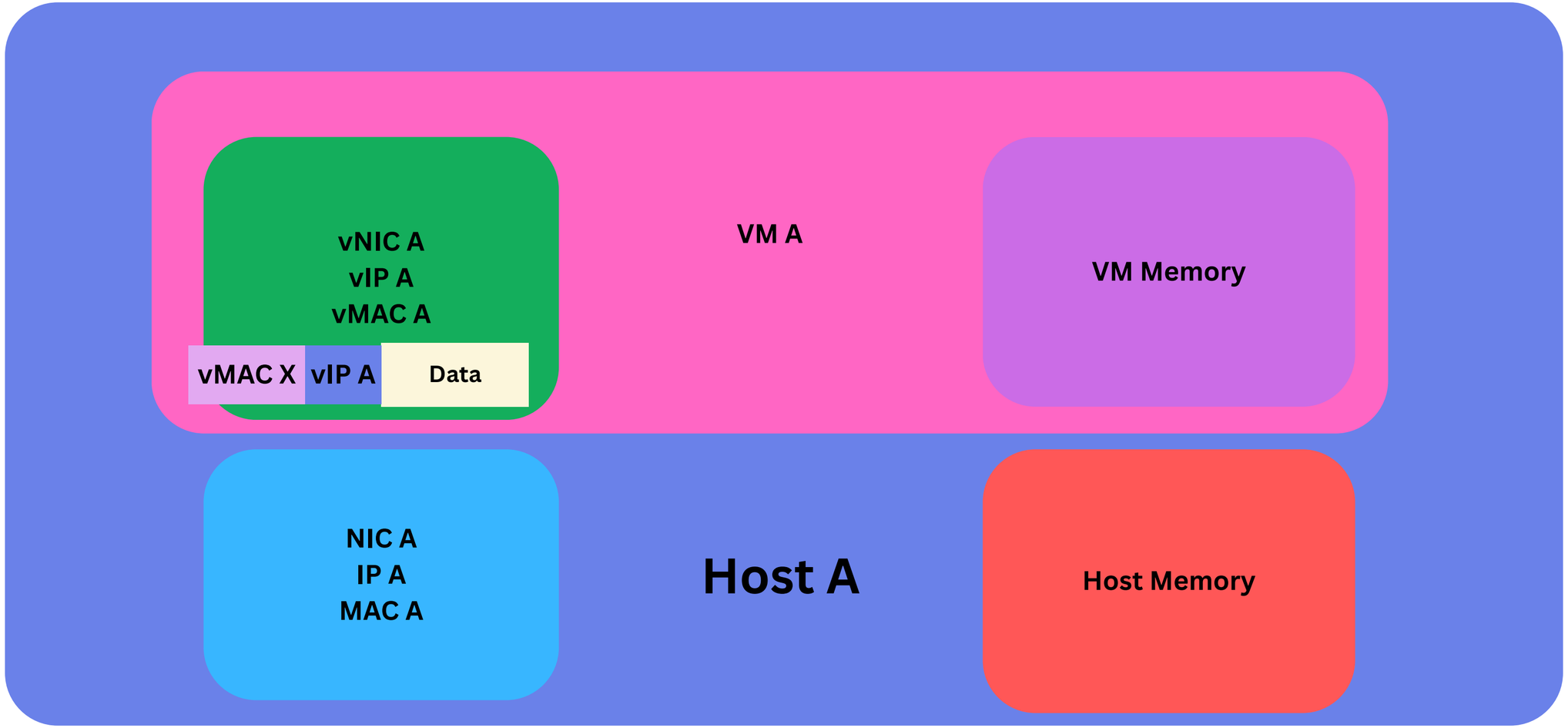
An Example


vNIC A
vIP A
vMAC A
Data
VM A
VM Memory
NIC A
IP A
MAC A
Host A
Host Memory

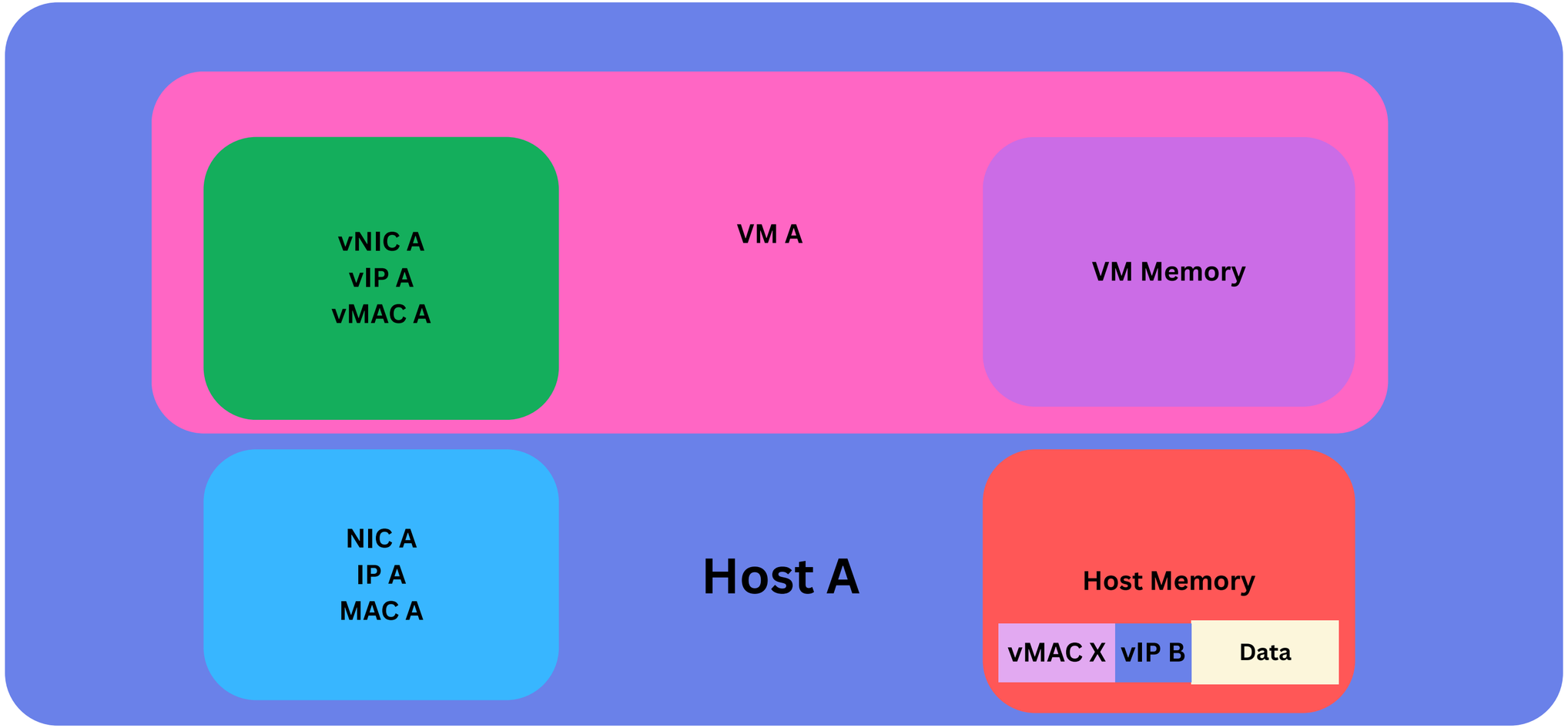
An Example


vNIC A
vIP A
vMAC A
vIP B
Data
VM A
VM Memory
NIC A
IP A
MAC A
Host A
Host Memory

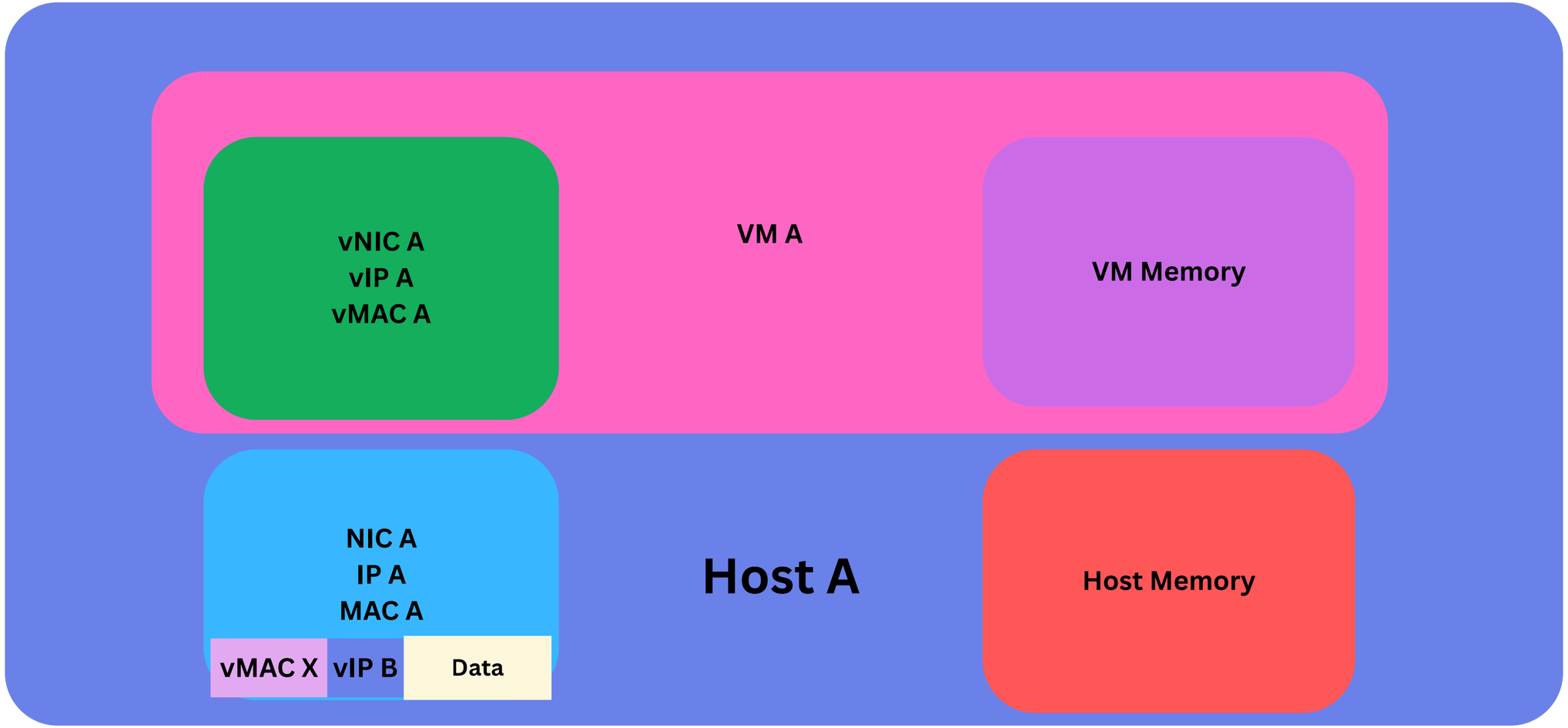
An Example


vNIC A
vIP A
vMAC A
vMAC X
vIP A
Data
VM A
VM Memory
NIC A
IP A
MAC A
Host A
Host Memory

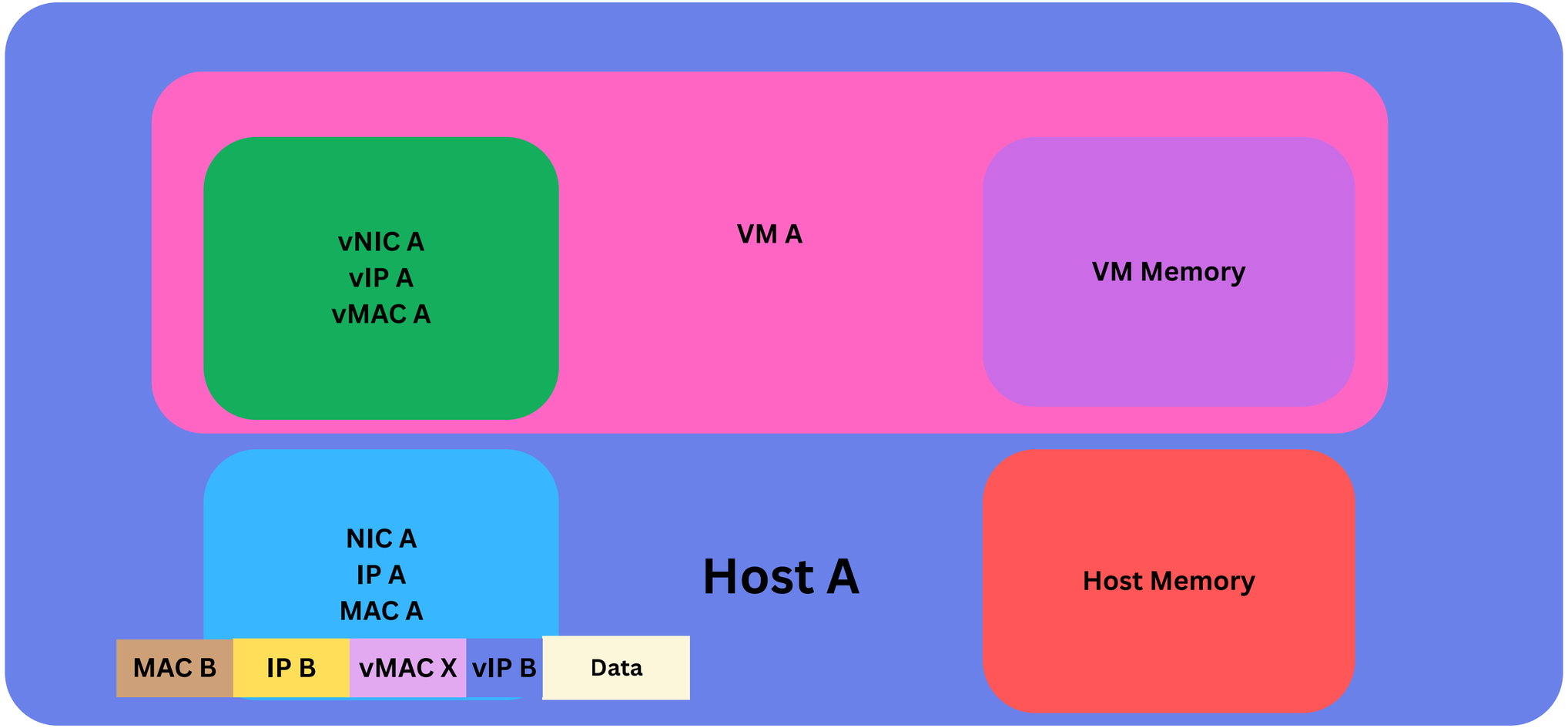
An Example


vNIC A
vIP A
vMAC A
VM A
VM Memory
NIC A
IP A
MAC A
Host A
Host Memory
vMAC X
vIP B
Data

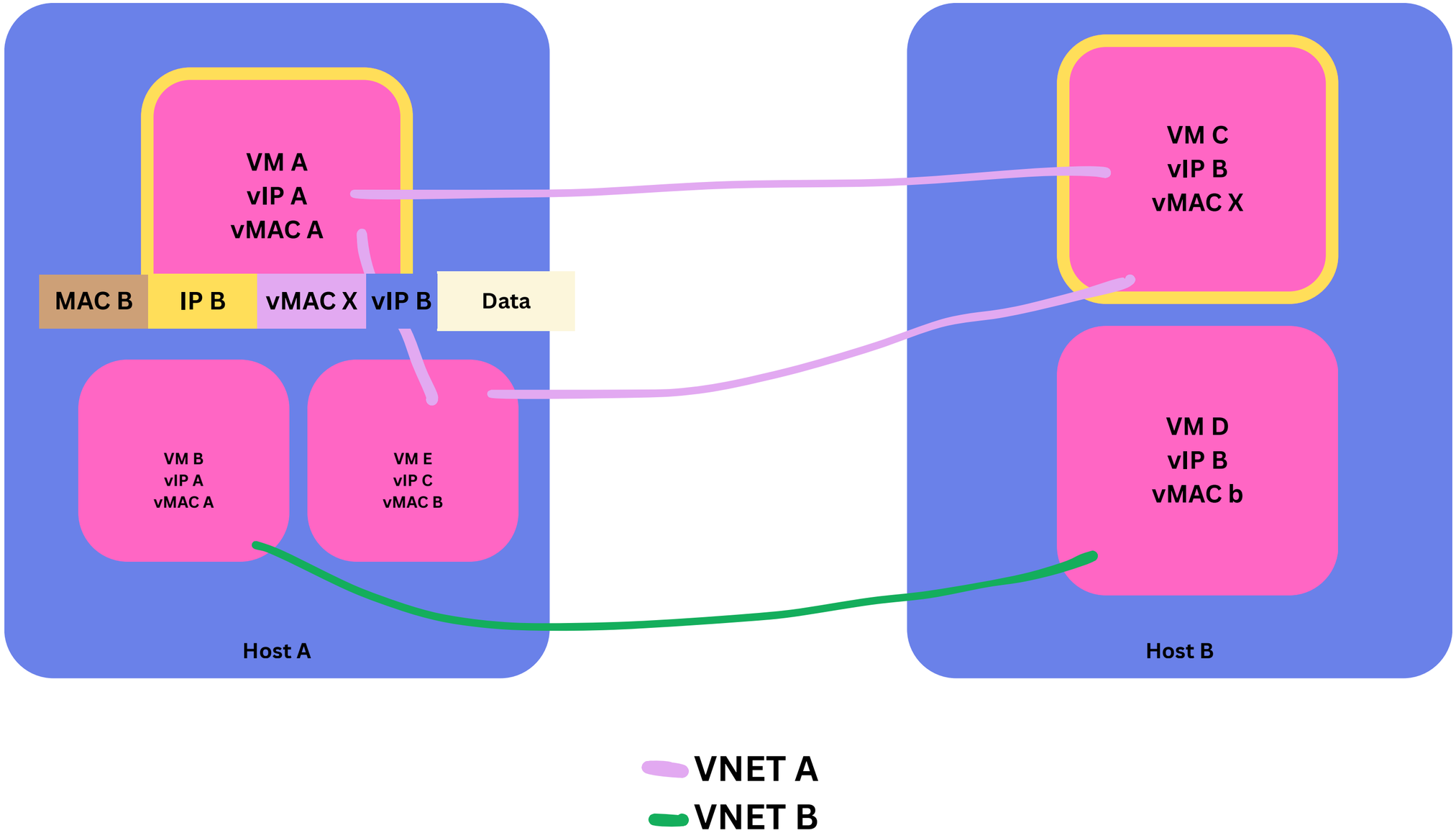
An Example


vNIC A
vIP A
vMAC A
VM A
VM Memory
NIC A
IP A
MAC A
vMAC X
vIP B
Data
Host A
Host Memory

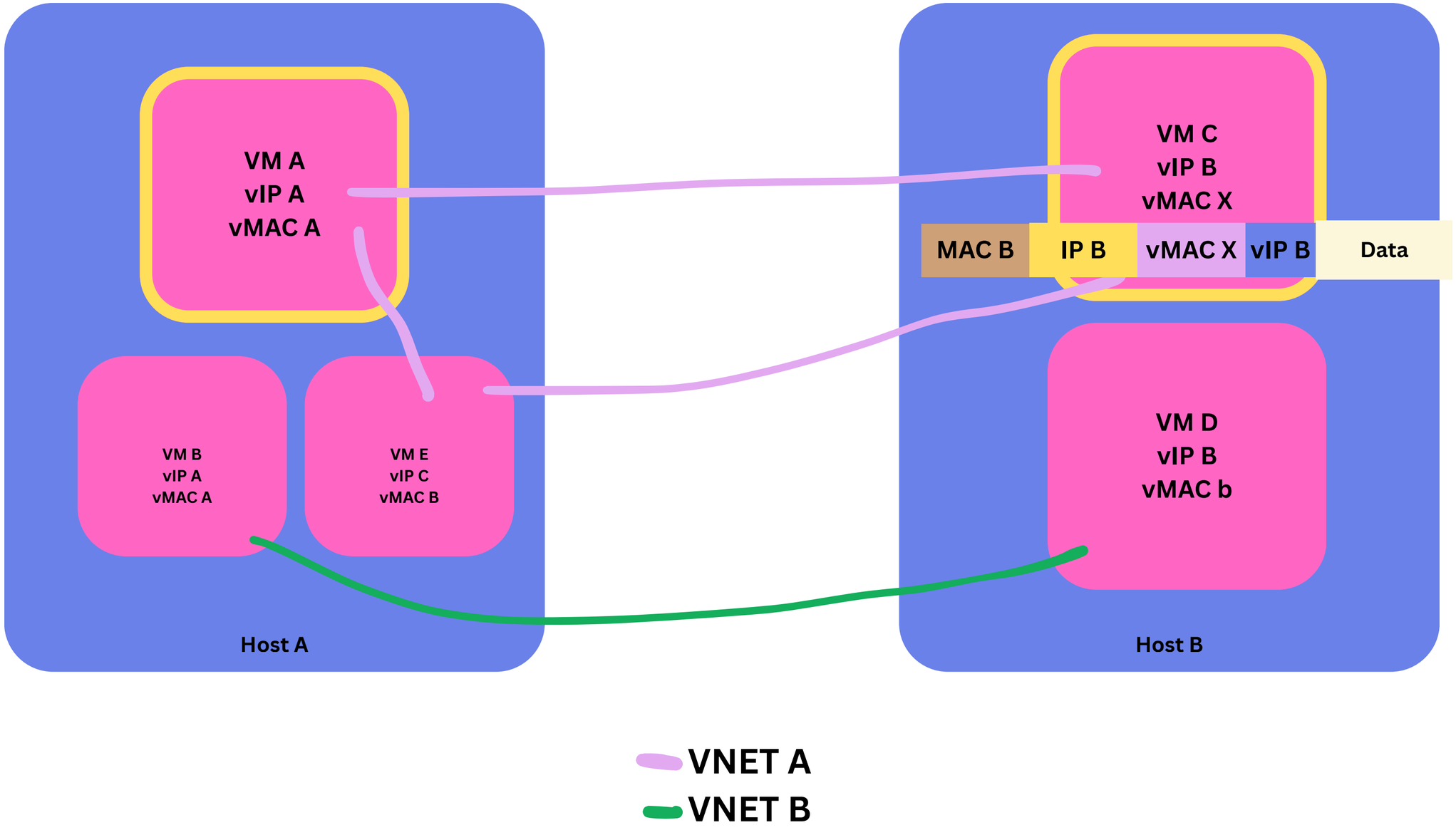
An Example


vNIC A
vIP A
vMAC A
VM A
VM Memory
NIC A
IP A
MAC A
Host A
Host Memory
MAC B
IP B
vMAC X
vIP B
Data

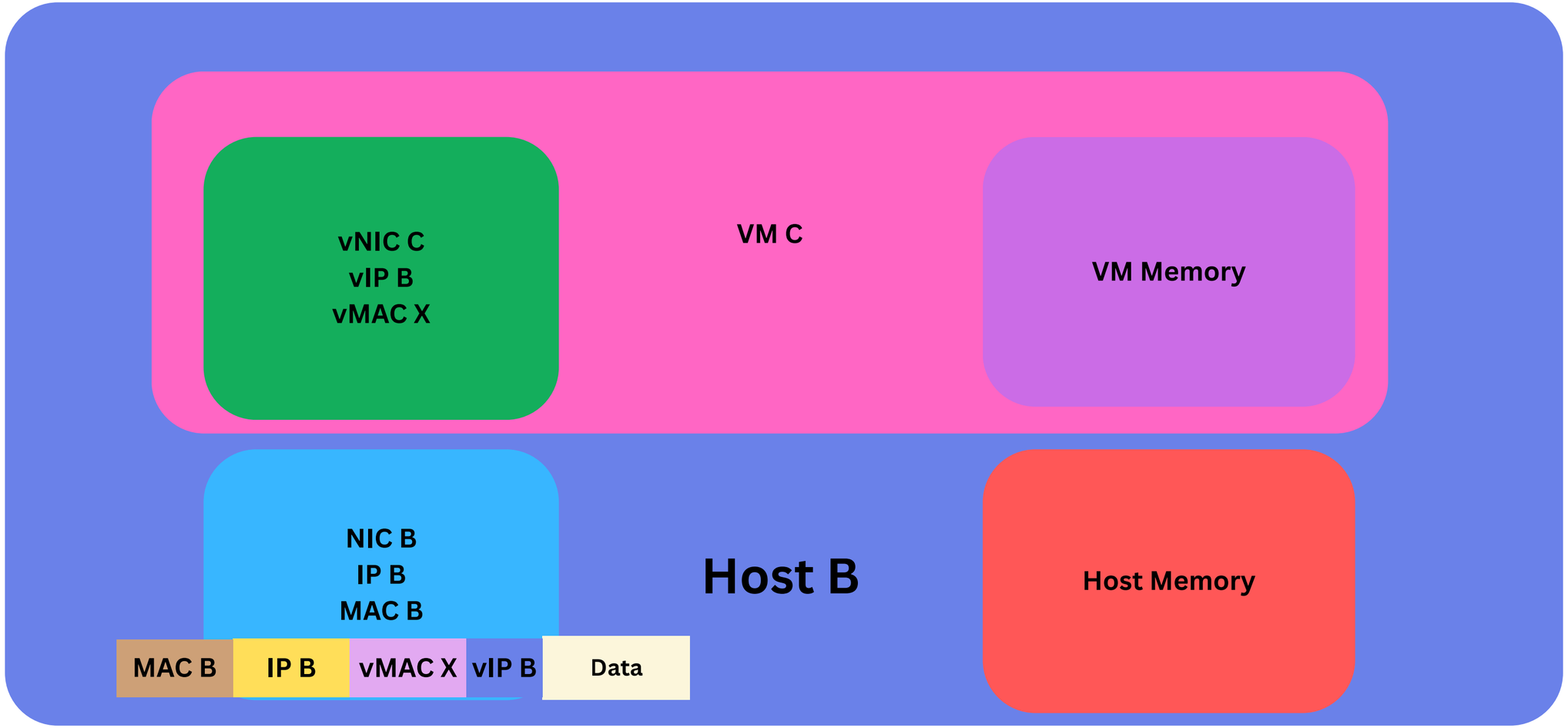
An Example


VM A
vIP A
vMAC A
MAC B
IP B
vMAC X
vIP B
Data
VM B
vIP A
vMAC A
VM E
vIP C
vMAC B
Host A
VM C
vIP B
vMAC X
VM D
vIP B
vMAC b
Host B
VNET A
VNET B

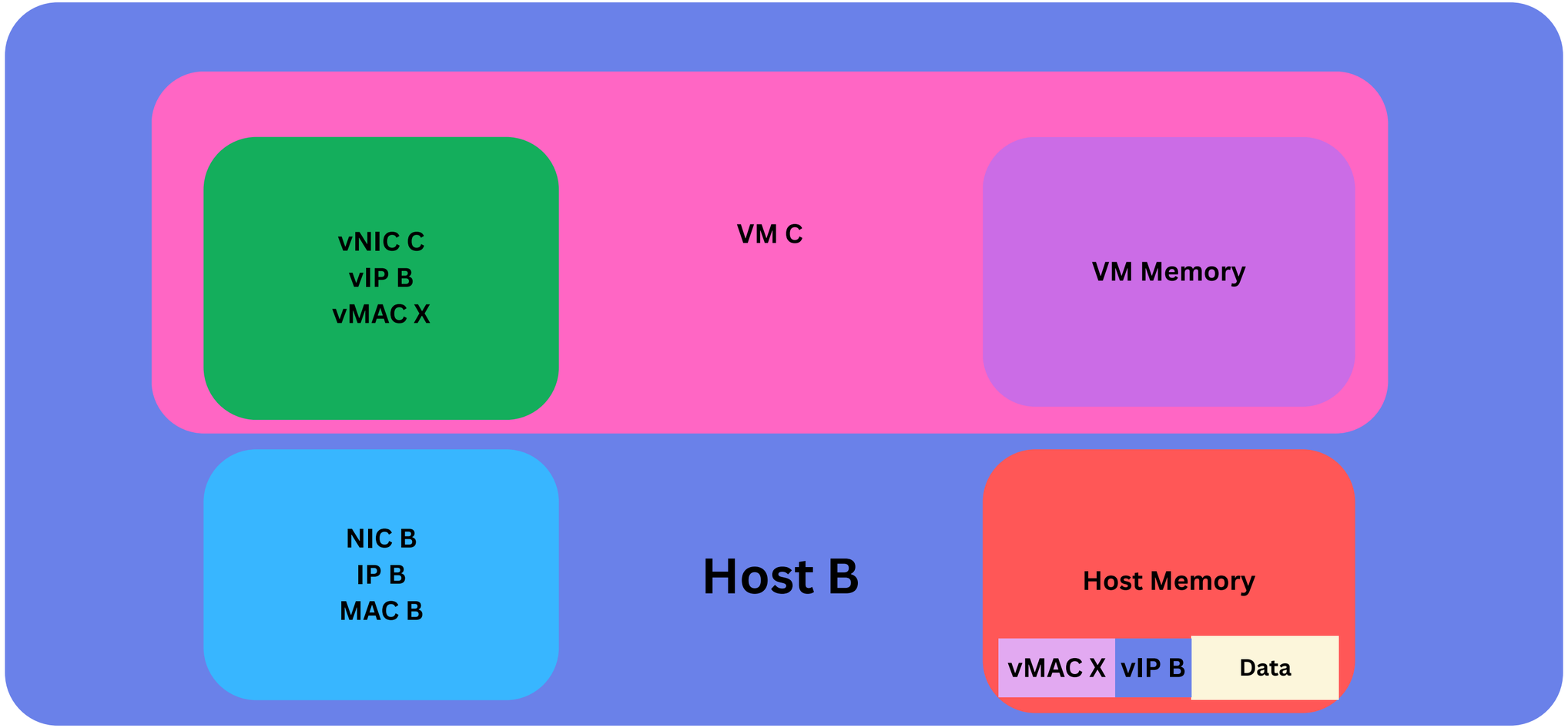
An Example


VM A
vIP A
vMAC A
VM B
vIP A
vMAC A
VM E
vIP C
vMAC B
Host A
VM C
vIP B
vMAC X
MAC B
IP B
vMAC X
vIP B
Data
VM D
vIP B
vMAC b
Host B
VNET A
VNET B

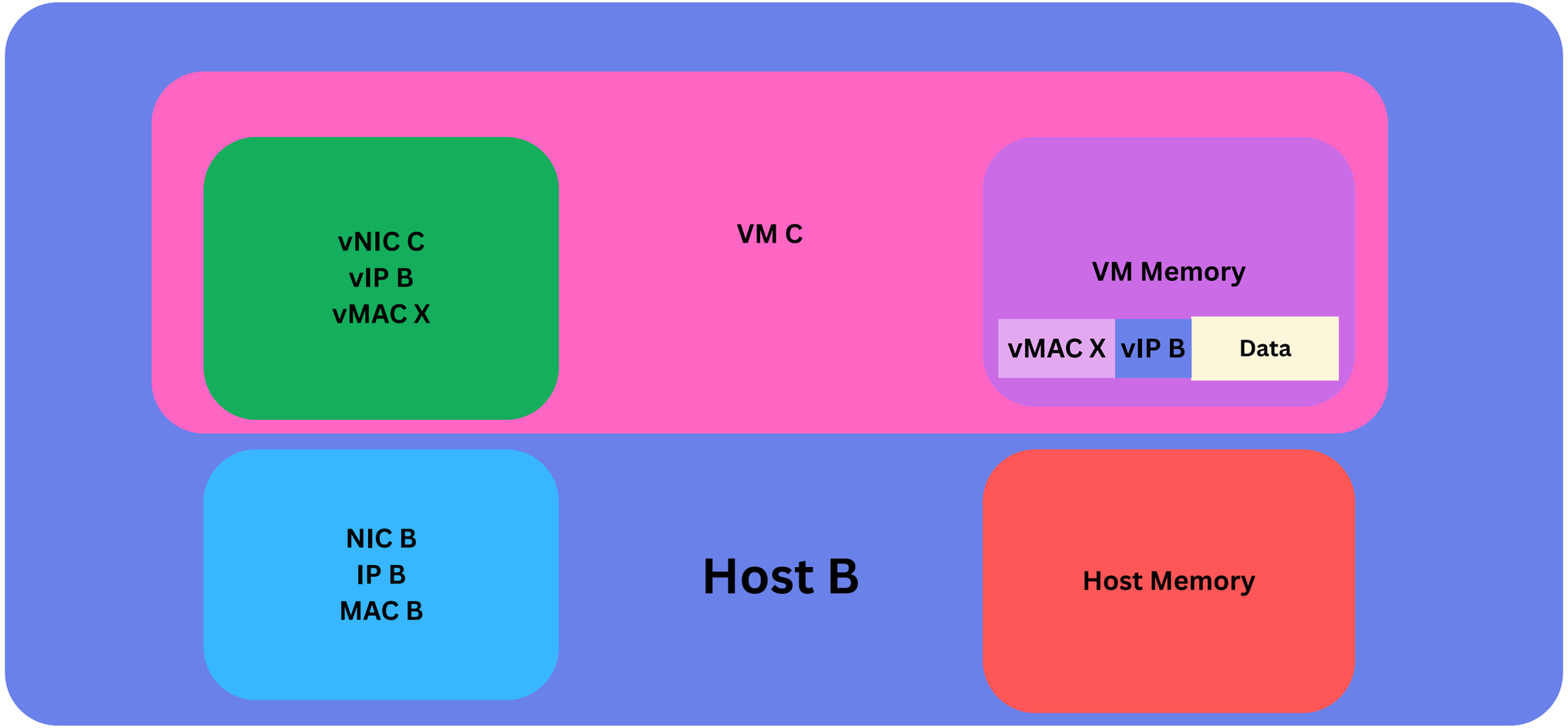
An Example


vNIC C
vIP B
vMAC X
VM C
VM Memory
NIC B
IP B
MAC B
Host B
Host Memory
MAC B
IP B
vMAC X
vIP B
Data

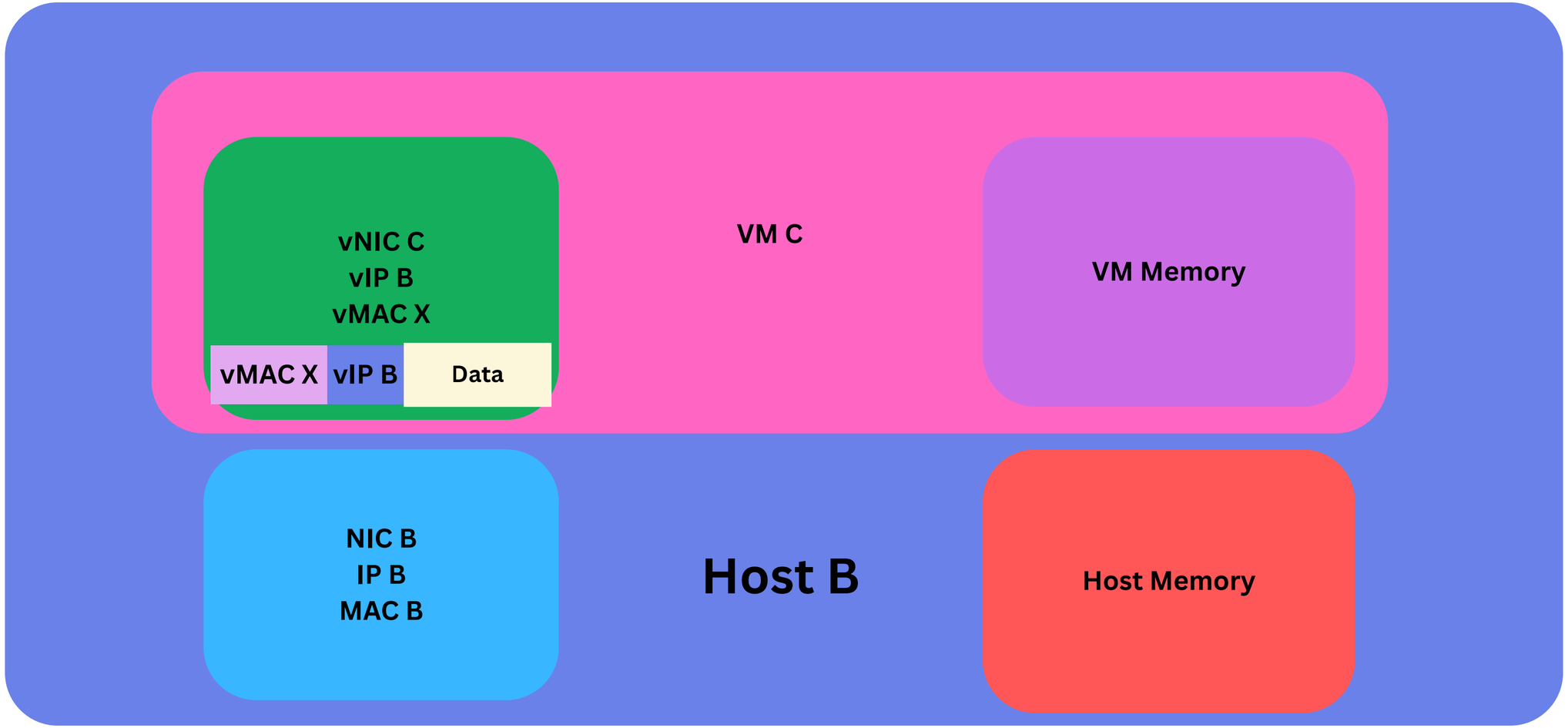
An Example


vNIC C
vIP B
vMAC X
VM C
VM Memory
NIC B
IP B
MAC B
Host B
Host Memory
vMAC X
vIP B
Data

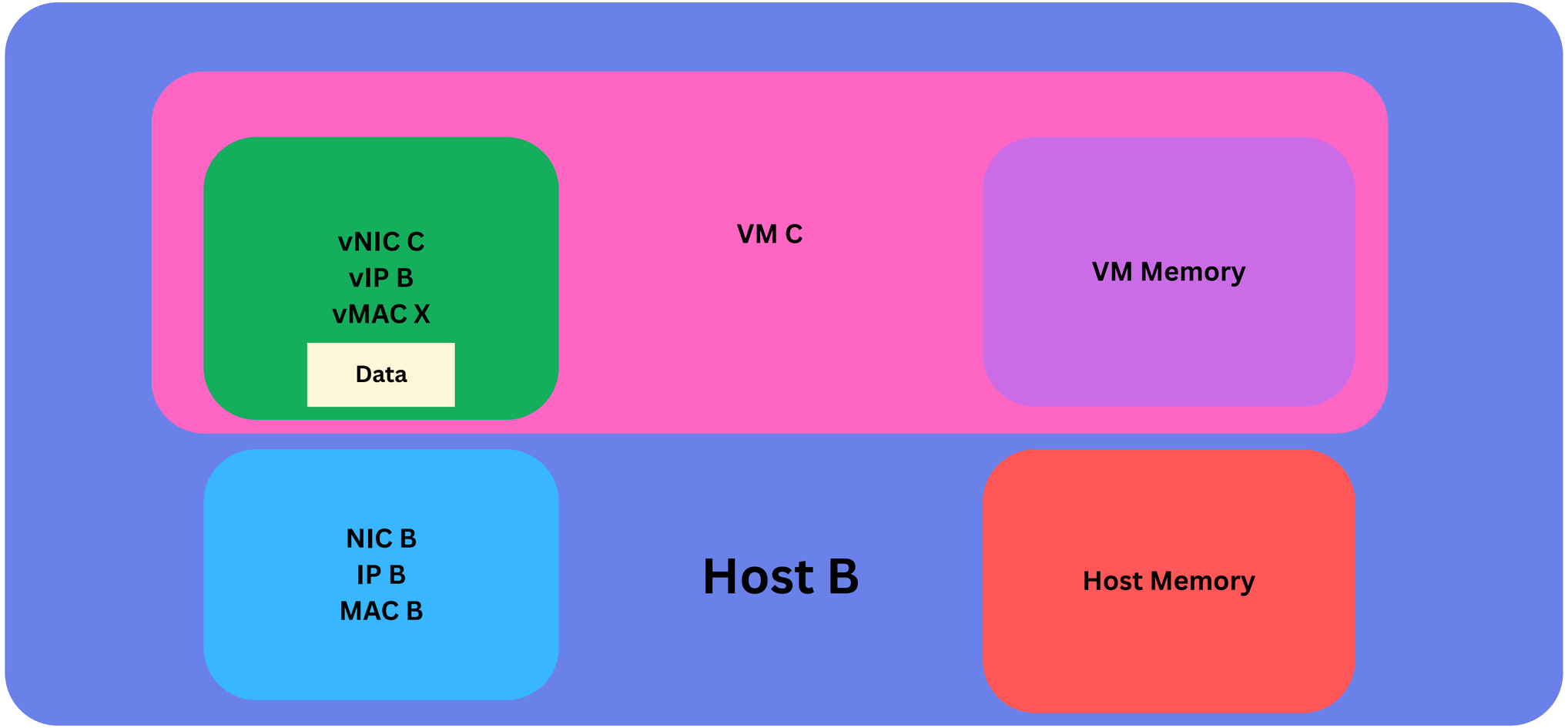
An Example


vNIC C
vIP B
vMAC X
VM C
VM Memory
vMAC X
vIP B
Data
NIC B
IP B
MAC B
Host B
Host Memory

An Example

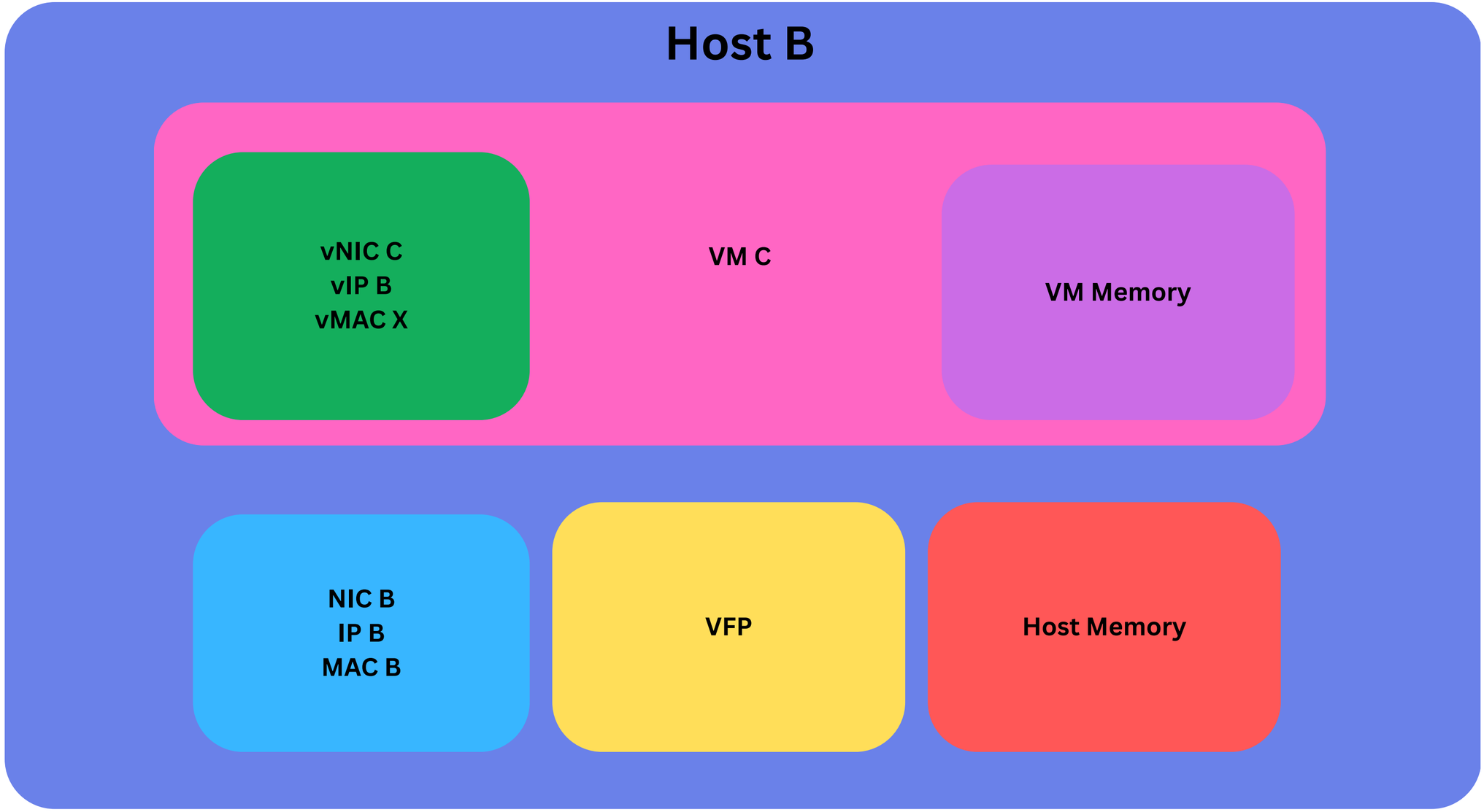

vNIC C
vIP B
vMAC X
vMAC X
vIP B
Data
VM C
VM Memory
NIC B
IP B
MAC B
Host B
Host Memory

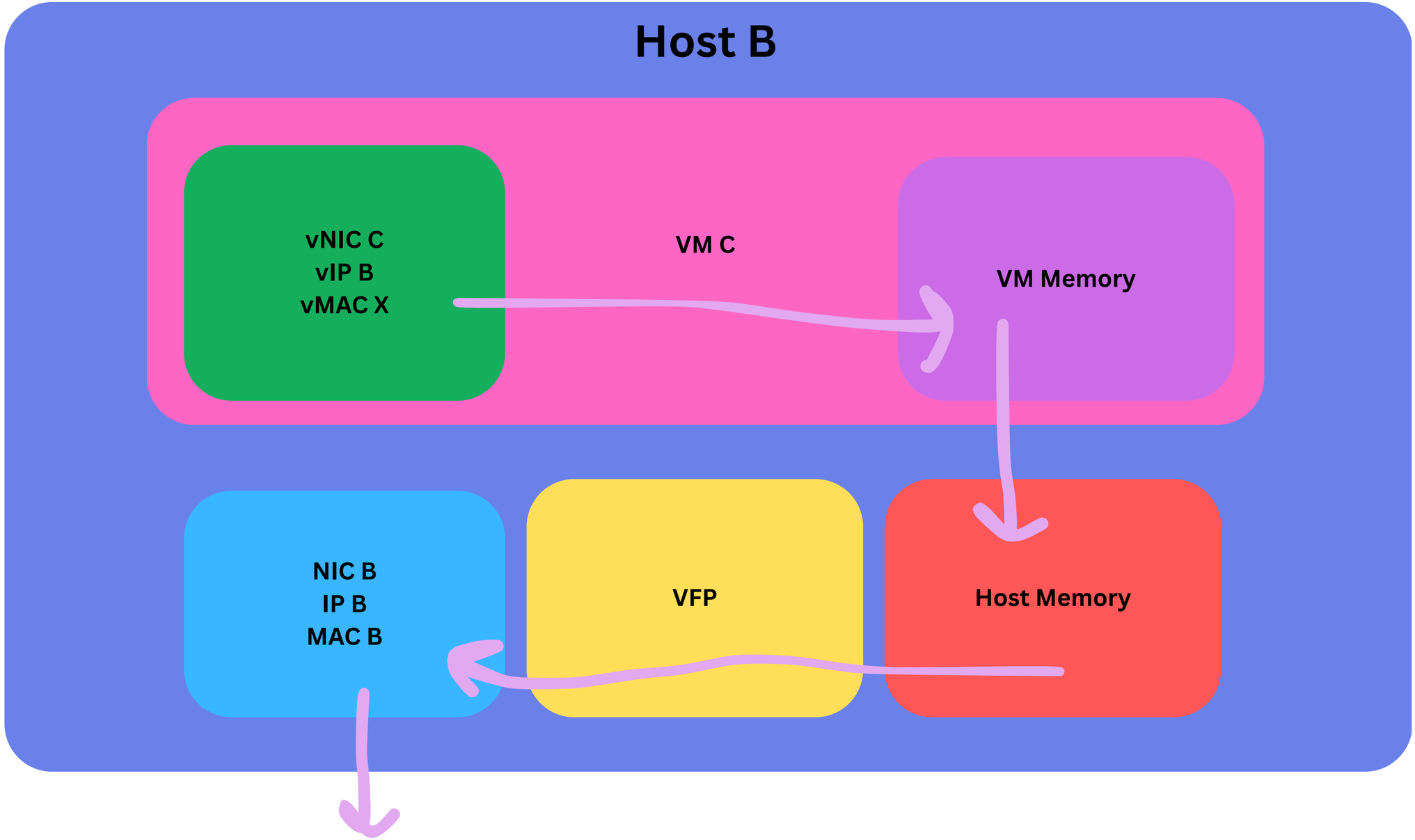
An Example


vNIC C
vIP B
vMAC X
Data
VM C
VM Memory
NIC B
IP B
MAC B
Host B
Host Memory

How do the extra headers get added?

The ✨**Virtual Filtering Platform (VFP)**✨


Host B
vNIC C
vIP B
vMAC X
VM C
VM Memory
NIC B
IP B
MAC B
VFP
Host Memory


Host B
vNIC C
vIP B
vMAC X
VM C
VM Memory
NIC B
IP B
MAC B
VFP
Host Memory

VFP isn’t just for VNETS:

- Load Balancers
- Firewalls
- Quality-of-Service
  - Giving some traffic priority over others
- Etc.

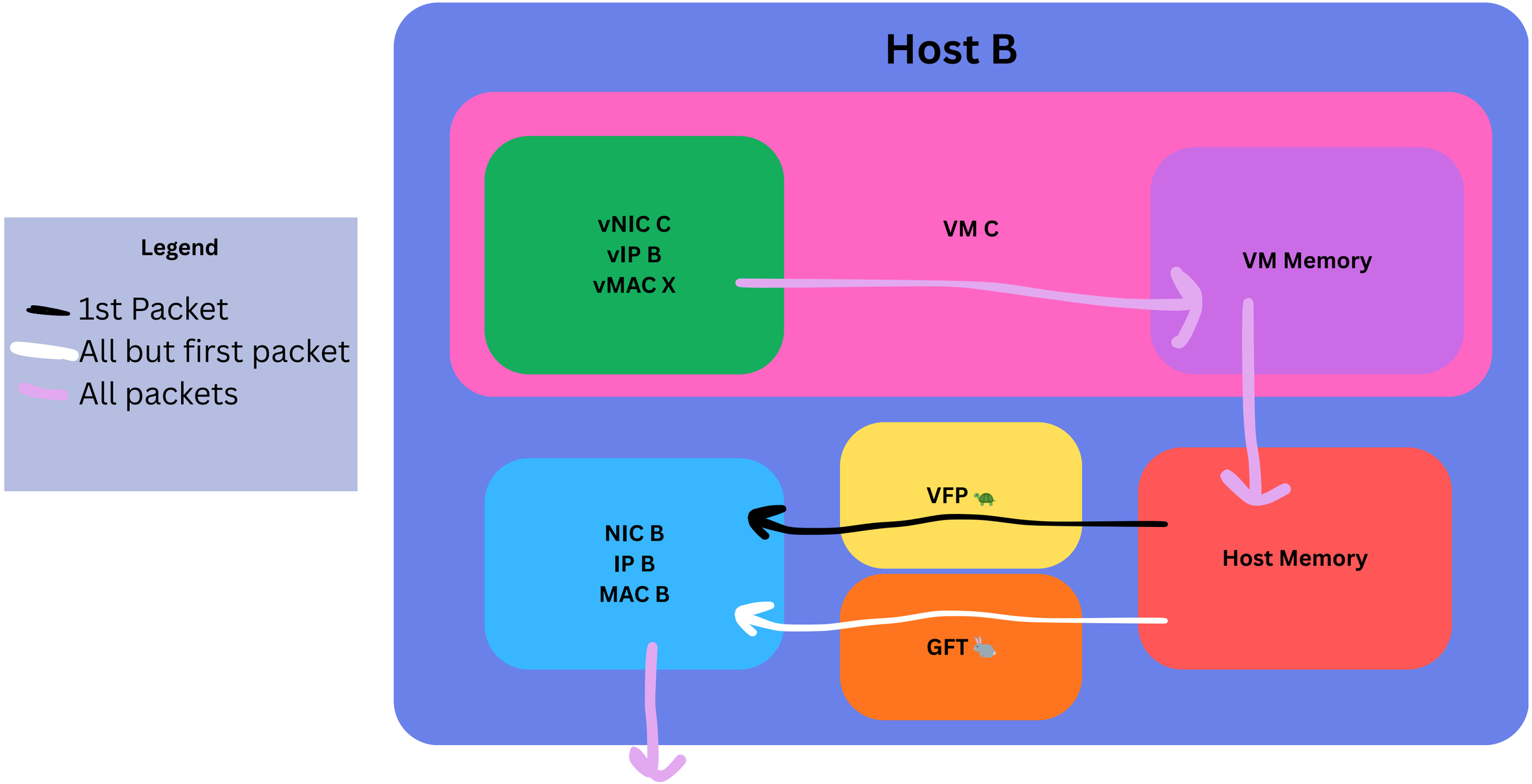
For a stream of packets[1], wouldn’t all this extra processing slow things down?? 🐢

A TCP connection specifically[1]

Great inuition! Yes!

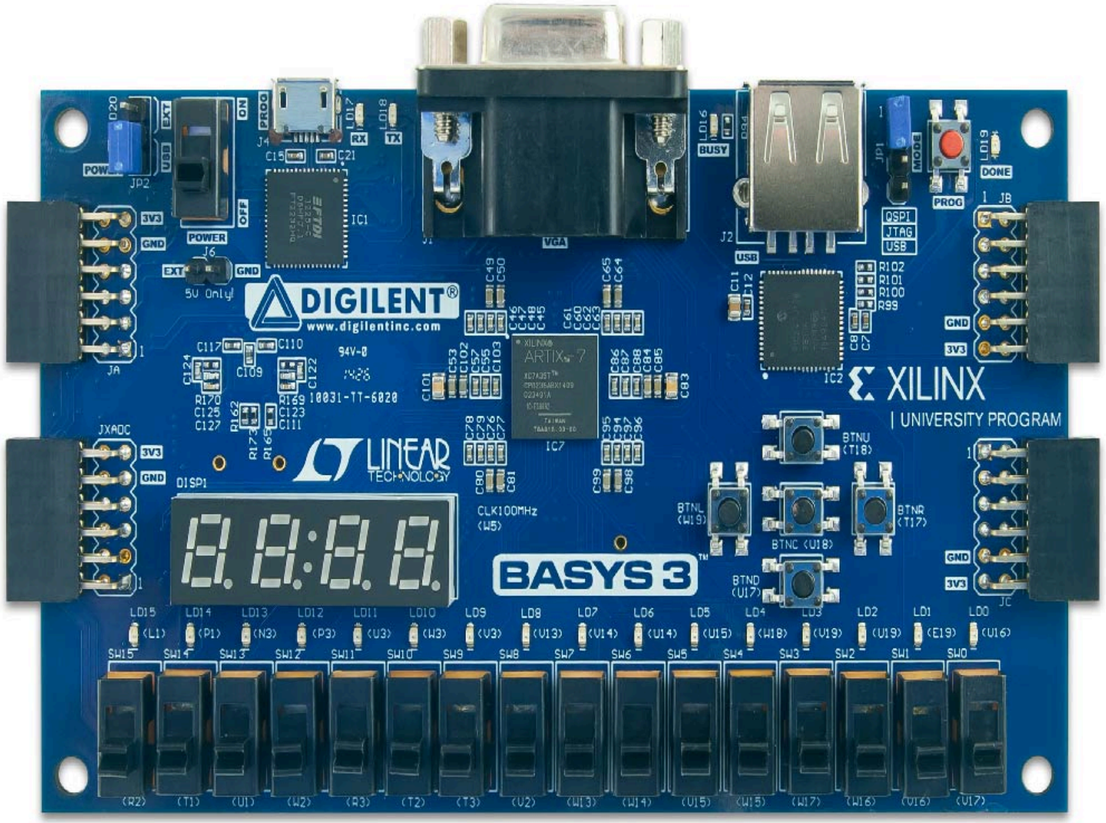
What if we did some of this in hardware? 😈

Introducing Generic Flow Tables! (GFT)



Fun Fact: GFT Lives on an FPGA!

Reconfigurability is important to improve perf and fix bugs as time goes on!



Yes, we use Vivado...

Remember this this from 140? That's part of my job now

Disclaimer: What I just showed you is an extremely simplified (and slightly incorrect) depecition of how it works... the fundamental principles remain the same though!

New Product Introduction (NPI)

We’ve finally arrived at what I do!

We are responsible for verifying the functionality and performance of the host networking components on new hardware to be added to the Azure Cloud

HOME > NEWS > THE CLOUD & HYBRID CHANNEL

Microsoft becomes first cloud to offer Nvidia Blackwell system

Cloud giant adds GB200 AI servers to its mix

October 09, 2024 By: Georgia Butler Have your say

Microsoft Azure has Nvidia's Blackwell system up and running on GB200-powered AI servers.

The company shared news of the deployment in a post to X, formerly known as Twitter.

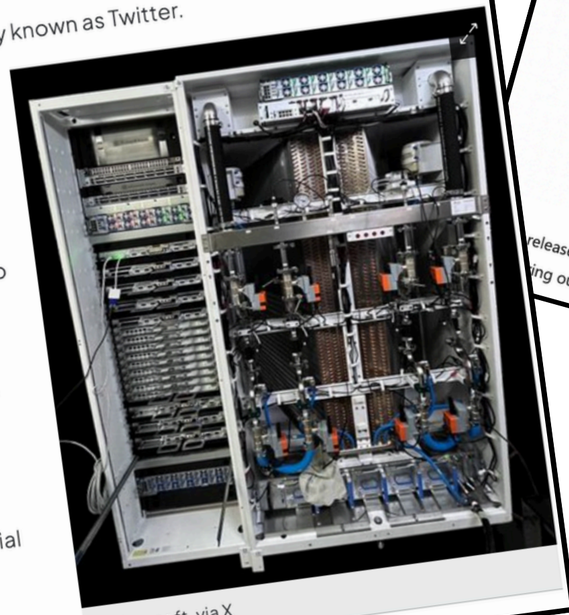
The post said: "Microsoft Azure is the 1st cloud running Nvidia's Blackwell system with GB200-powered AI servers. We're optimizing at every layer to power the world's most advanced AI models, leveraging Infiniband networking and innovative closed-loop liquid cooling. Learn more at MS Ignite."

Microsoft was previously rumored to be the first to gain access to the Blackwell servers.

According to a report from Tom's Hardware, the machine Microsoft has deployed is not NVL72 GB200, but is at least one GB200-based server rack with an unknown number of B200 processors.

The rack is likely to be used for testing purposes of both the Blackwell GPUs, and the liquid cooling system, and commercial deployment will follow in the coming months.

More details are expected to be shared at Microsoft's Chicago event, MS Ignite.



– Microsoft, via X

RUNNING SAP APPLICATIONS ON THE MICROSOFT PLATFORM 2 MIN READ

Azure delivers the first cloud VM with Intel Xeon 6 and CXL memory - now in Private Preview

Elyse_Ge_Hylander MICROSOFT
Nov 18, 2025

CXL on Azure

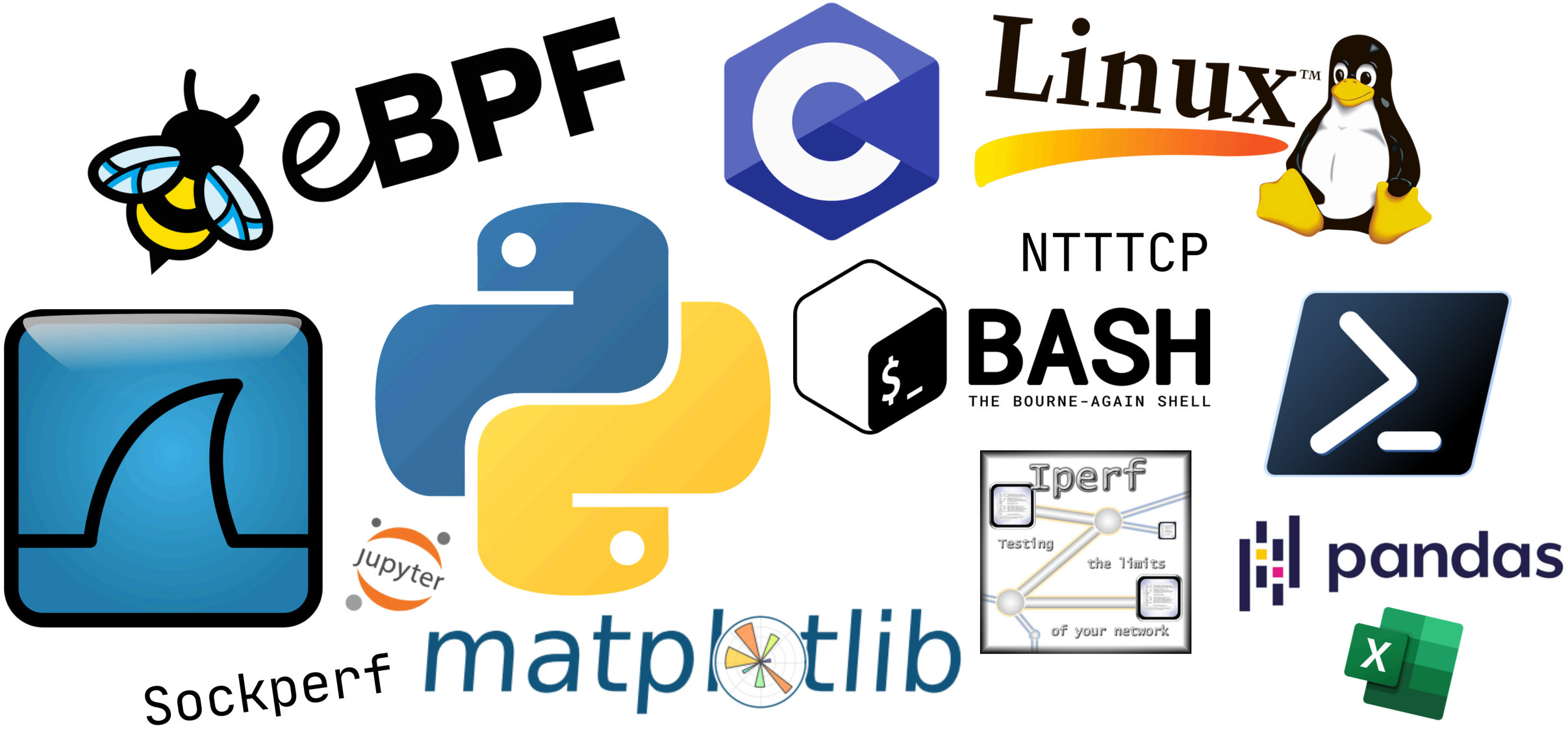
released their new Intel Xeon 6 6500/6700 series processor with P-cores this year. Intel Xeon 6 processors provide performance and scalability by ing outstanding performance for transactional and analytical workloads and provide scale-up capacities of up to 64TB of memory.

I led this one! My first big qual!

Why NPI is Hard 🥲

- Defining what is acceptable performance
- These systems are complicated! How do you test everything?
- Lots of data... 😵‍💫
- Speed
  - Longer tests take longer until customers can run workloads.
- Performance investigations can be tricky!

Tools I Use


eBPF
Linux™
NTTTCP
BASH
THE BOURNE-AGAIN SHELL
Iperf
Testing
the limits
of your network
jupyter
pandas
Sockperf
matplotlib

Thank you! ❤️

jameshurd2718 [at]
gmail [dot] com

Questions?

jameshurd.net/talks/sdan_at_msft

fin.