Voice over IP – A Beginners Guide

Dr. Jonathan Rosenberg
Vice President and Chief Technology Officer of Collaboration at Cisco Systems

About Me

Bachelors and Masters 1994
Electrical Engineering

PhD, Voice over IP 2001

Member of Technical Staff, Bell Laboratories, Holmdel

Birthplace of the Transistor

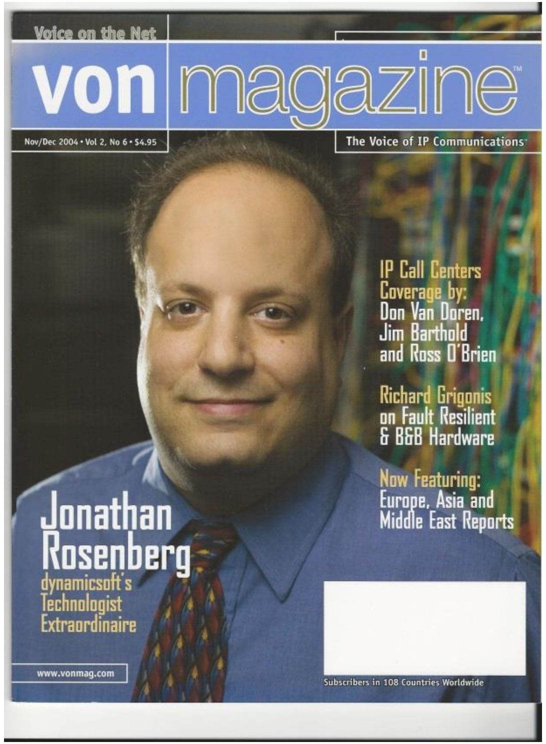
Water Tower at Holmdel Site

Coincidence or on Purpose?

You decide!

Transistor

Chief Technology Officer, dynamicsoft



dynamicsoft.™

Acquired 2004

CISCO™

Left 2009

Chief Technologist, Skype

Left 2012

CTO, Collaboration at Cisco

Co-Inventor and Lead Author of the Session Initiation Protocol (SIP)

SIP is the fabric that connects all modern telecom products and services

Billions and billions of minutes of calls

The Phone Network is ~50% converted to SIP and will be completely built on SIP by 2020ish

My Career Mission

To Reimagine Telecommunications in the Image of the Internet

Awards and Accomplishments

~125 filed,
89 Granted US Patents

Technology Review
TR35 *The World's Top 35 Innovators Under 35*

Named to TR35 in 2002, along with Sergey Brin, Larry Page, John Carmack

I E T F

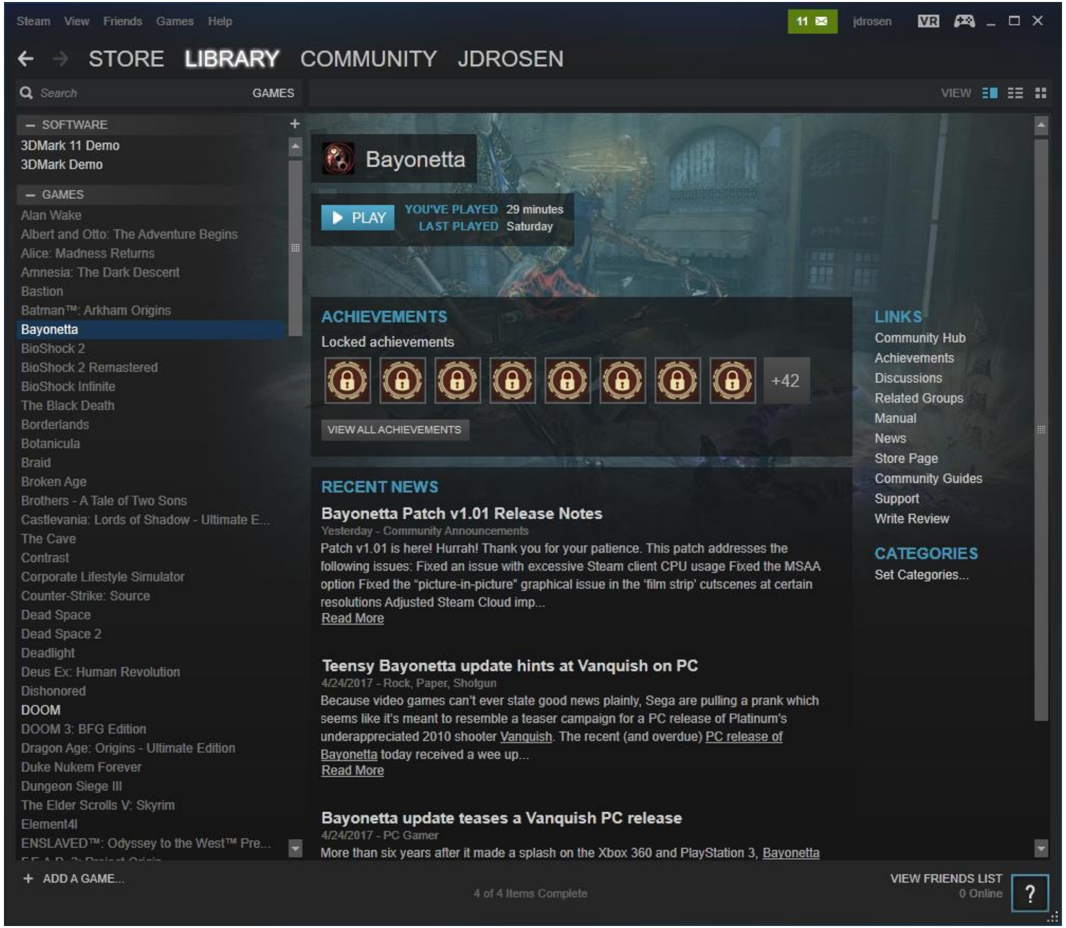
71 Internet Standards, 8th most prolific creator of Internet standards of all time

2015 Ohio University Strowger award for pioneering contributions to telecommunications

cisco

jdrosen on Steam
135 titles!

1. Half Life 2
2. World of Warcraft
3. Portal 2
4. Ori and the Blind Forest
5. Alice: Madness Returns
6. The Last of Us
7. Uncharted 3
8. Rise of the Tomb Raider
9. Tomb Raider (2013)
10. Portal 2
11. Prince of Persia: Sands of Time
12. Doom (2016)
13. Prince of Persia: The Forgotten Sands
14. Limbo
15. Portal
16. Legend of Zelda: Link Between Worlds
17. Guitar Hero III
18. God of War III
19. Bioshock
20. The Talos Principle



Where I work Now: Cisco

Not This Sysco

Where I work Now: Cisco

This Cisco

Cisco Fast Facts

CEO: Chuck Robbins

Headquarters: San Jose, CA

Yearly Revenue: $49.1 Billion 2016

Market Capitalization: $158B

Employees: ~75,000

So… what does it make?

Cisco Makes all of the Stuff Under the Hood that Makes the Internet Work

WiFi Access Points for Businesses

All the Stuff that AT&T use to Provide you Internet access

A Data Center

The Stuff that Walmart uses to connect its stores and customers

What My Business Unit Does: Collaboration

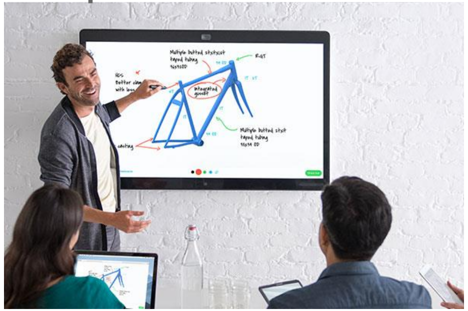
Phones and Phone Control Software
(Freehold Township School District is a customer!)

Web Conference Calling

Cisco Spark

Telepresence Systems

Spark Board



$4B Yearly Revenue

A Quick Pitch…

Cisco Spark

GET IT ON Google Play

Available on the App Store

Cisco Spark – FREE!

- 1-1 and persistent group chat up to 5000 people
- Voice and video calling
- Schedule meetings, ad-hoc meetings
- File sharing and in-app preview
- Photo sharing

What is Voice over IP?

Making Voice and Video Calls using

Your Internet Connection

Instead of your Phone Line

IP == The Internet Protocol – the Core Technology of the Internet

The Internet vs. the Phone Network

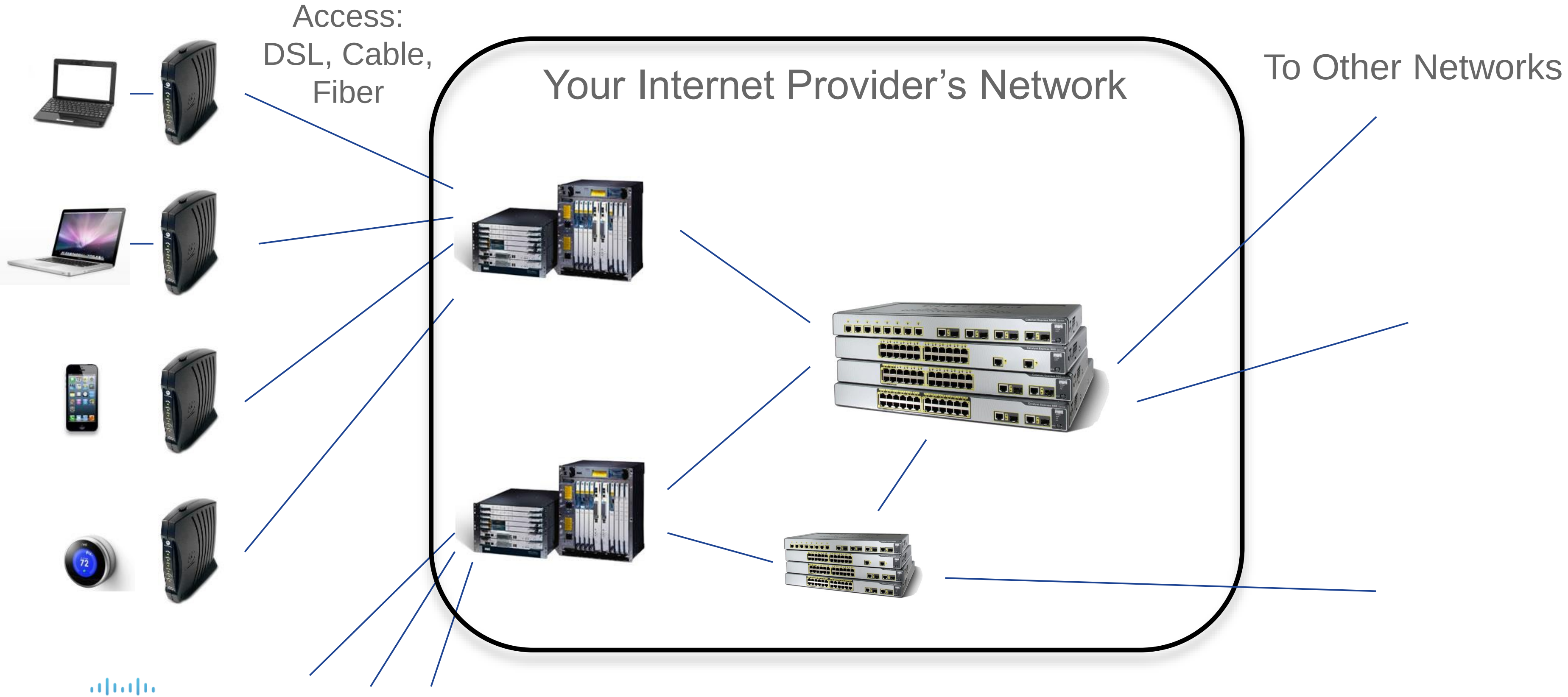
Phone Network: Circuit Switched
Like a wire between users

Internet: Packet Switched
Routes messages between users

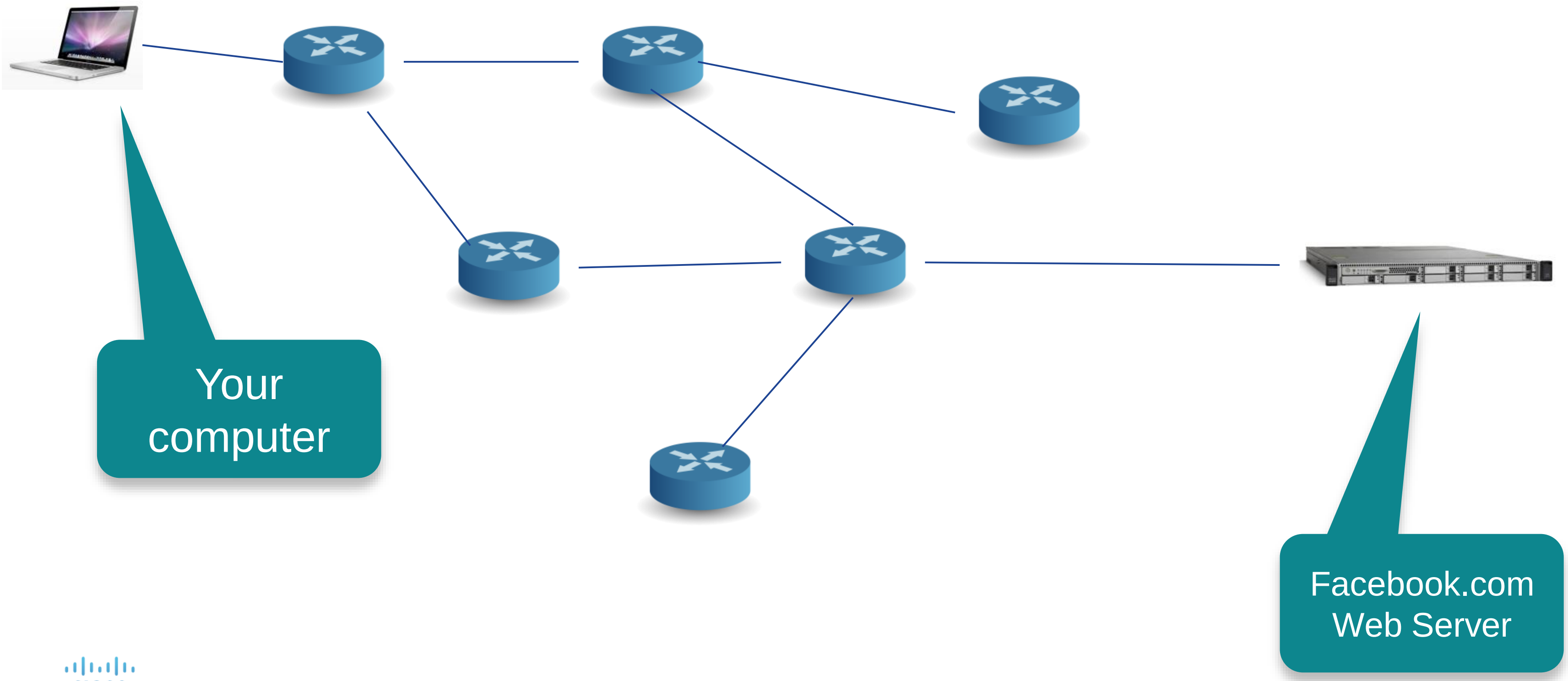
Why Use Internet for Calling Instead of Phone Network??

- Cheaper calls
- As a business, one network instead of two
- Better experiences (video!)
- Mobility – receive my calls anywhere I have Internet

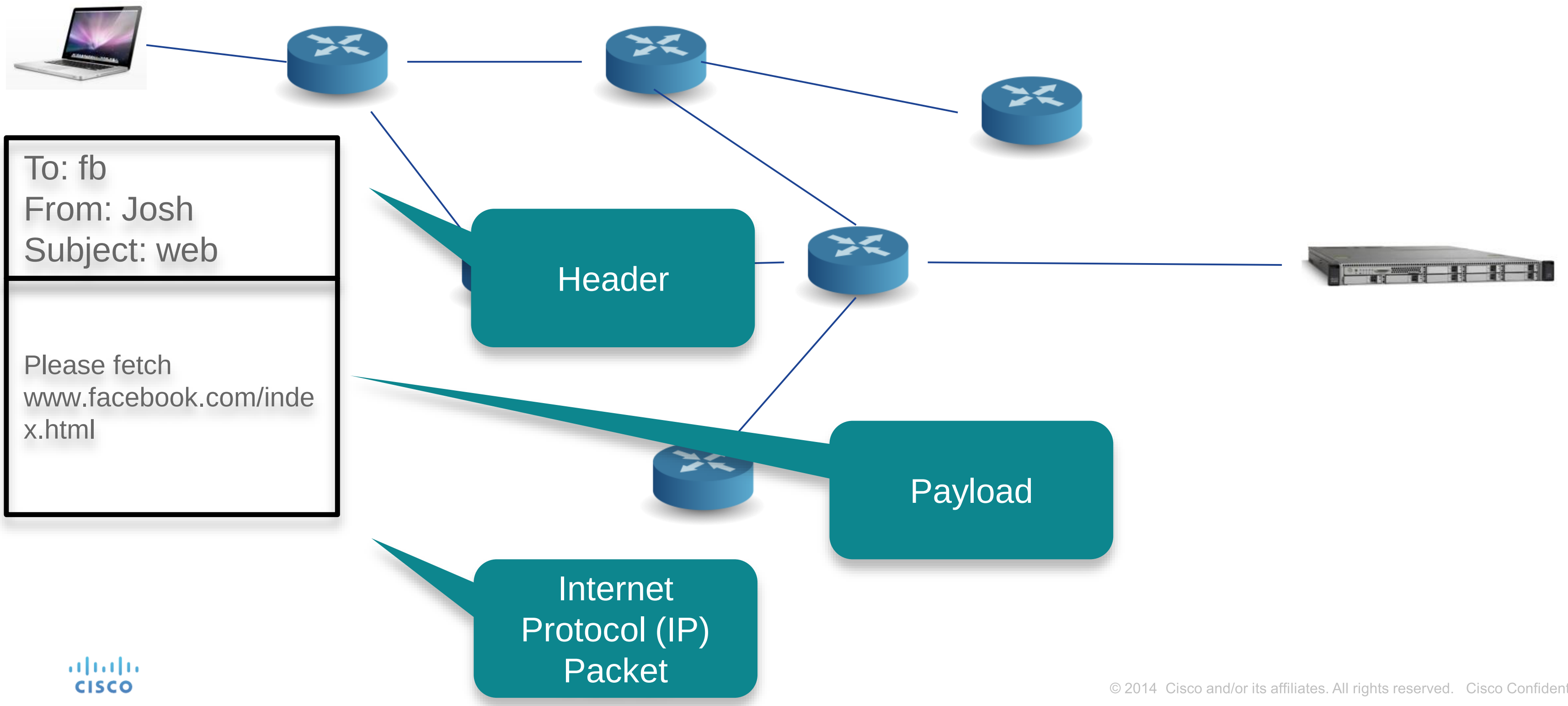
How does it work?



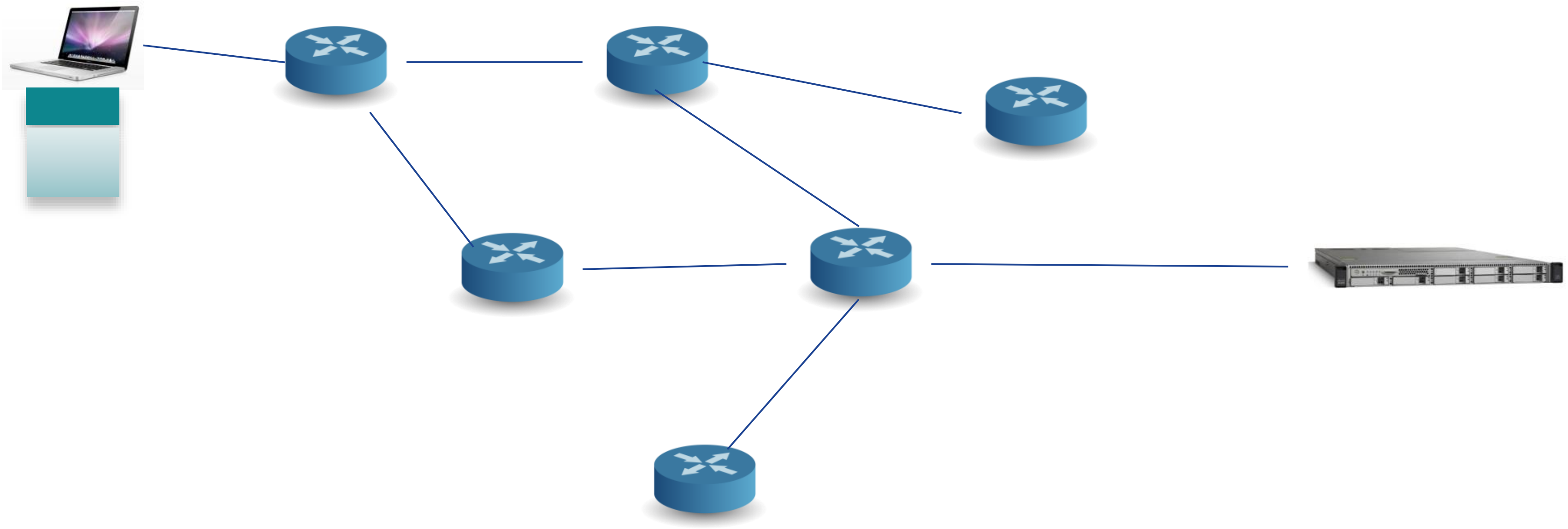
Packet Switching



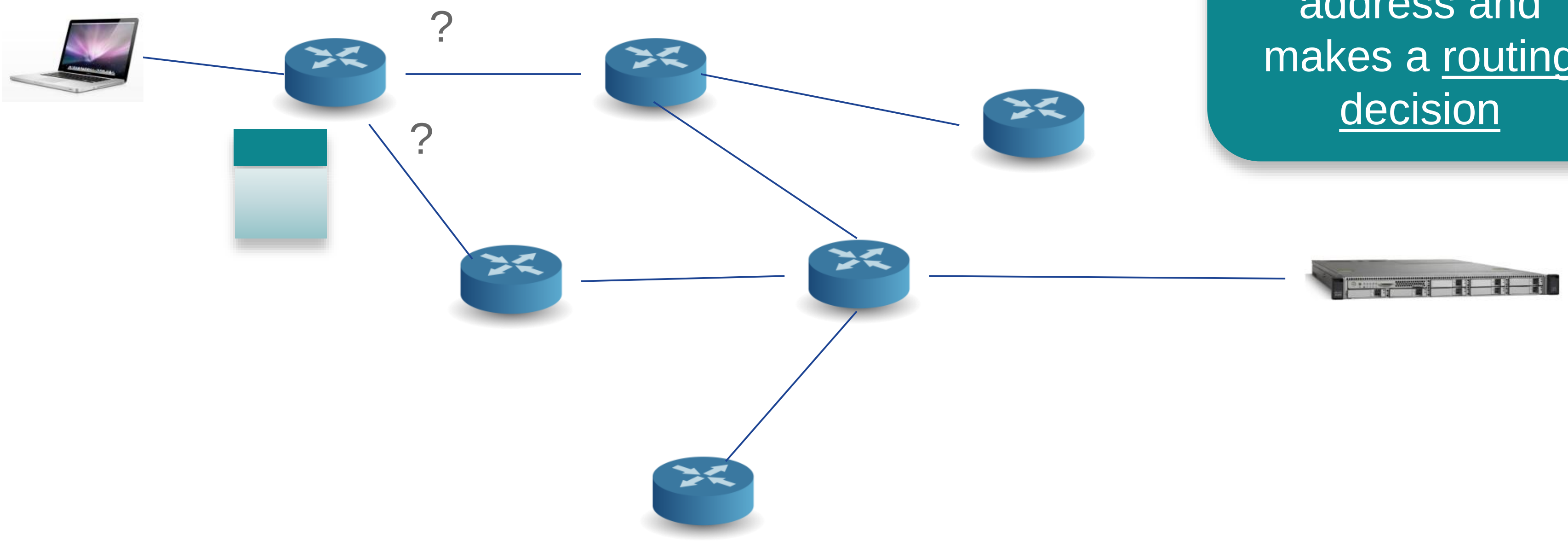
Packet Switching



Packet Switching



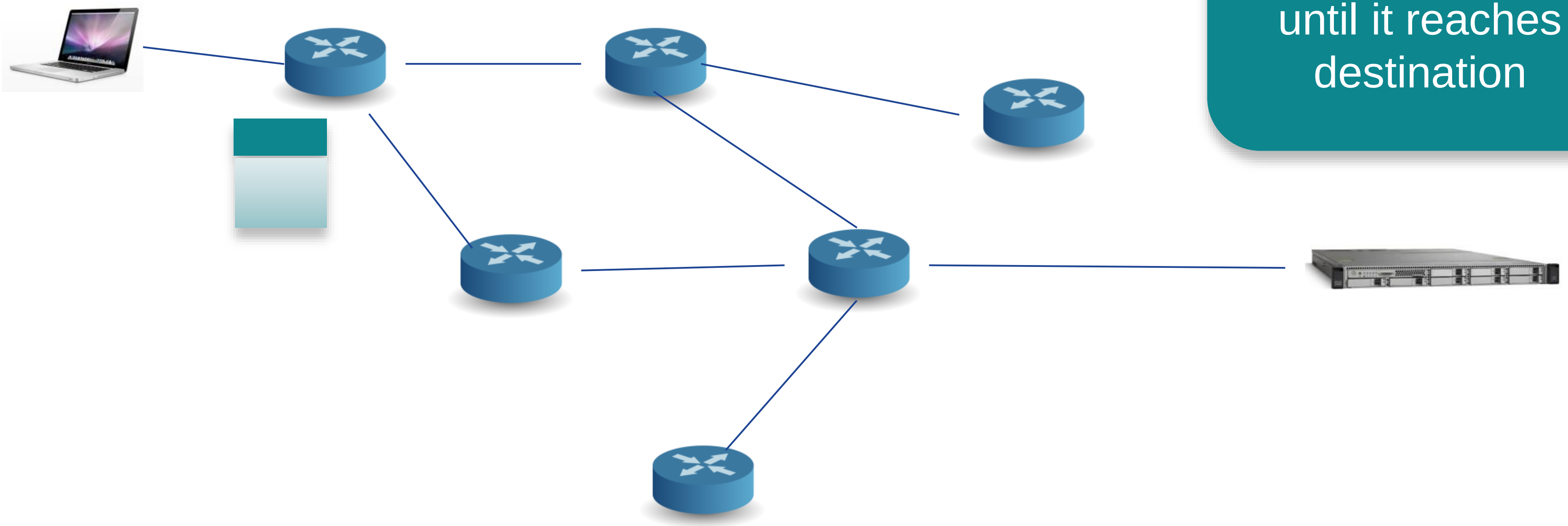
Packet Switching



Router looks at the destination address and makes a <u>routing decision</u>

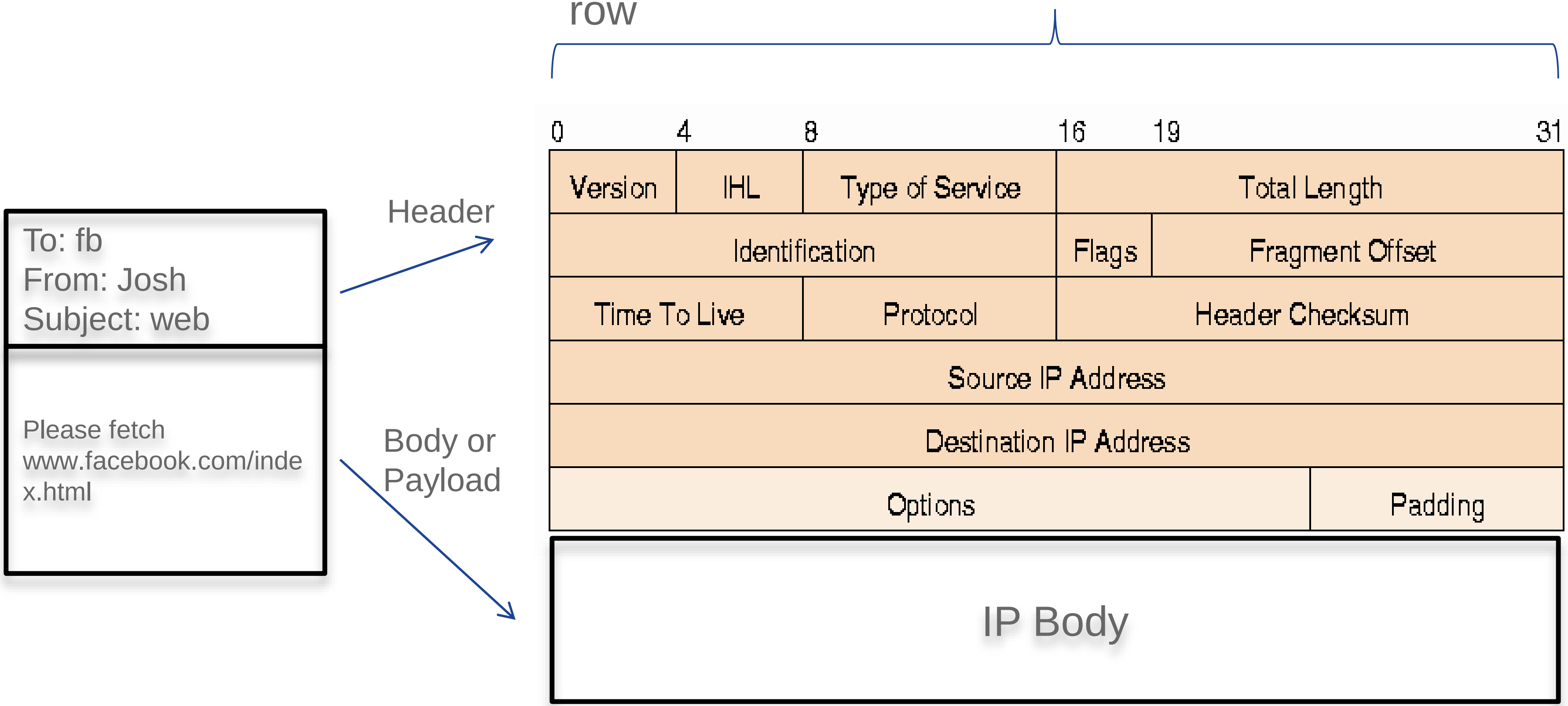
Packet Switching

Packet moves from hop to hop until it reaches destination



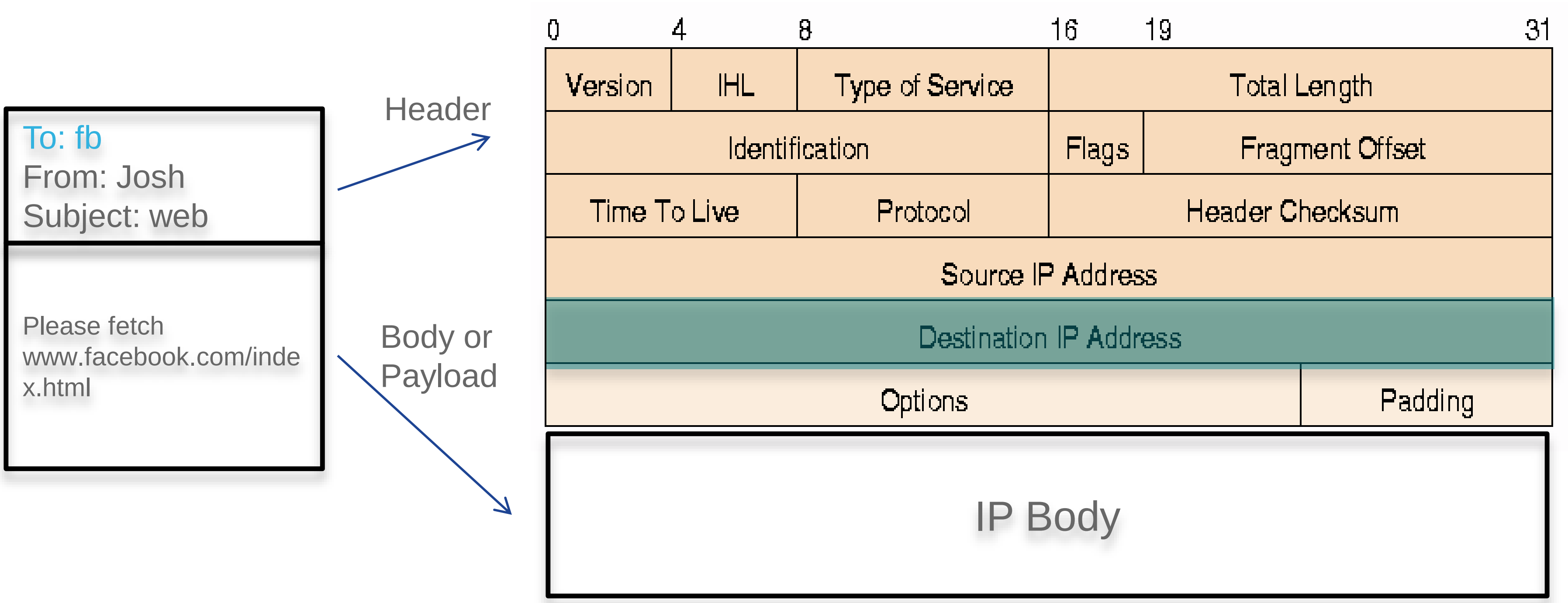
A Real Packet

The packet is sent bit at a time. We write it this way – showing each consecutive 32 bits as a row



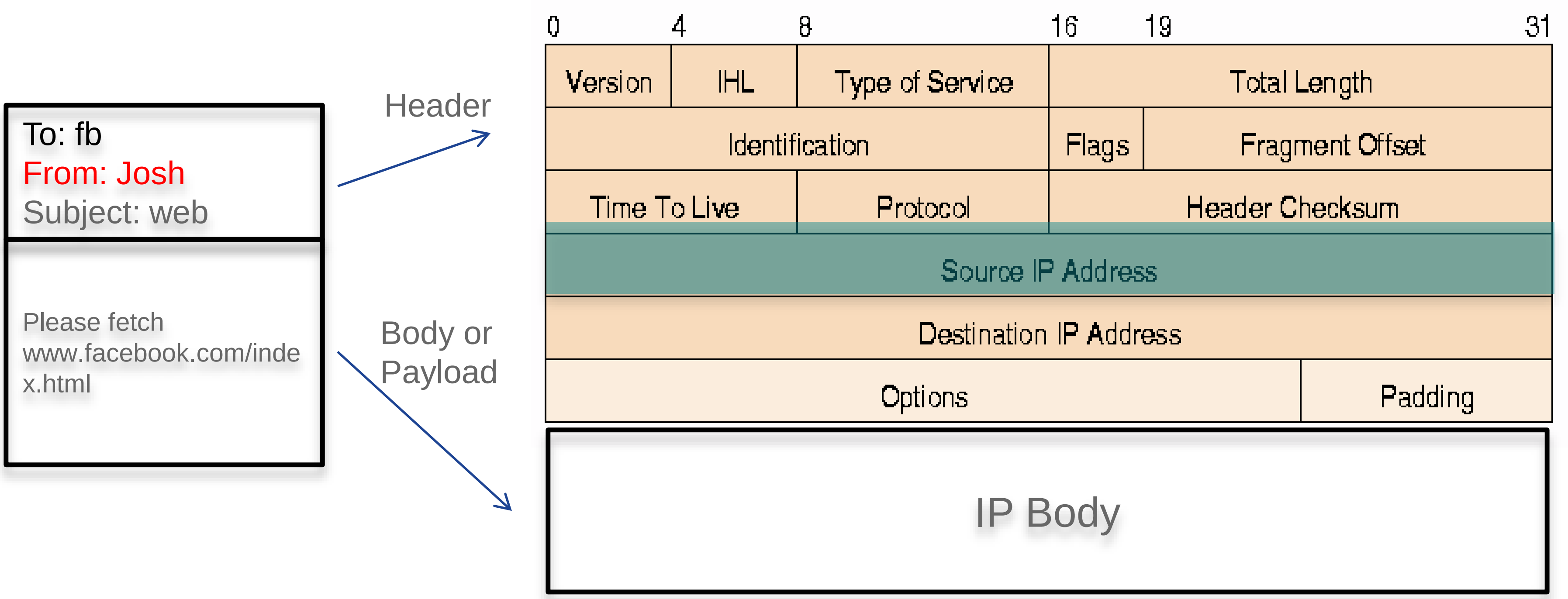
A Real Packet

This set of 32 bits is the destination address – it's the "To" part of the packet.



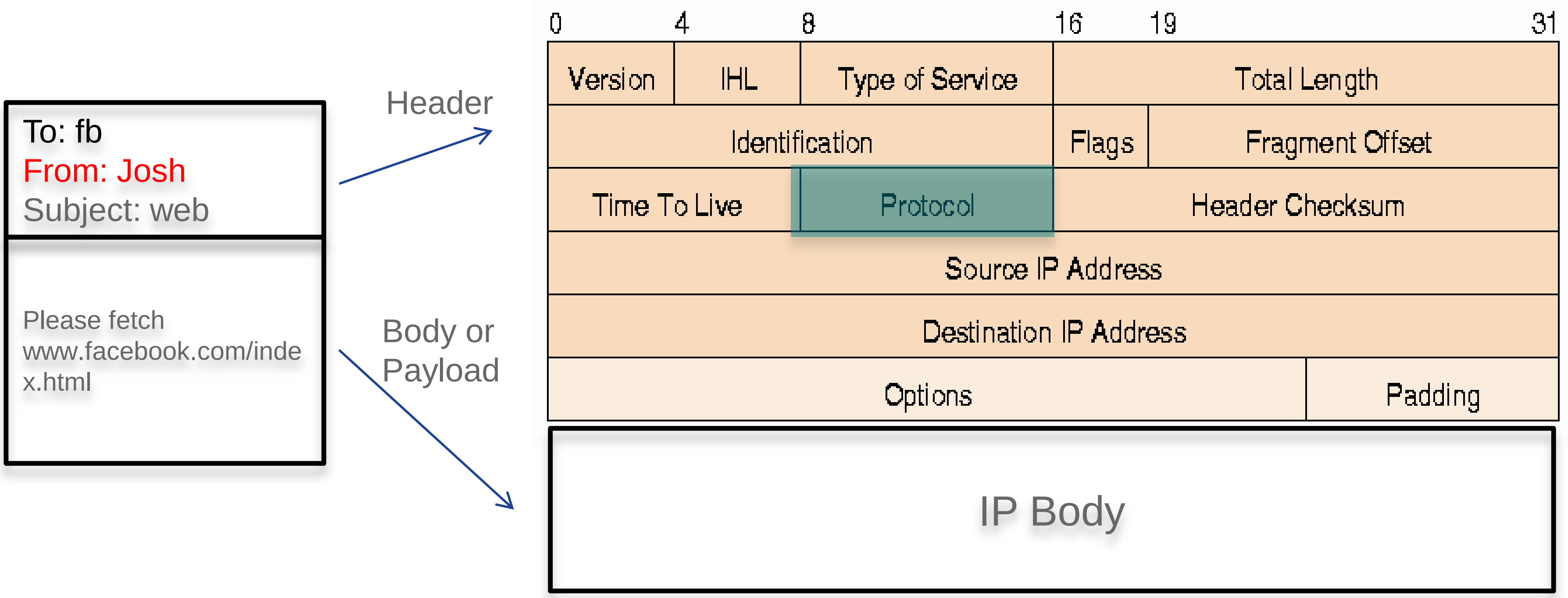
A Real Packet

This set of 32 bits is the source address – it's the "From" part of the packet.



A Real Packet

This set of 16 bits is the protocol – it describes what comes next in the packet. It's the "Subject" part of the packet – part of it at least.



Sidebar - what is a Protocol?

A communications protocol is a system of digital rules for data exchange between computers through a computer network

Syntax

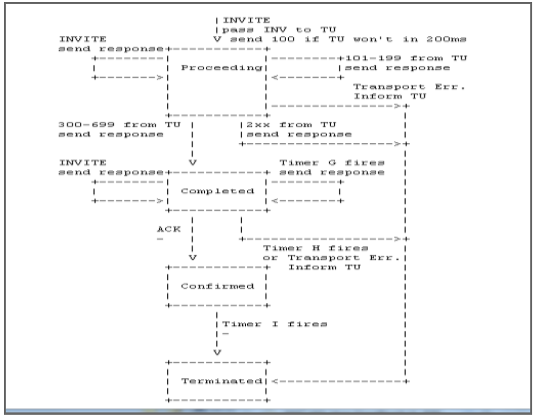
rules to encode and decode data

```
Rosenberg, et. al.          Standards Track

RFC 3261             SIP: Session Initiation Protocol


cnonce              =  "cnonce" EQUAL cnonce-value
cnonce-value        =  nonce-value
nonce-count         =  "nc" EQUAL nc-value
nc-value            =  8LHEX
dresponse           =  "response" EQUAL request-digest
request-digest      =  LDQUOT 32LHEX RDQUOT
auth-param          =  auth-param-name EQUAL
                       ( token / quoted-string )
auth-param-name     =  token
other-response      =  auth-scheme LWS auth-param
                       *(COMMA auth-param)
auth-scheme         =  token
```

Type of language called ABNF which specifies protocol syntax, here for the SIP protocol

Semantics

rules to process data



A formal model of protocol semantics using a Finite State Machine (FSM) – here for SIP transactions.

Why do we need Protocols?

Protocols describe the behavior of software which allows computers to talk to each other successfully to accomplish a task

Protocols need to be rigorous since computers are literal – protocols specify rules that are meant for a computer to process

Protocols need to deal with problems that might happen when computers talk to each other – loss, delay, jitter, different versions of software

Protocols are defined by a combination of computer programs (often in special languages) combined with English prose

Protocols are often standardized – agreed upon by a formalized community of computer scientists – enabling different companies to make products that work together

You Use Protocols Everyday

Protocol Name	Used For	Standardized In
IP – the Internet Protocol	Exchanging packets between routers	http://www.ietf.org/rfc/rfc791.txt
HTTP – the Hypertext Transfer Protocol	Requesting and transferring documents from a web server to browser	http://www.ietf.org/rfc/rfc2616.txt
DNS – the Domain Name System	Allowing a computer to convert a human readable name (like www.twitter.com) to an IP address	https://www.ietf.org/rfc/rfc1034.txt
SMTP – the Simple Mail Transfer Protocol	Sending mail between two computers	http://www.ietf.org/rfc/rfc2821.txt
TCP – the Tranport Control Protocol	Reliable transmission of a stream of bytes	

Where do the protocols on the Internet come from?

The Internet Engineering Task Force (IETF)

An informal collection of engineers and computer scientists from around the world that define the rules that govern the Internet and oversee its operation

In existence since 1986

The standards it produces are called RFCs

8174 RFCs produced so far

An Example Page of SIP

After a new request has been created, and the header fields described above have been properly constructed, any additional optional header fields are added, as are any header fields specific to the method.

SIP requests MAY contain a MIME-encoded message-body. Regardless of the type of body that a request contains, certain header fields must be formulated to characterize the contents of the body. For further information on these header fields, see Sections 20.11 through 20.15.

8.1.2 Sending the Request

The destination for the request is then computed. Unless there is local policy specifying otherwise, the destination MUST be determined by applying the DNS procedures described in [4] as follows. If the first element in the route set indicated a strict router (resulting in forming the request as described in Section 12.2.1.1), the procedures MUST be applied to the Request-URI of the request. Otherwise, the procedures are applied to the first Route header field value in the request (if one exists), or to the request's Request-URI if there is no Route header field present. These procedures yield an ordered set of address, port, and transports to attempt. Independent of which URI is used as input to the procedures of [4], if the Request-URI specifies a SIPS resource, the UAC MUST follow the procedures of [4] as if the input URI were a SIPS URI.

And now back to the IP packet…..

The IP Address

- Every device on the Internet has an IP address
- The IP address is 32 bits long and uniquely identifies the device on the Internet
- If a device has multiple connections to the Internet (think: your iPhone has WiFi and LTE), it has one IP address per connection – this is called an interface
- The routers and switches themselves also have an IP address – one on each interface
- With 32 bits there are 2^{32} IP addresses available – or 4.2 Billion addresses – and we're almost out of them!!

IP Address Notation

31.13.69.228

This is the actual IP address of facebook.com in the US

We write it as four numbers separated by periods

Four numbers, each of Which has 256 values, gives us all 2^{32} addresses

Each number is between 0 and 255. That's 256 numbers, or 2^8

IP Address in Iron Man 3!

Hmm – what's wrong with this?

IP Address Blocks

Notation	Means	Number of addresses in this block	Known as
173/8	The 2^{24} addresses from 173.0.0.0 to 173.255.255.255	2^{24} = 16.7 million	Class A network
173.252/16	The 2^{16} addresses from 173.252.0.0 to 173.252.255.255	65,636	Class B network
173.252.110/24	The 2^{8} addresses from 173.252.110.0 to 173.252.110.255	256	Class C network

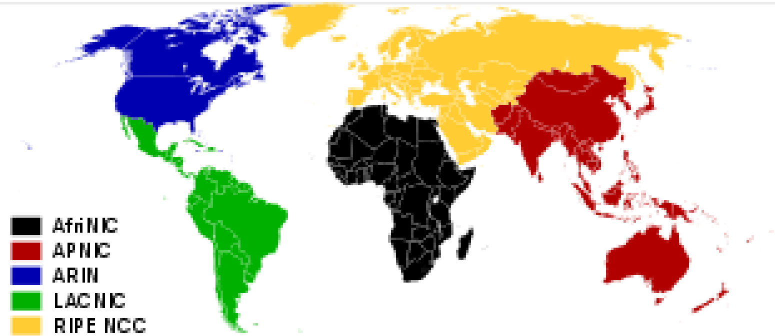
IP Address Assignment

Internet Assigned Numbers Authority

IANA is the international body that oversees all IP address allocation. They allocate large blocks to:

The RIRs, which manage geo-specific allocation to:

Regional Internet Registries (RIRs)



comcast

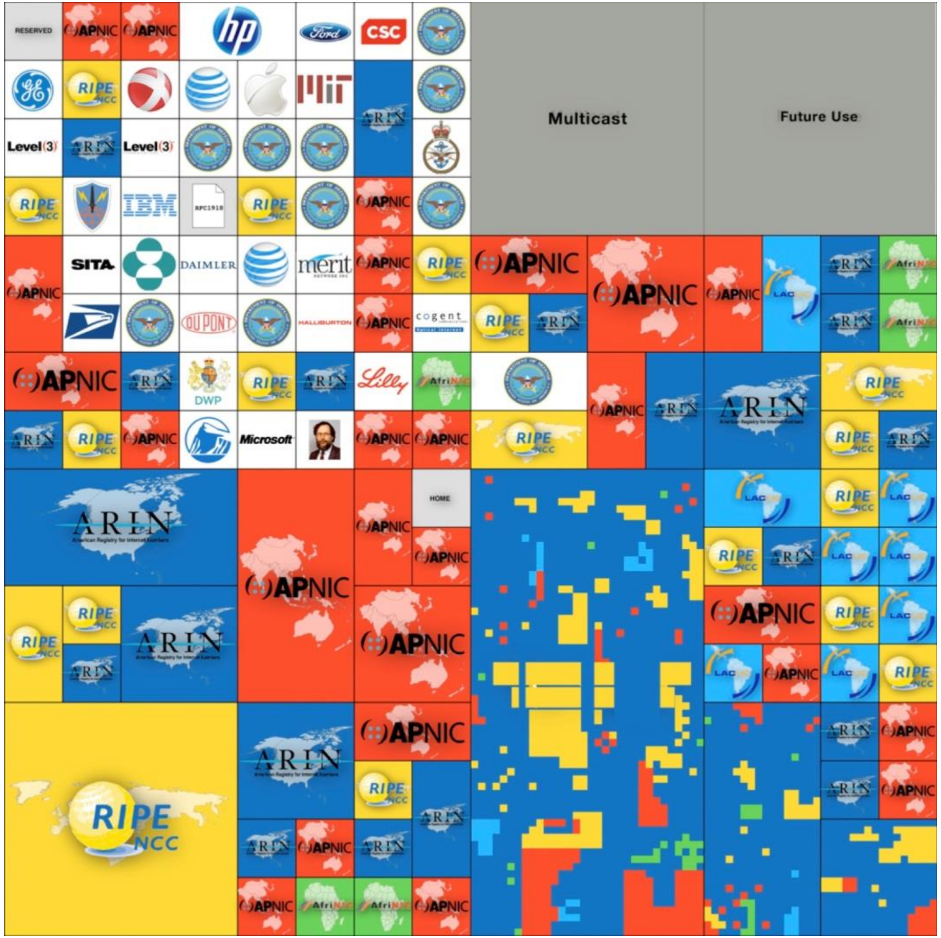
verizon

ISPs, which then allocates one to

Your internet router

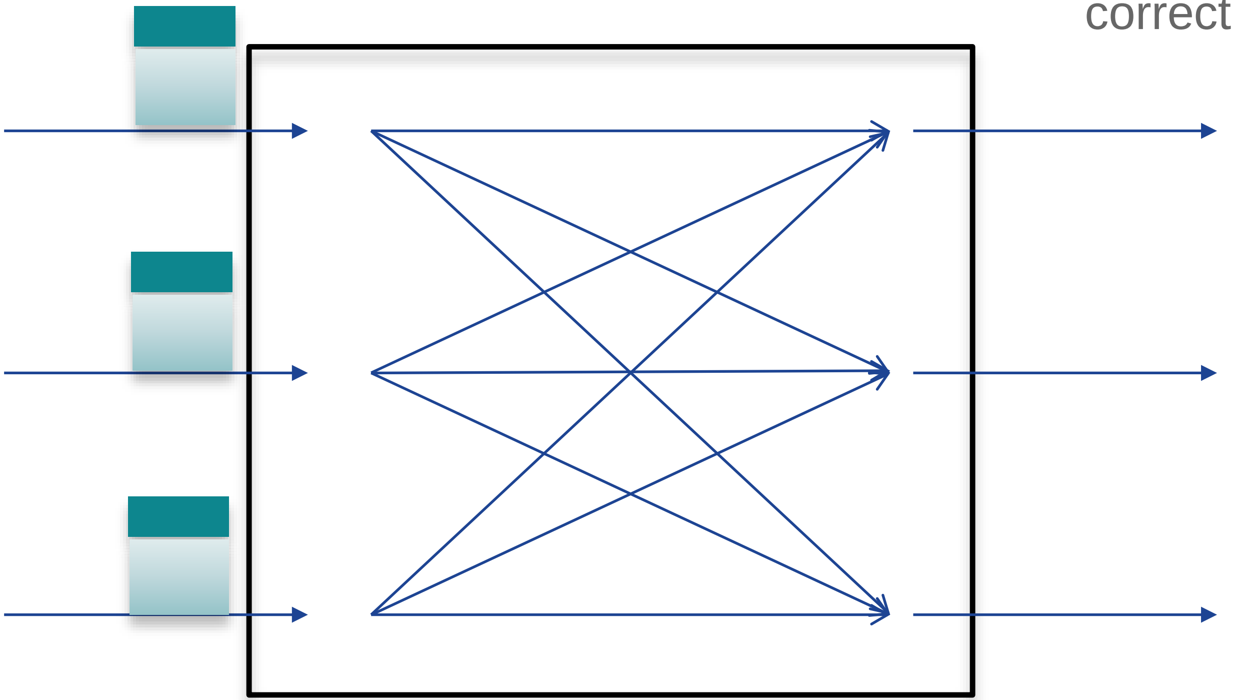
IP Address Map

Each little block represents a /8 – They are now all allocated.



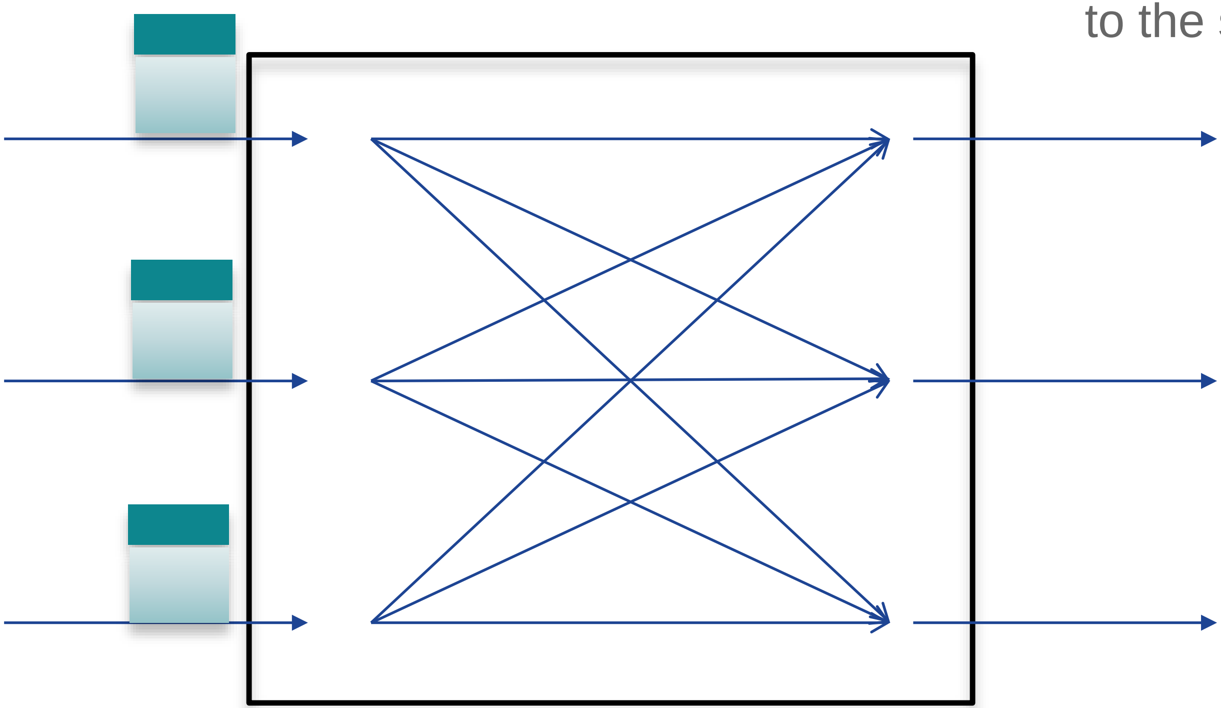
Back to Routers for a Moment

Packets arrive at the inputs and are moved to the correct output



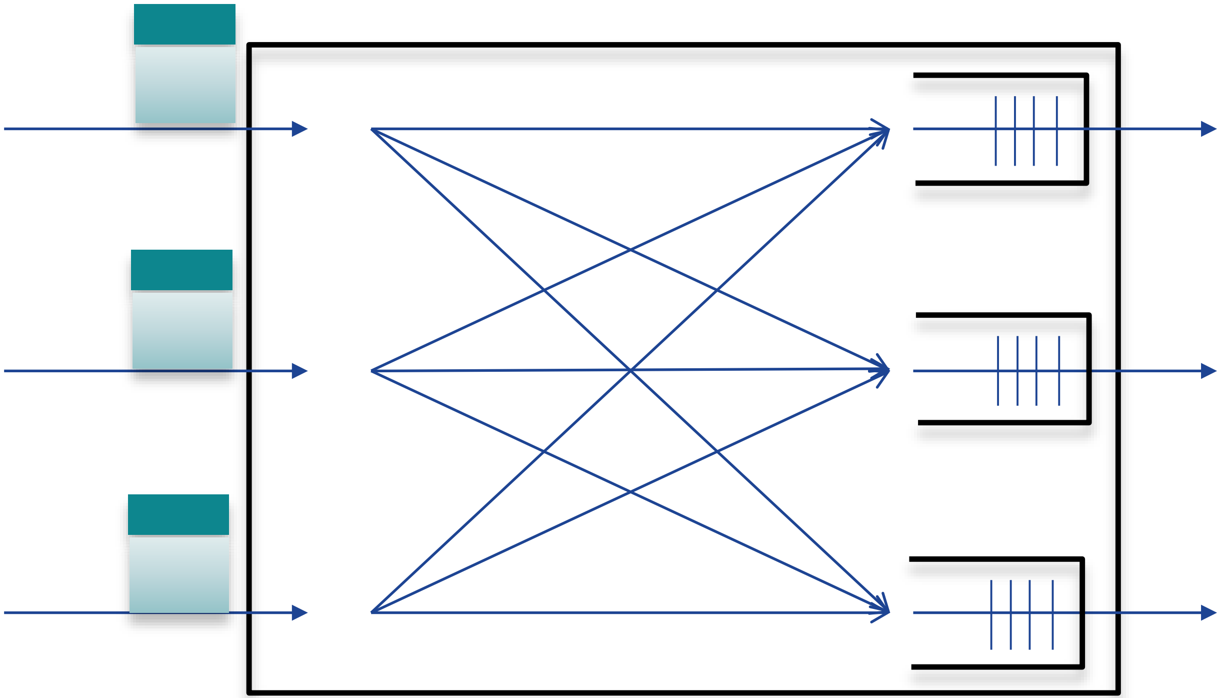
Back to Routers for a Moment

But what if more than one of the inputs needs to go to the same output?



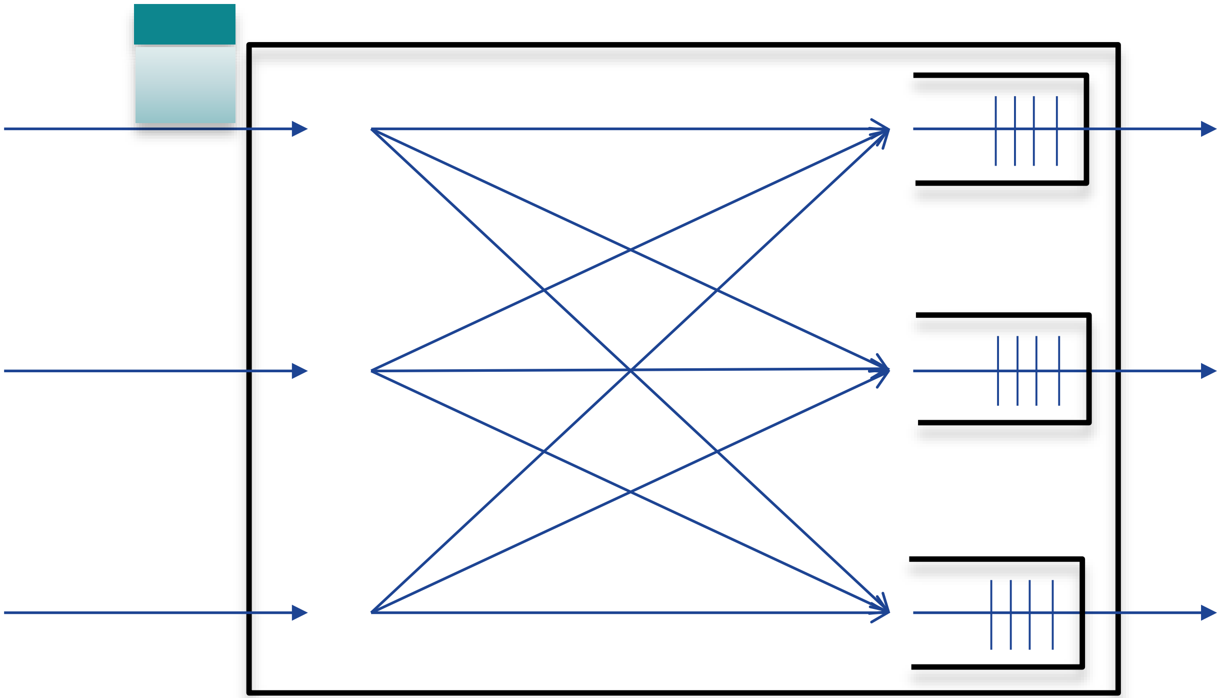
Queues

The router has buffers, also called queues – which holds packets until they can be sent



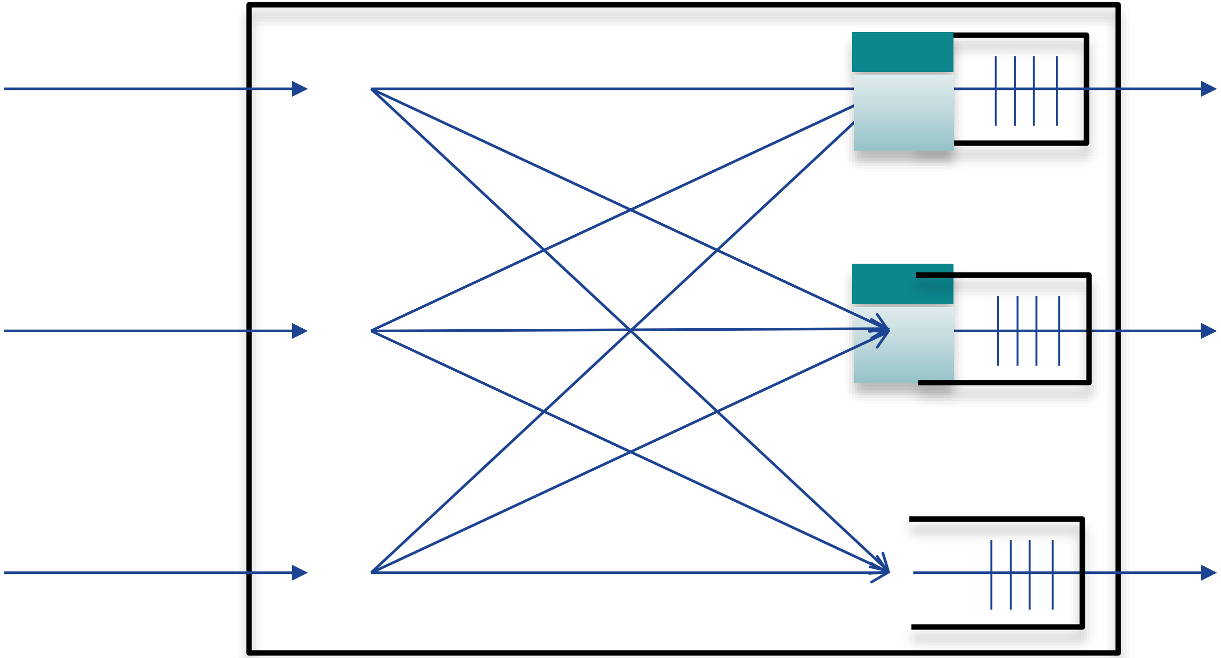
Uh Oh - Congestion

If a the queue is full when the next packet shows up – its dropped! This is called congestion, and it causes packet loss.



Uh Oh - Delay

Congestion also causes delay. If there is a lot of packets in the queue, the new packet has to wait a bit to get sent. If there arent any packets, it can get sent fast.



More Delay – that Pesky Speed of Light!

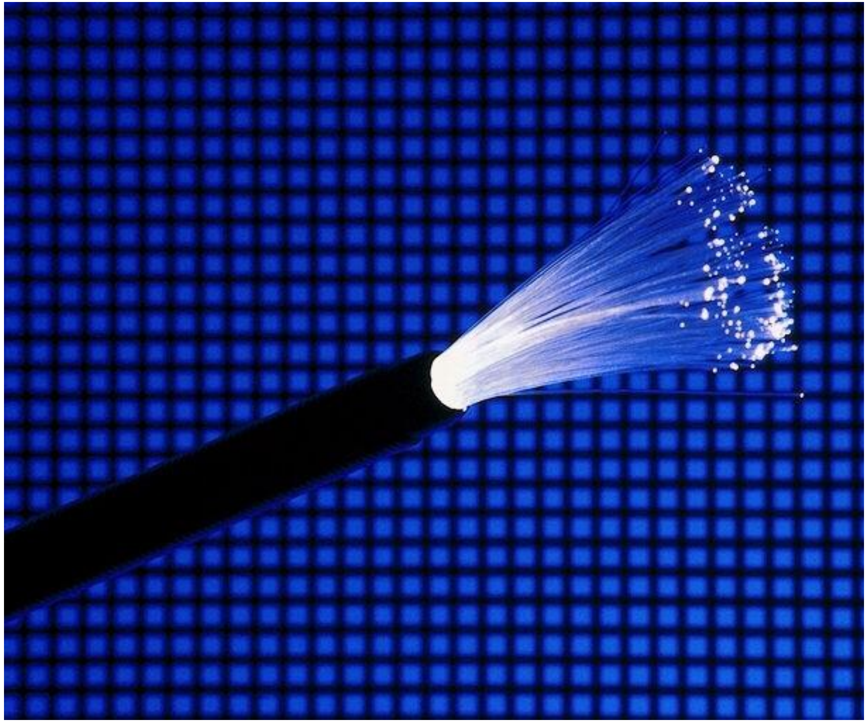
Speed of light in vacuum = 300,000 km/s (186,000 miles/sec)

Speed of light in an optical fiber = 200,000 km/s

Distance from Sydney to New York = 16,000 km

Time for speed of light to travel from Sydney – New York = 80ms

80ms is a lot – more soon!



More Uh Ohs - Jitter

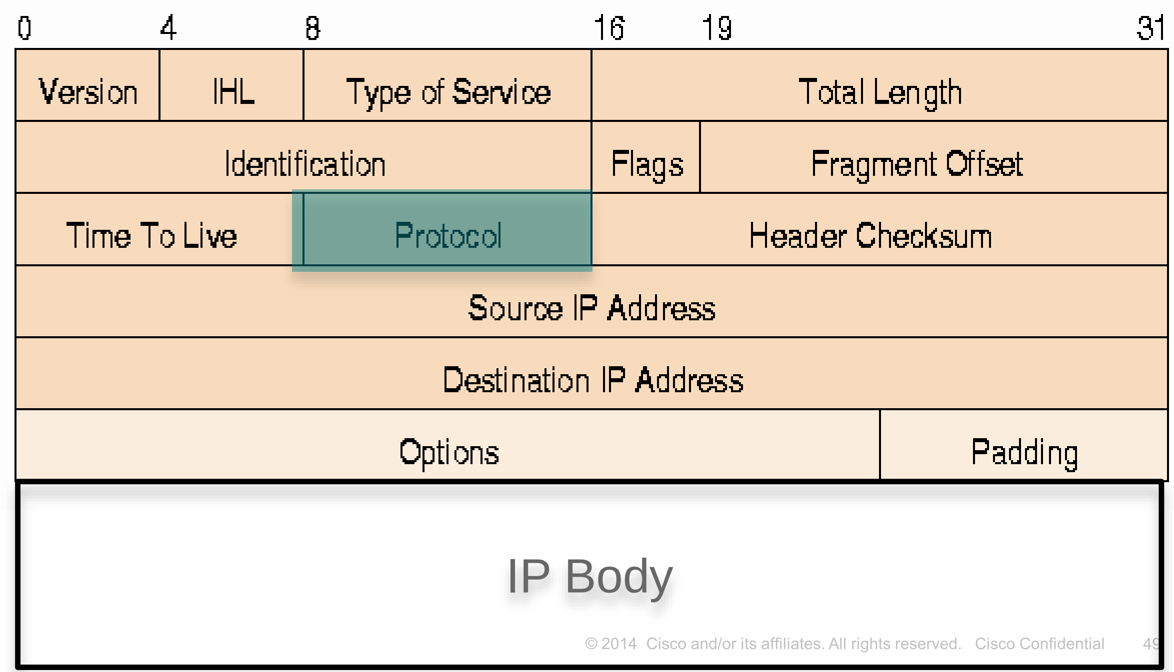
- Jitter is defined as variation in delay

Packet Number	Delay
1	33ms
2	31ms
3	40ms
4	39ms

Jitter is 9 ms
(40ms – 31ms)

Back to The IP Packet

This set of 16 bits is the protocol – it describes what comes next in the packet. It's the "Subject" part of the packet – part of it at least.



The Two Big Protocols – TCP and UDP – the Middleware of the Internet

The Transmission Control Protocol (TCP)

Stream Oriented – sequence of bytes

Reliable – Retransmit lost data

Rate control – limits how fast you can send to not clog the Internet

Connection oriented – computers exchange messages to “set it up” first and “end it” when done

The Unreliable Datagram Protocol (UDP)

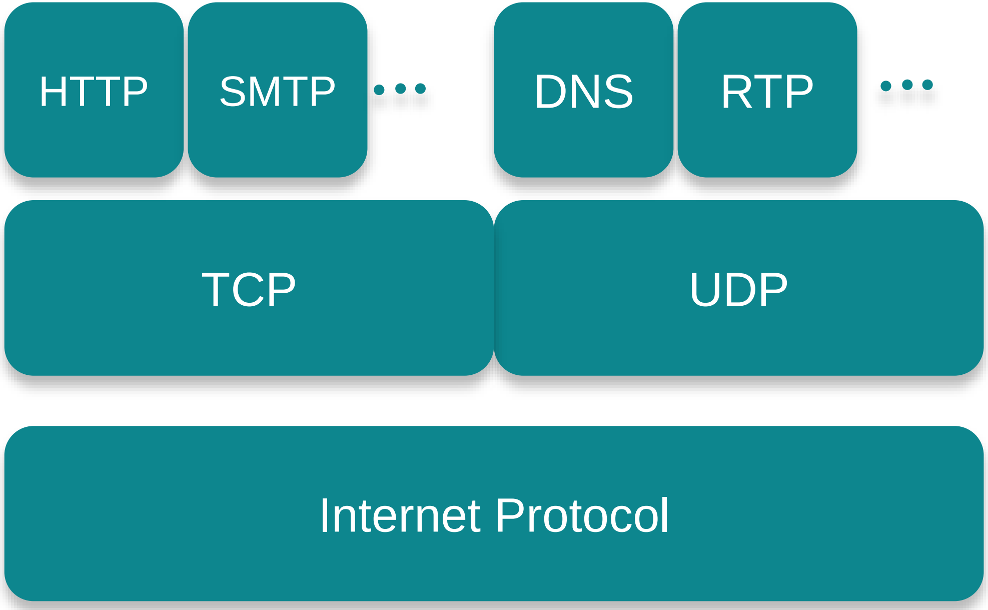
Packet Oriented – sequence of packets

Unreliable – no retransmissions

No Rate control – care needed not to clog the Internet

Connectionless – just send it and go

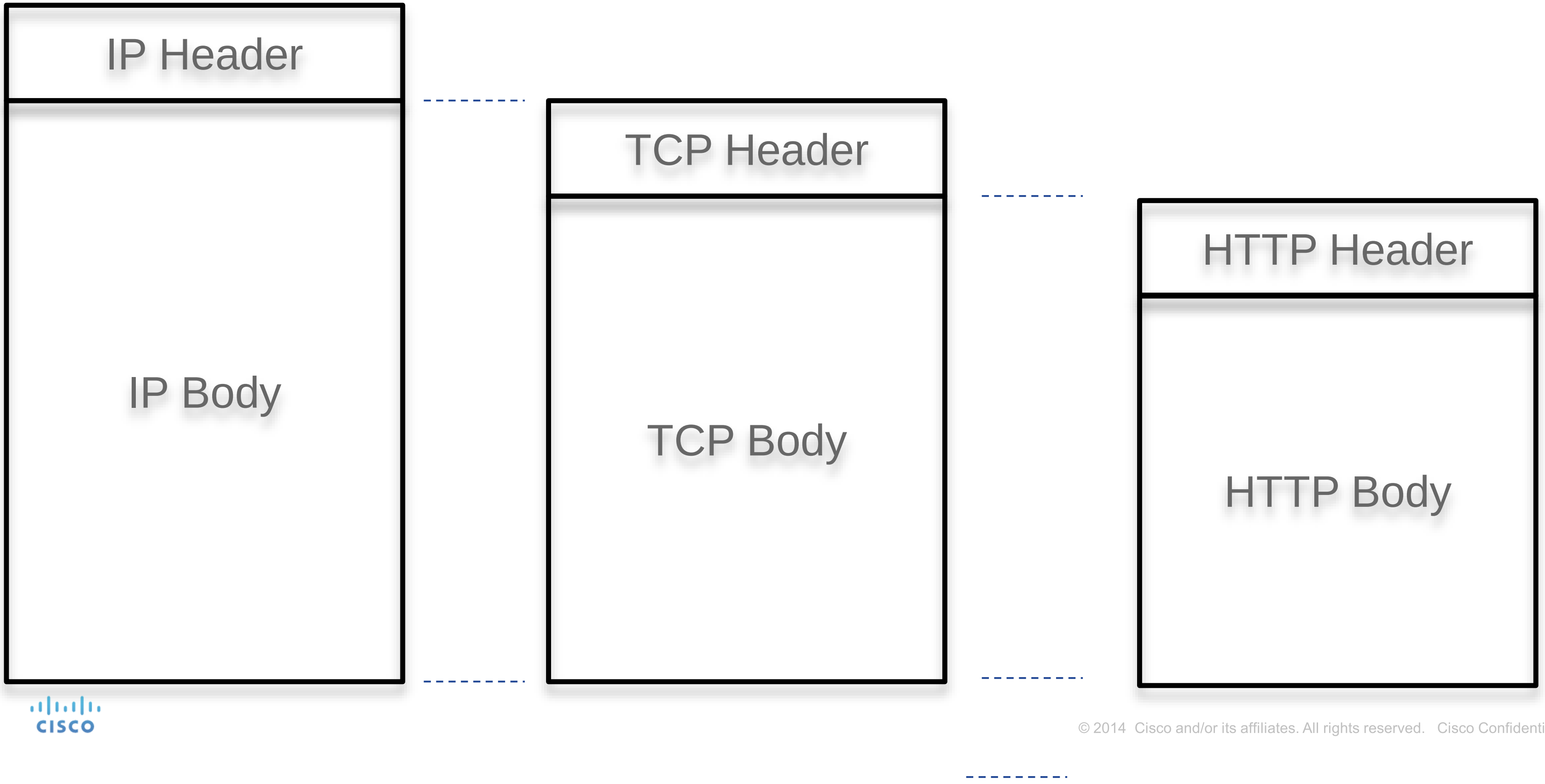
Protocol Layering



Each protocol runs "on top of" the ones below

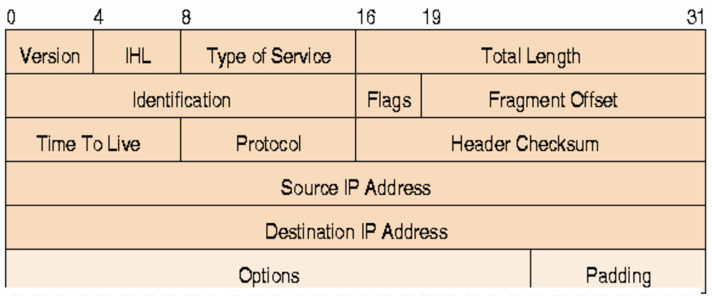
Analogy for Protocol Layering

Protocol Layering

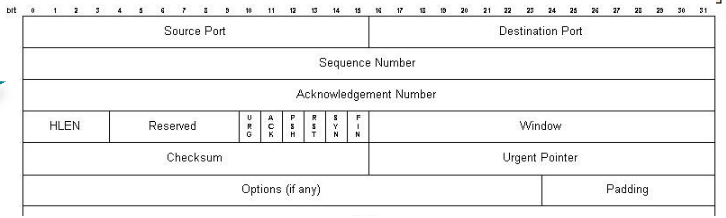


On-the-Wire

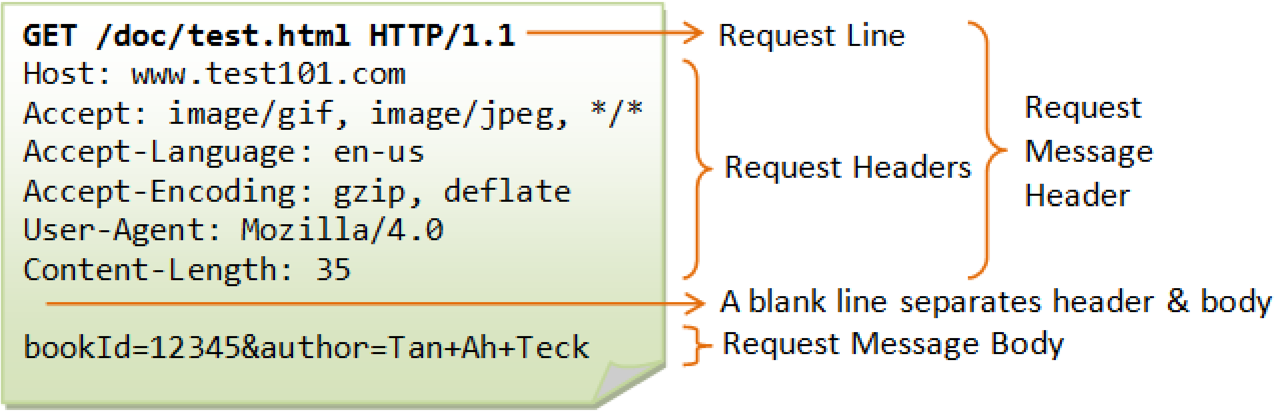
First we send these bits – the IP header



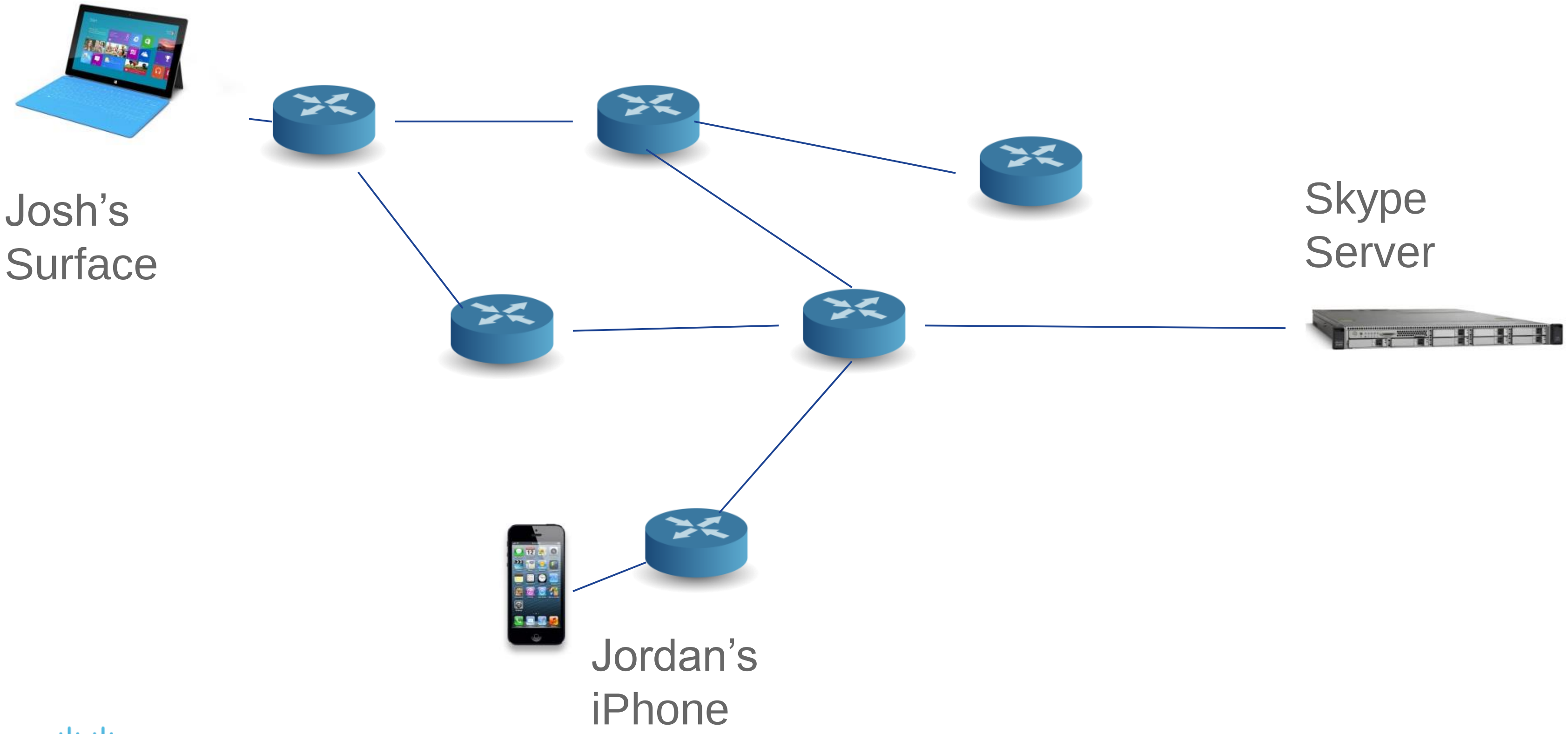
The body of the IP packet is a TCP message – so next we send the TCP header



The body of the TCP message is the content to send – an HTTP message. It is the HTTP header, followed by the HTTP body



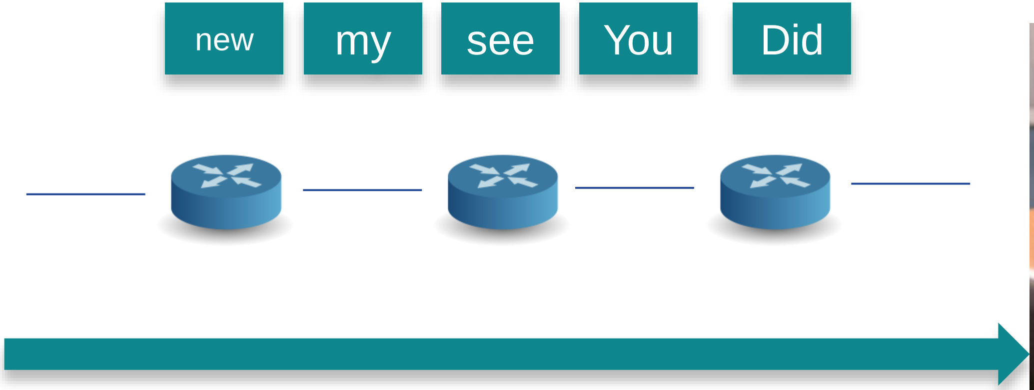
And Now – Voice over IP



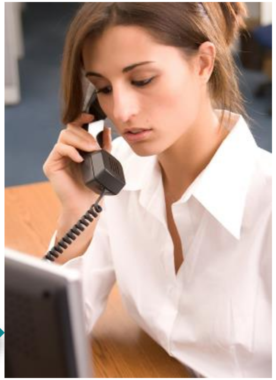
What we want to Happen

Did you see my new computer?

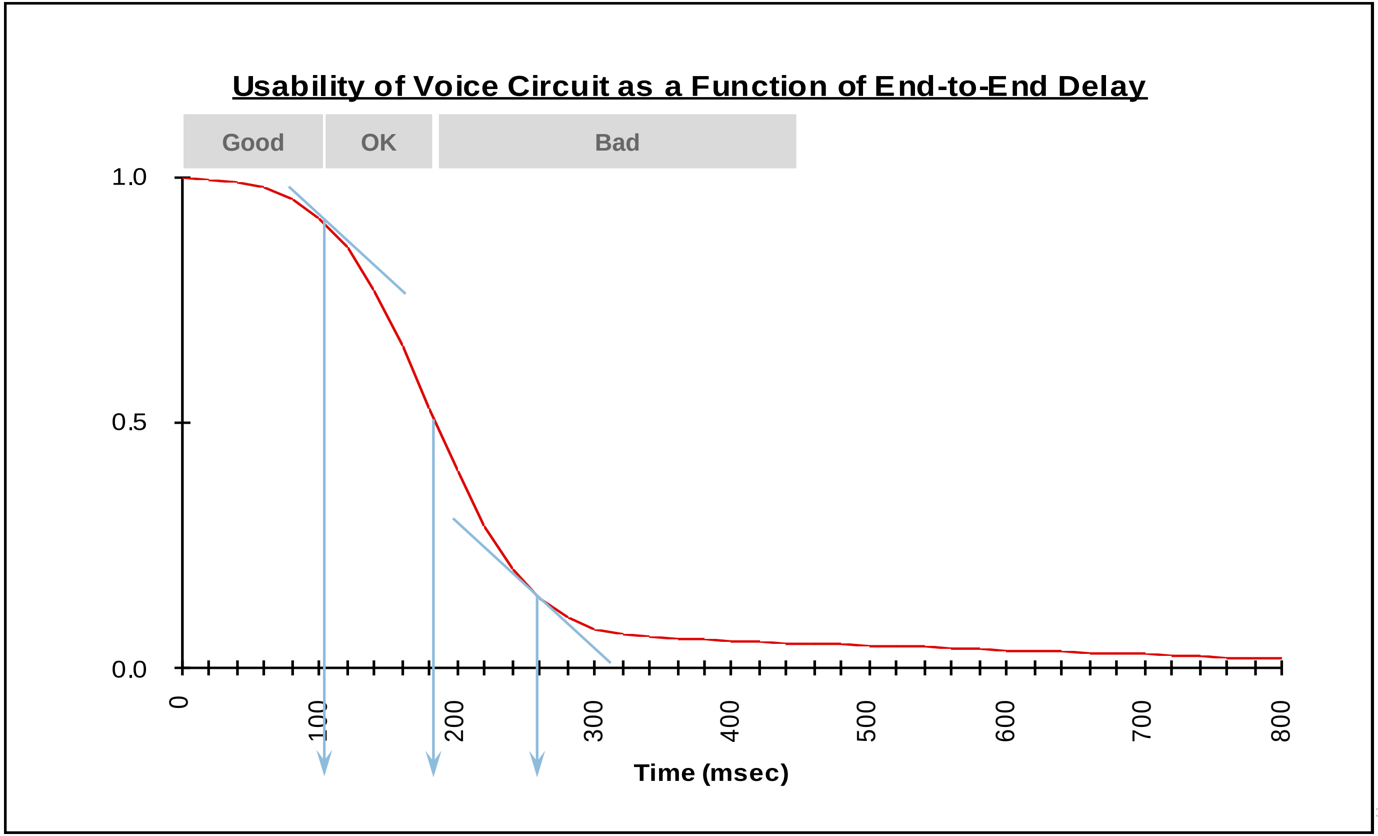
Packets containing Josh's words



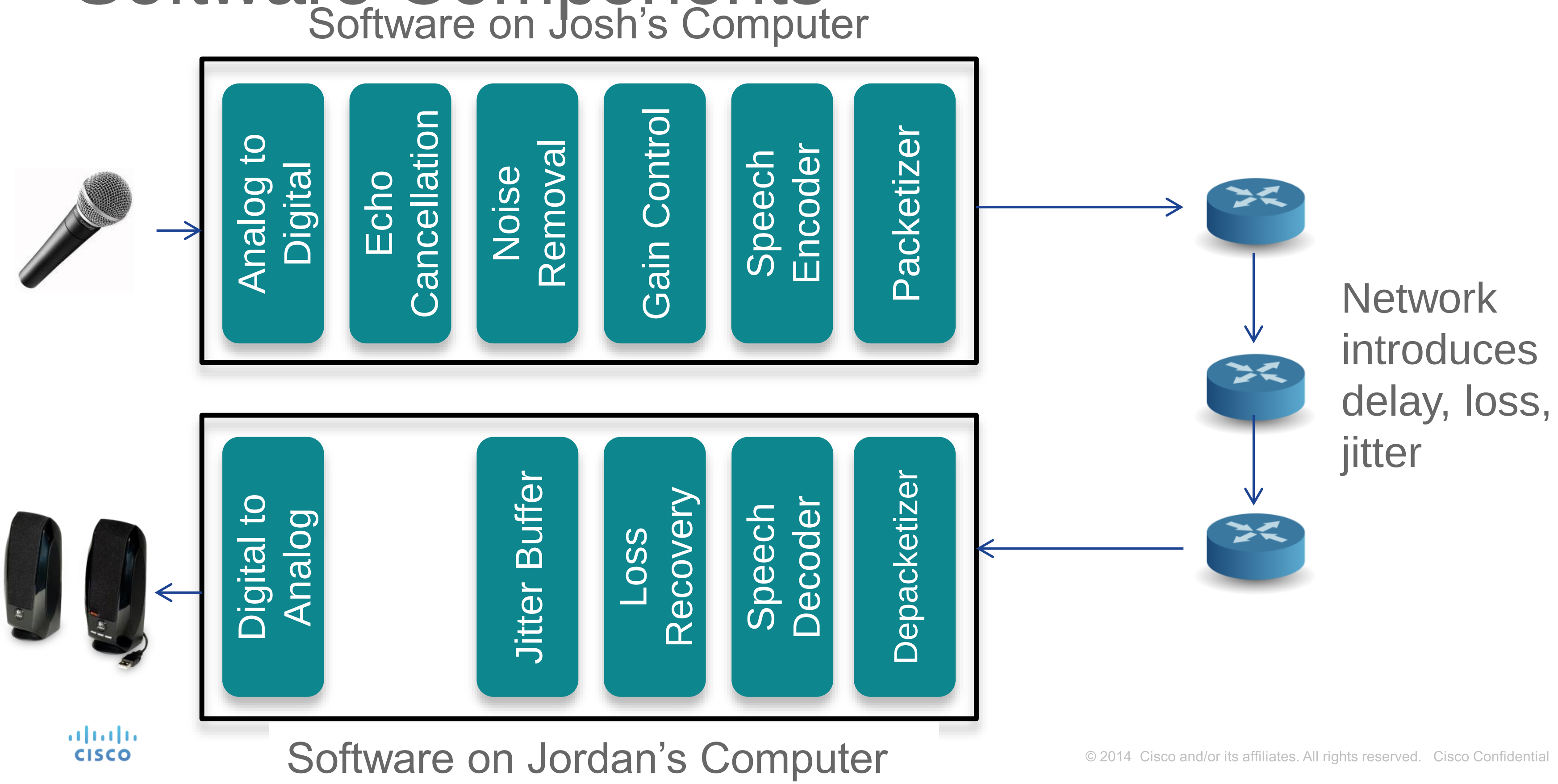
Fast!



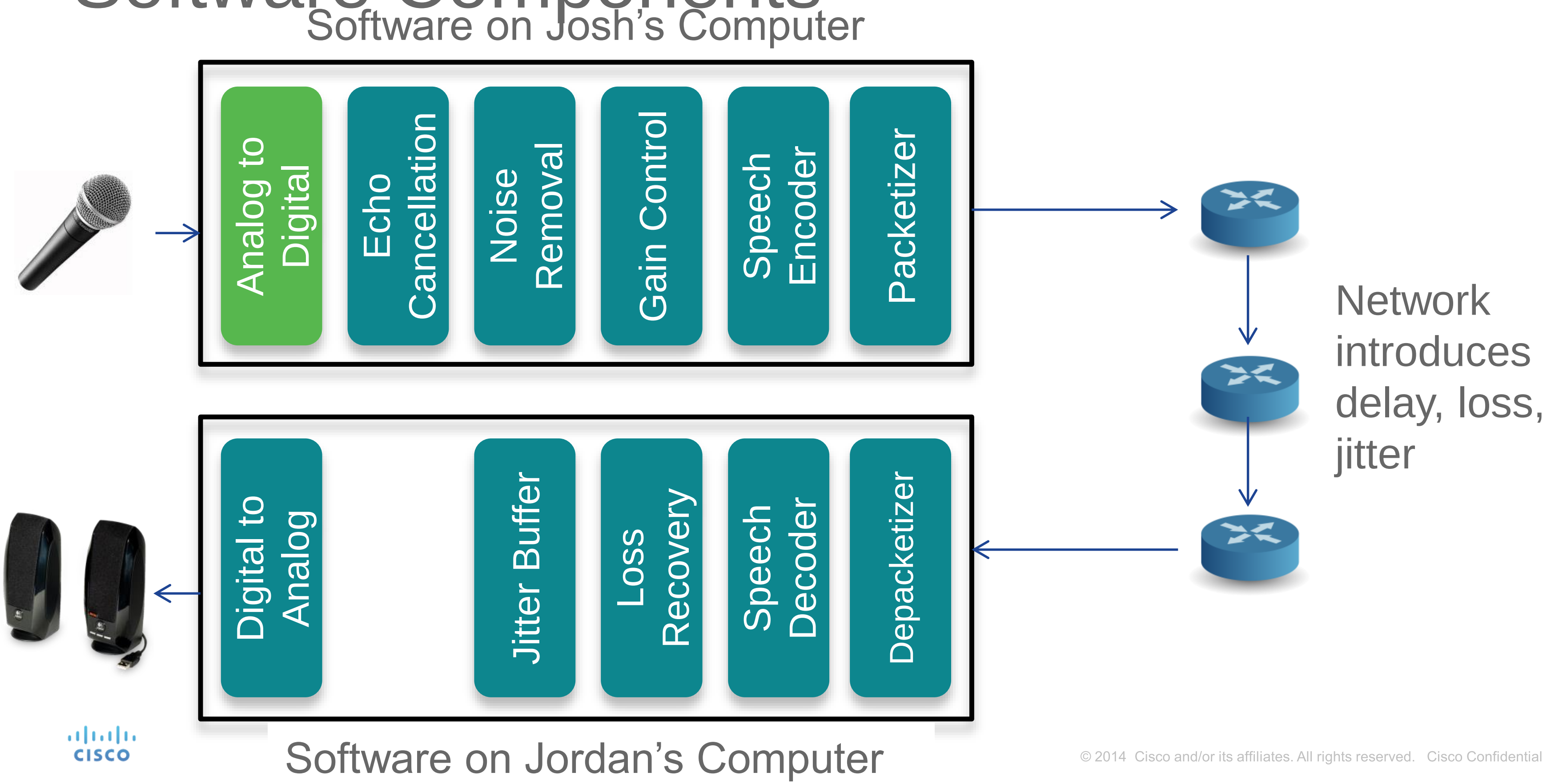
How Fast?


Usability of Voice Circuit as a Function of End-to-End Delay
Good
OK
Bad
1.0
0.5
0.0
0
100
200
300
400
500
600
700
800
Time (msec)

Software Components

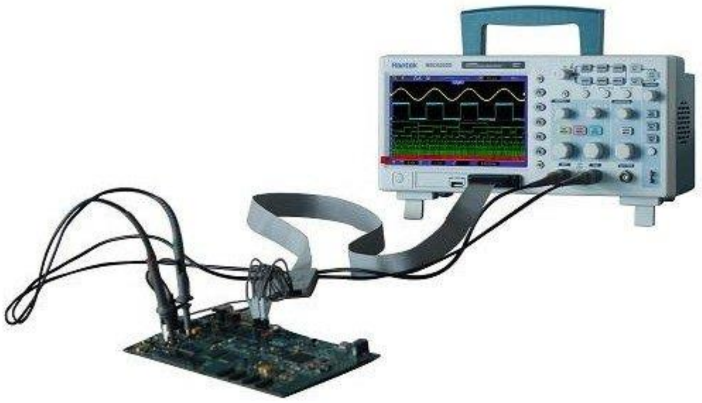


Software Components

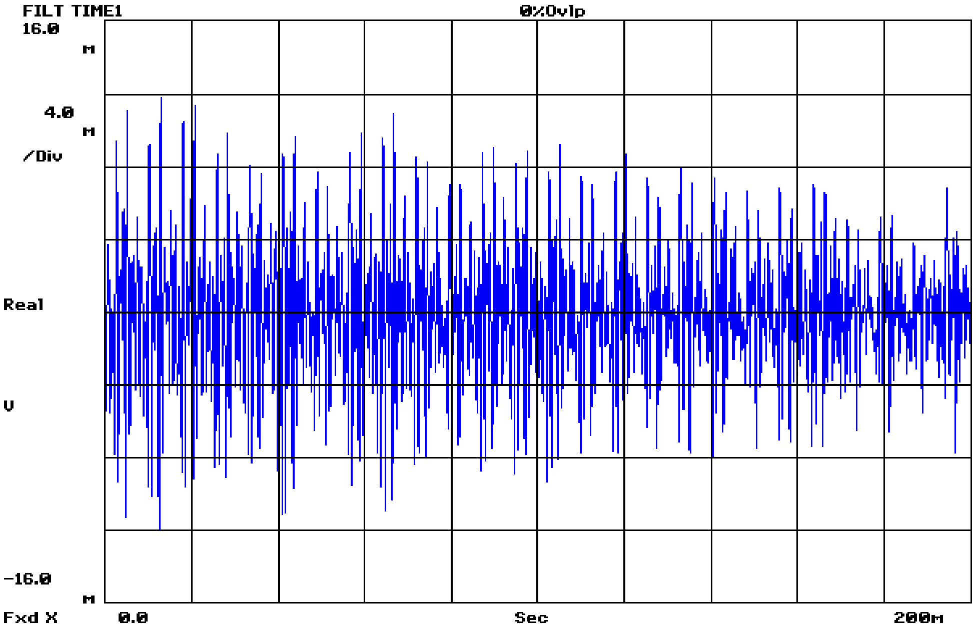


What is "Analog"?

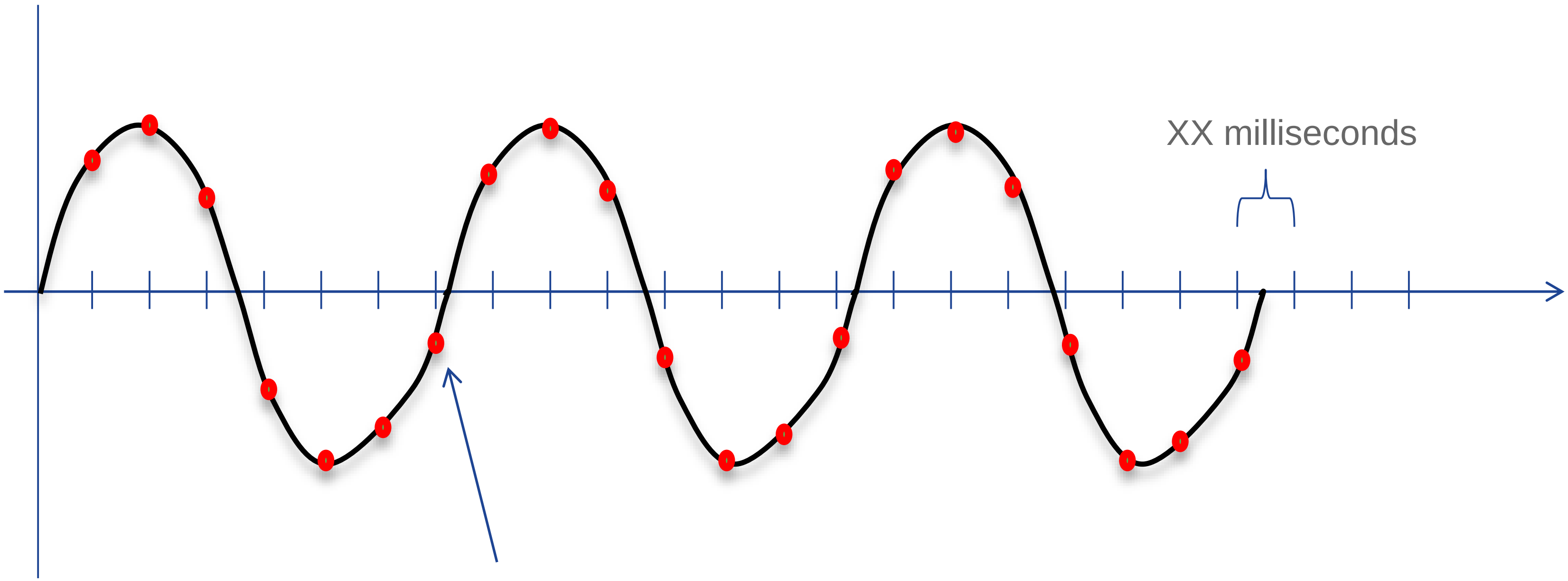
- An analog signal is a variable voltage on a wire, often measuring something in the real world
- Examples:
    - The wiring in your home
    - The output of an electro cardiogram that measures your heartbeat



Speech Starts out as Analog Too

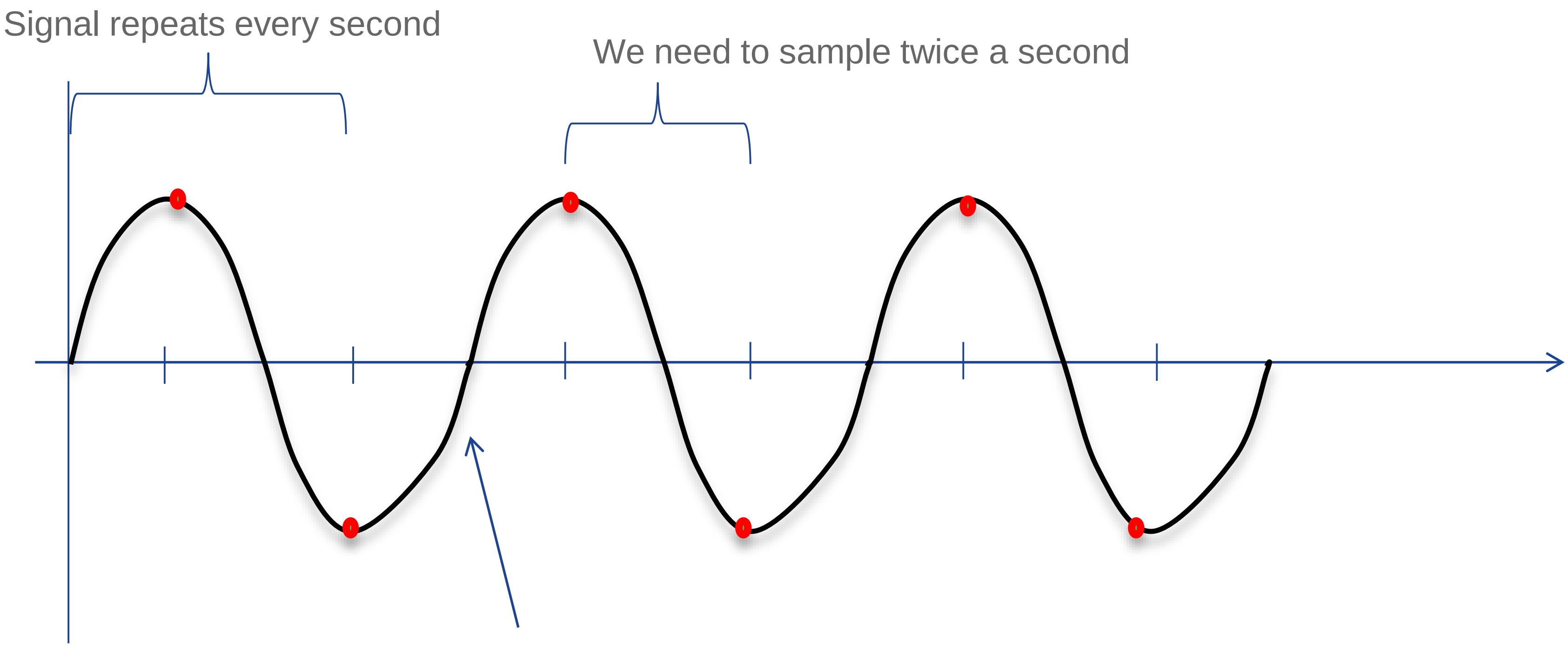


Analog to Digital - Sampling



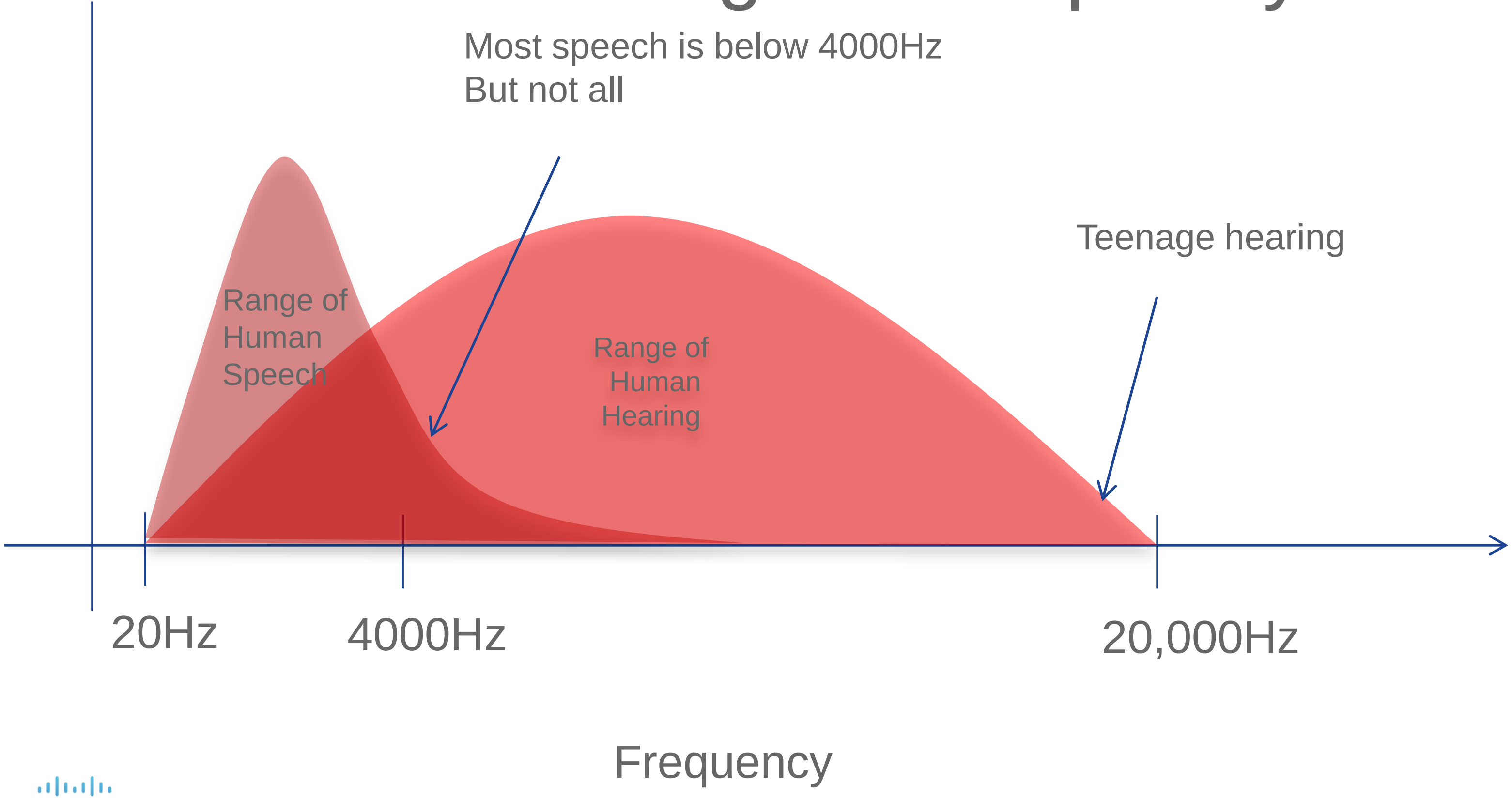
Every XX milliseconds, record the value of the analog signal using a binary number of length YY bits

How often do we need to sample?

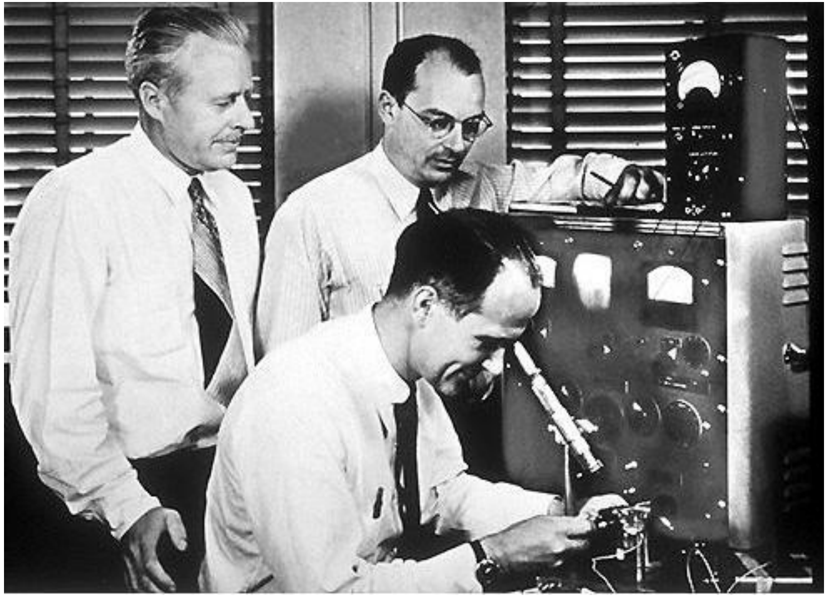


A famous math theorem shows that – we need to sample twice as often as the highest frequency of the signal

So – what is the highest frequency in voice?

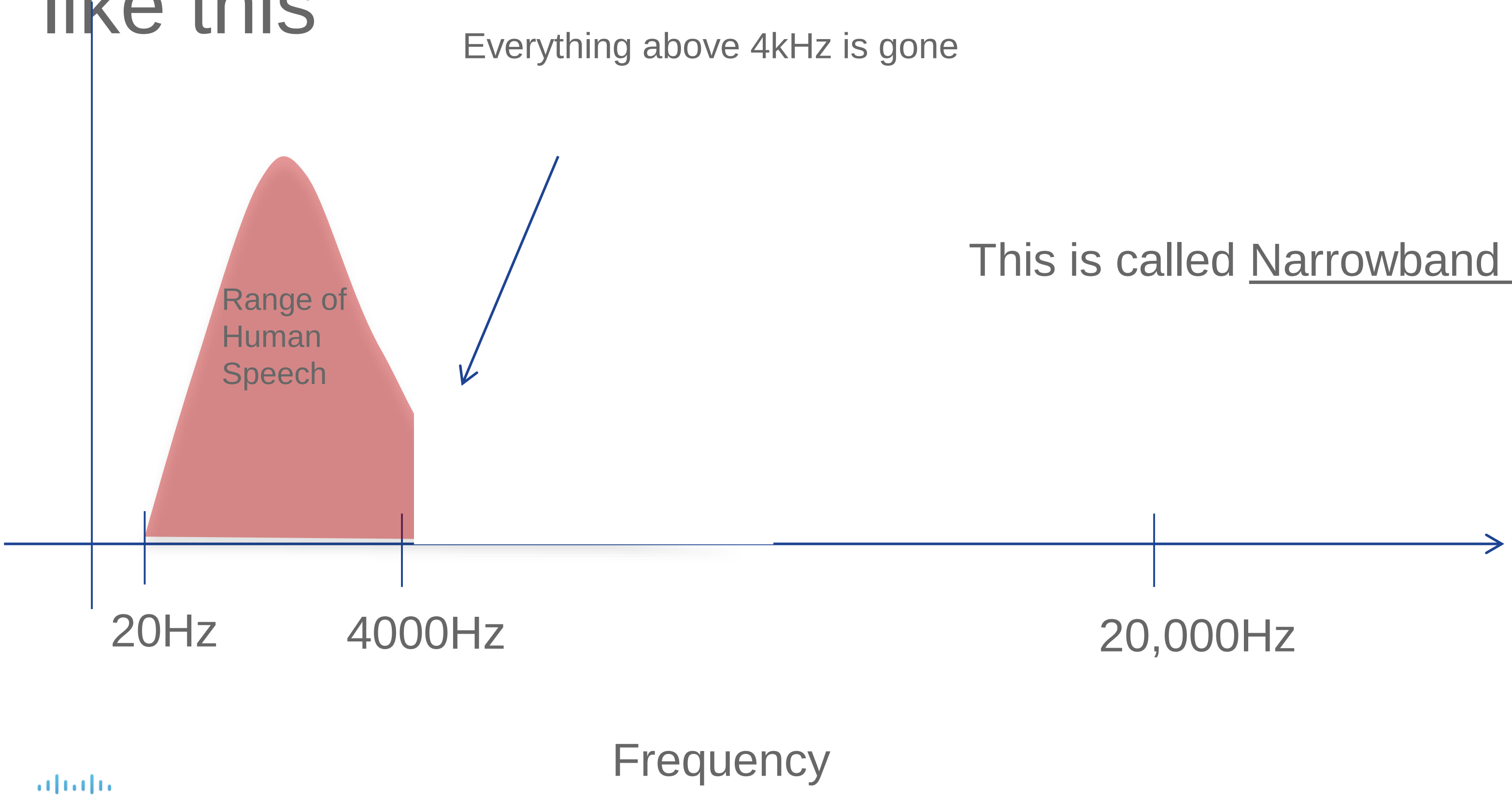


The Big Decision 60 Years Ago



- Conversion of telephone network from analog to digital in 1950s
- AT&T Engineers decide to cut off all speech above 4kHz
- Send digital speech through phone network with 8 bits per sample
- (8 bits * 2 * 4 kHz) = 64 kbps

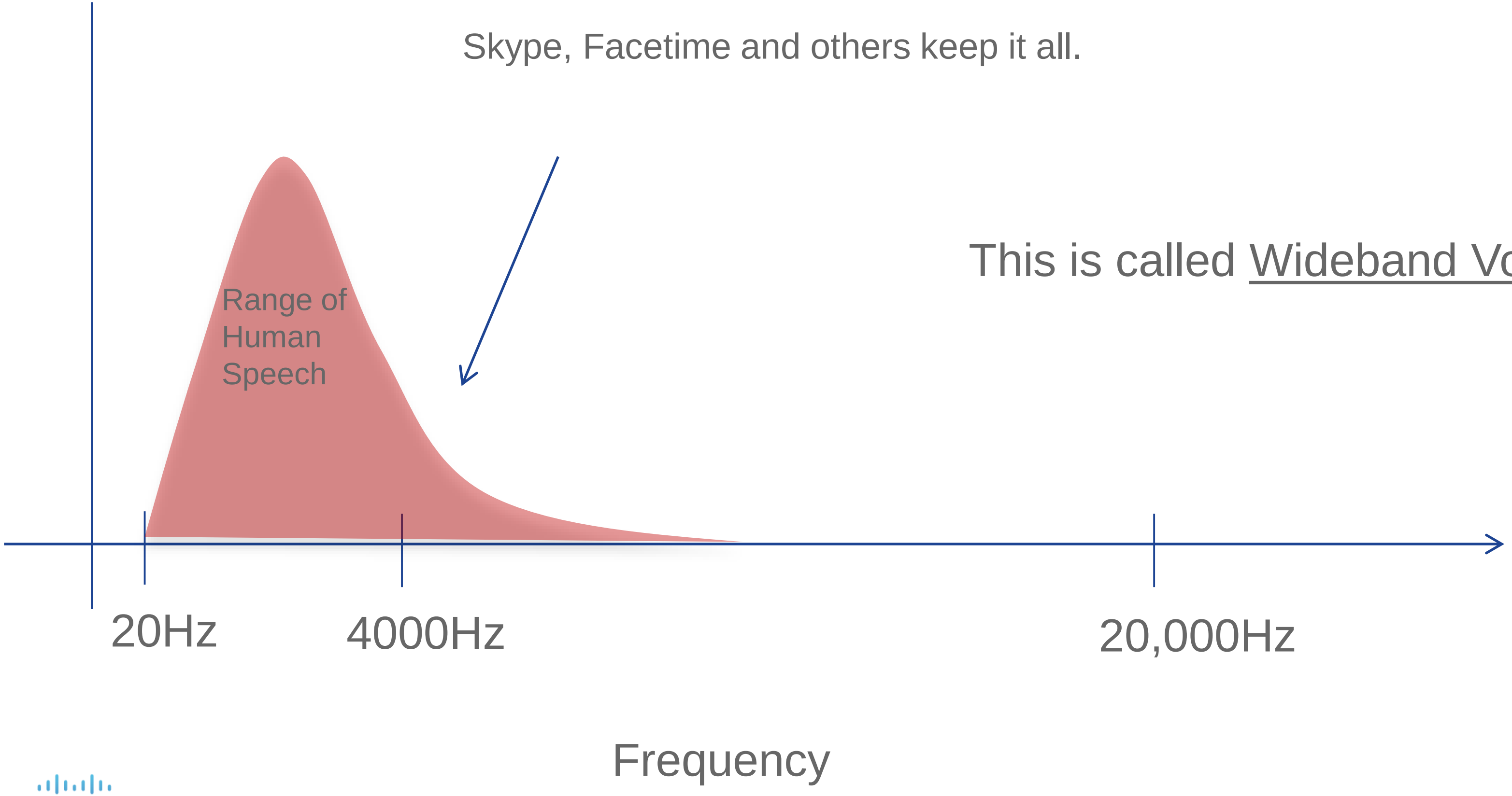
And so all speech on the phone network is like this



This is called Narrowband Voice.

With Voice over IP we can do better.

Skype, Facetime and others keep it all.



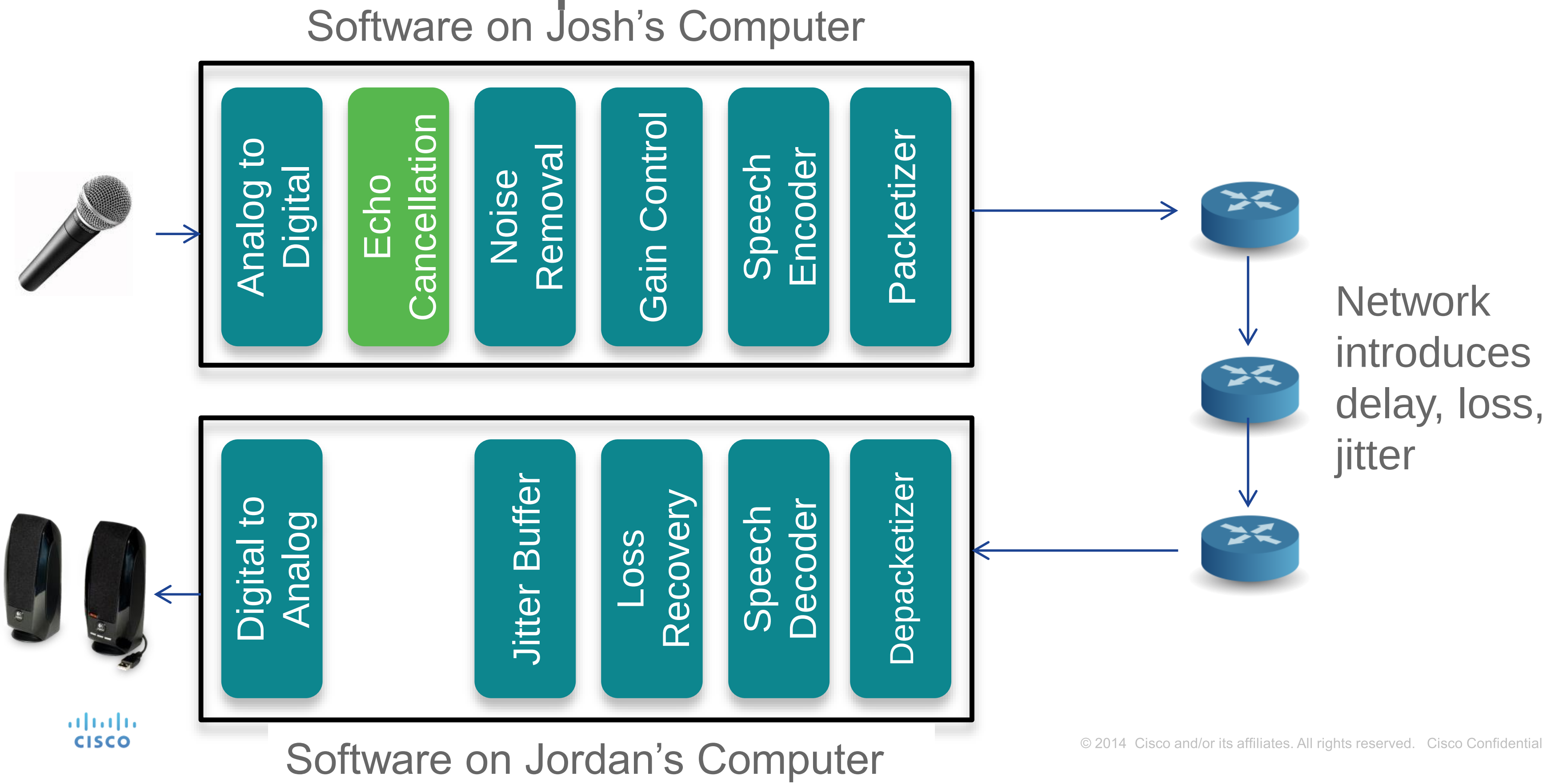
This is called Wideband Voice.

Does it Matter? You decide!

Narrowband Voice

Wideband Voice

Software Components



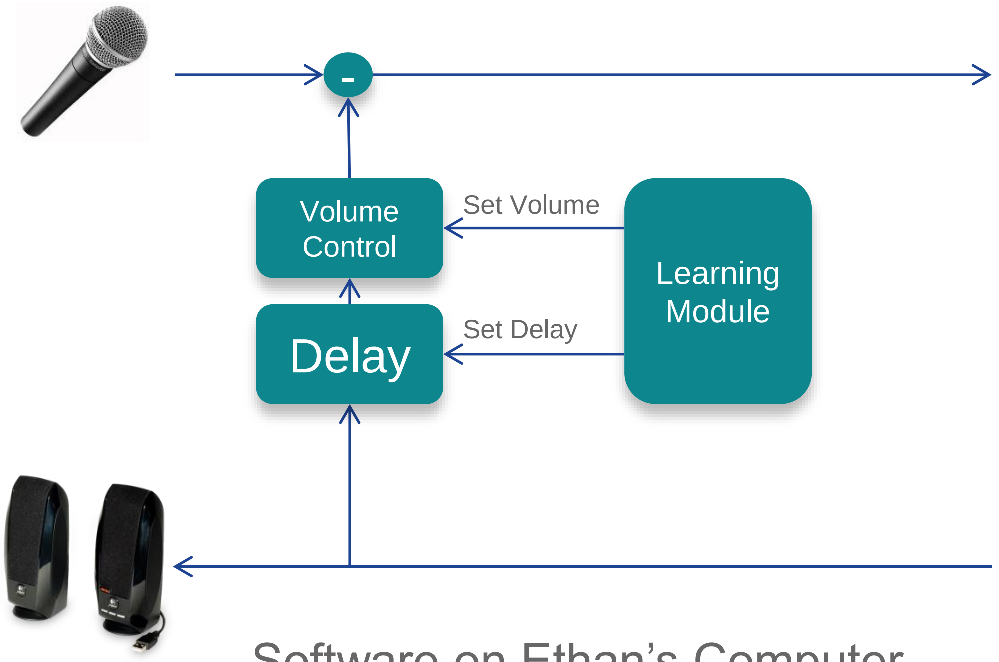
The Echo Problem

Sound comes out of the speaker

And the mic picks it up

The girl will hear what she said echoed back to her!

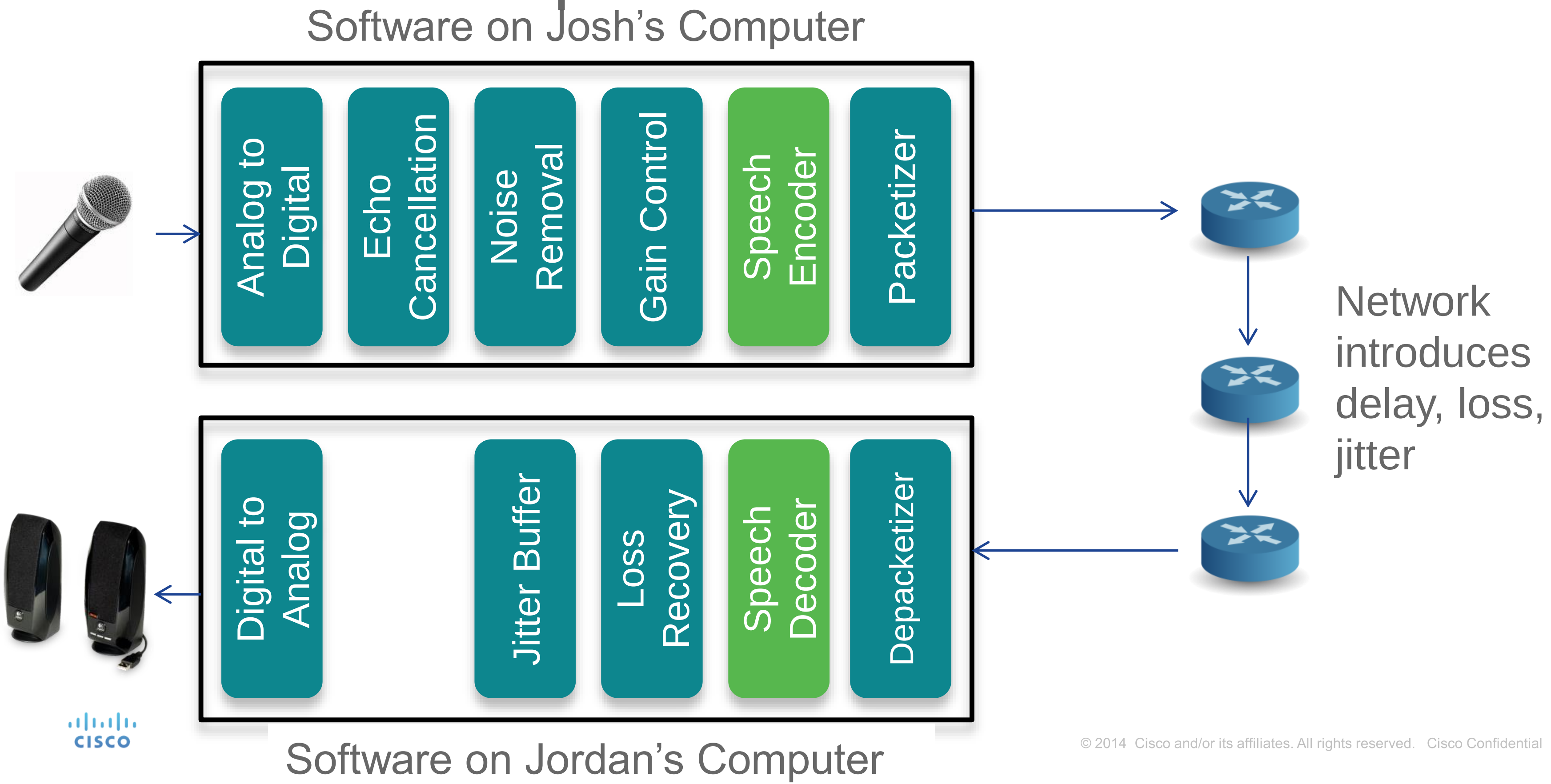
Echo Canceller Software



Software on Ethan's Computer

Learning Module figures out the acoustic properties of the room by listening

Software Components

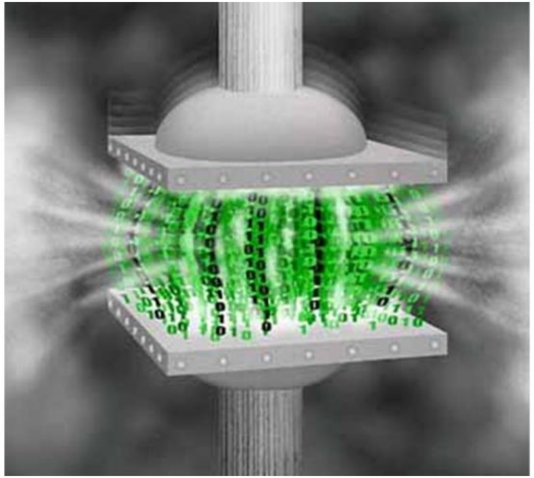


Speech Coding – What is it?

Wideband Speech

300 kbps

This is a lot! More than many home broadband speeds.

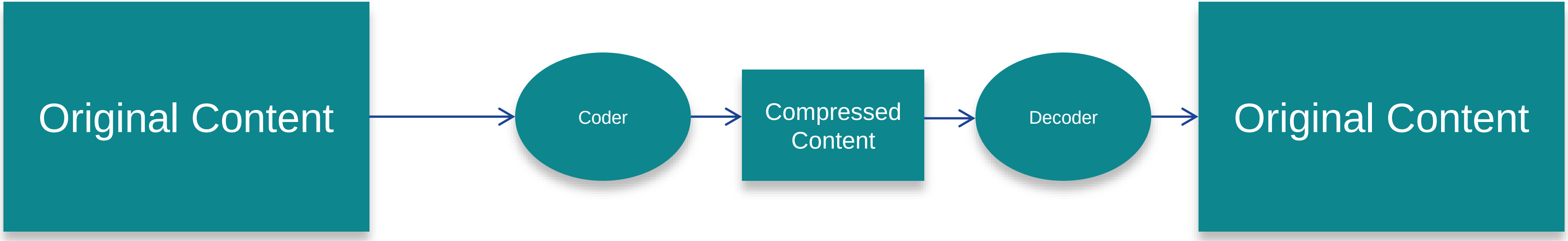


Compressed Wideband Speech

40 kbps

Compressed or coded speech takes up less bandwidth.

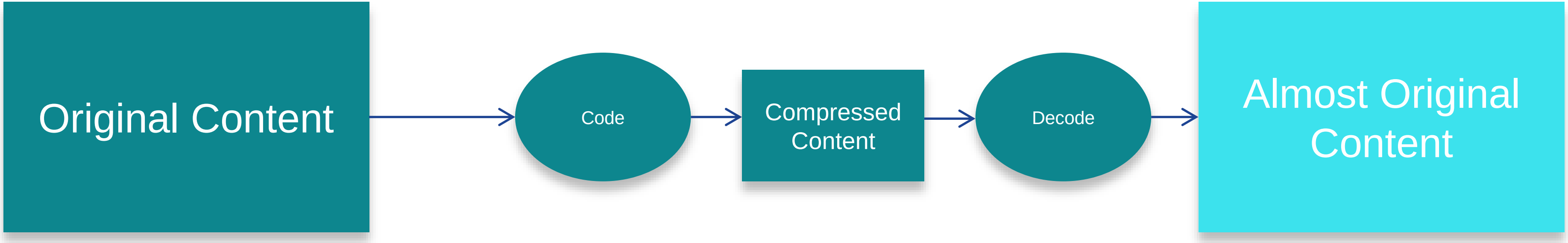
Lossless Coding



In Lossless Coding, you get back the exact same thing (bit by bit) after applying compression and then decompression.

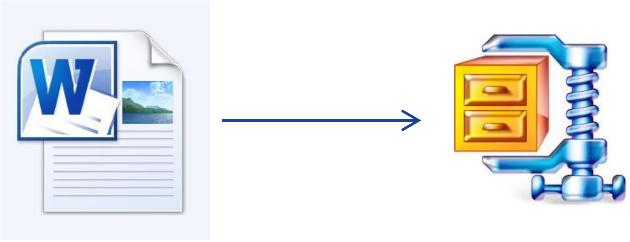
The combination of a coder and decoder is called a codec.

Lossy Coding



In Lossy Coding, you don't get back the exact same thing (bit by bit) after applying compression and then decompression. Lossless compression can compress even more than lossy.

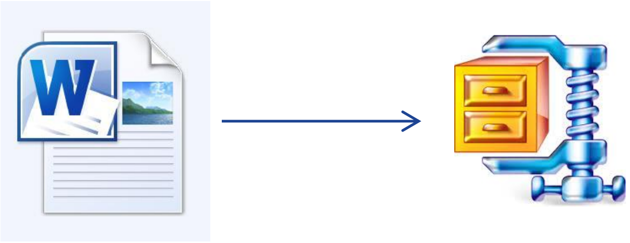
Guess the Type!



ZIP

??

Guess the Type!



ZIP

LOSSLESS

Guess the Type!

JPG

??

Guess the Type!

JPG

LOSSY

Guess the Type!

PNG

??

Guess the Type!

PNG

LOSSLESS

This is why PNG files are larger than JPG for the same image.

How do we do lossy compression of voice?

You might hear two sounds that – to you – sound alike – even though a computer would think they are different!

The behavior of the brain in recognizing speech is very well understood

Your brain is not like a computer.

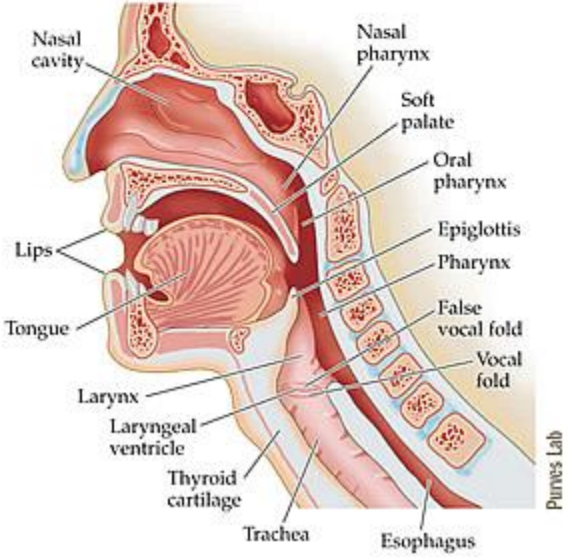
We can remove information from the speech that we know you’re brain won’t notice!

But how does it work?

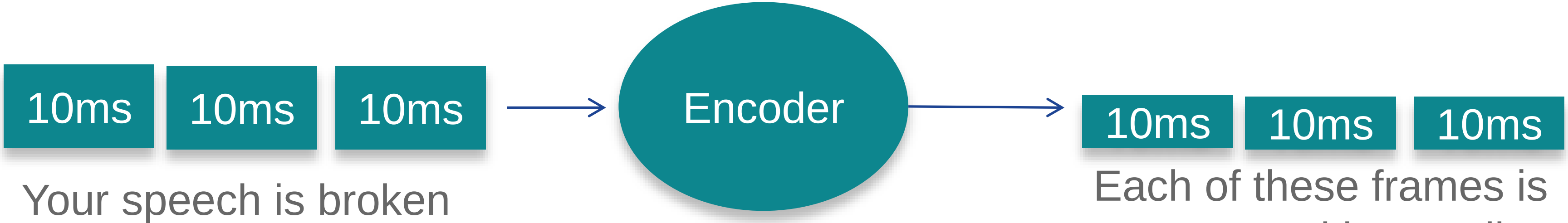
The encoder listens to you talk and builds a mathematical model of your vocal system!

It then sends the parameters of this model to the decoder which rebuilds the speech.

Example: your **pitch** – or core frequency of your voice.



Encoding Speech Frames



Your speech is broken
into 10ms
frames of uncompressed
speech

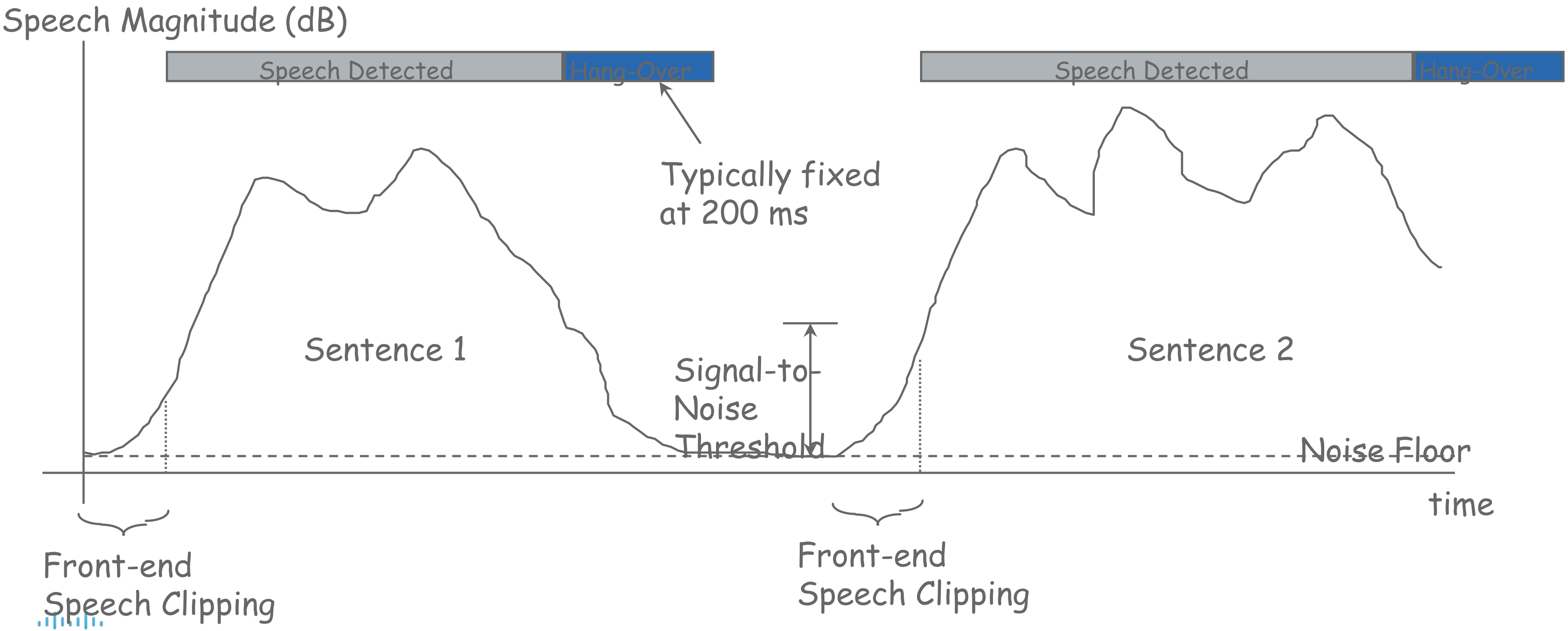
Each of these frames is compressed by sending the parameters of the mathematical model.

Decoding one frame however relies on having gotten the previous one!

Speech Coding Fun Facts

1. Speech coding is an arcane field of mathematics and computer science that is understood by very few people in the world
2. There are many speech codecs, and they usually have funny monikers like “G.729” and “OPUS”
3. The codec used in most cell phone networks is called AMR and it can compress your voice as low as 4.75 kbps
4. Speech codecs have traditionally been different from music codecs; speech codecs due a bad job at music. This is why music-on-hold on the telephone sounds so awful

Voice Activity Detection and Comfort Noise



Video Coding

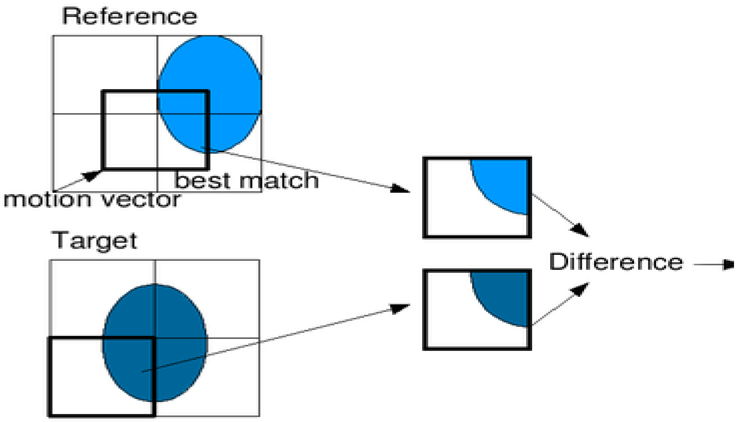
Key Terms

Term	Description
Frame	An individual picture in a sequence that makes up the video
Frame Rate	The number of frames per second in video. 30 is excellent (TV quality)
Resolution	The number of horizontal and vertical pixels. VGA=640x480.
Interlacing	A mechanism for transmitting video by splitting a frame into two fields, one field representing the odd lines, and one the even field. This is the "i" in 1080i
Progressive	As opposed to interlaced, a method for transmitting video by sending each frame as a whole.
HD	High Def resolutions – 720p is 1280x720 with 60fps. 1080i is 1920x1080 at 30fps

Key Concept: Macroblocks

Rectangular block in an image which is a basic unit of compression. Typically 16x16 pixels.

Key Concept: Inter-Frame Prediction



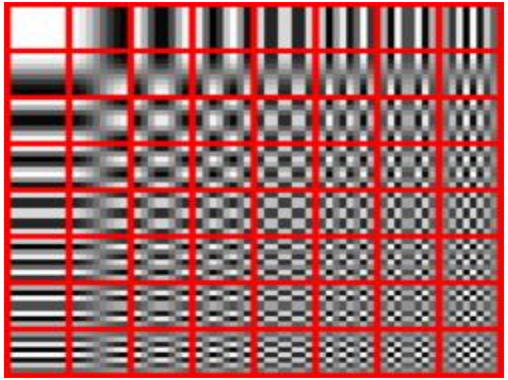
Encode

Predict information in the current frame by looking at previous frames, possibly taking into account motion.

Key Concept: Discrete Cosine Transform (DCT)

Increasing horizontal frequencies

Increasing vertical frequencies



A technique for representing a macroblock by its component frequencies. Discarding the higher frequencies throws away the finer details without losing the core image.

Video Coding Fun Facts

1. Encoding (compressing) requires much more computational horsepower than decoding (decompressing)
2. Video coding is a highly patented area and this has made it very difficult for everyone to agree on which one to use
3. The primary video code in usage today is called H.264 and most video on the Internet uses it
4. Speech coding is mature and many consider it “done”. Video coding is still an area of research and innovation.

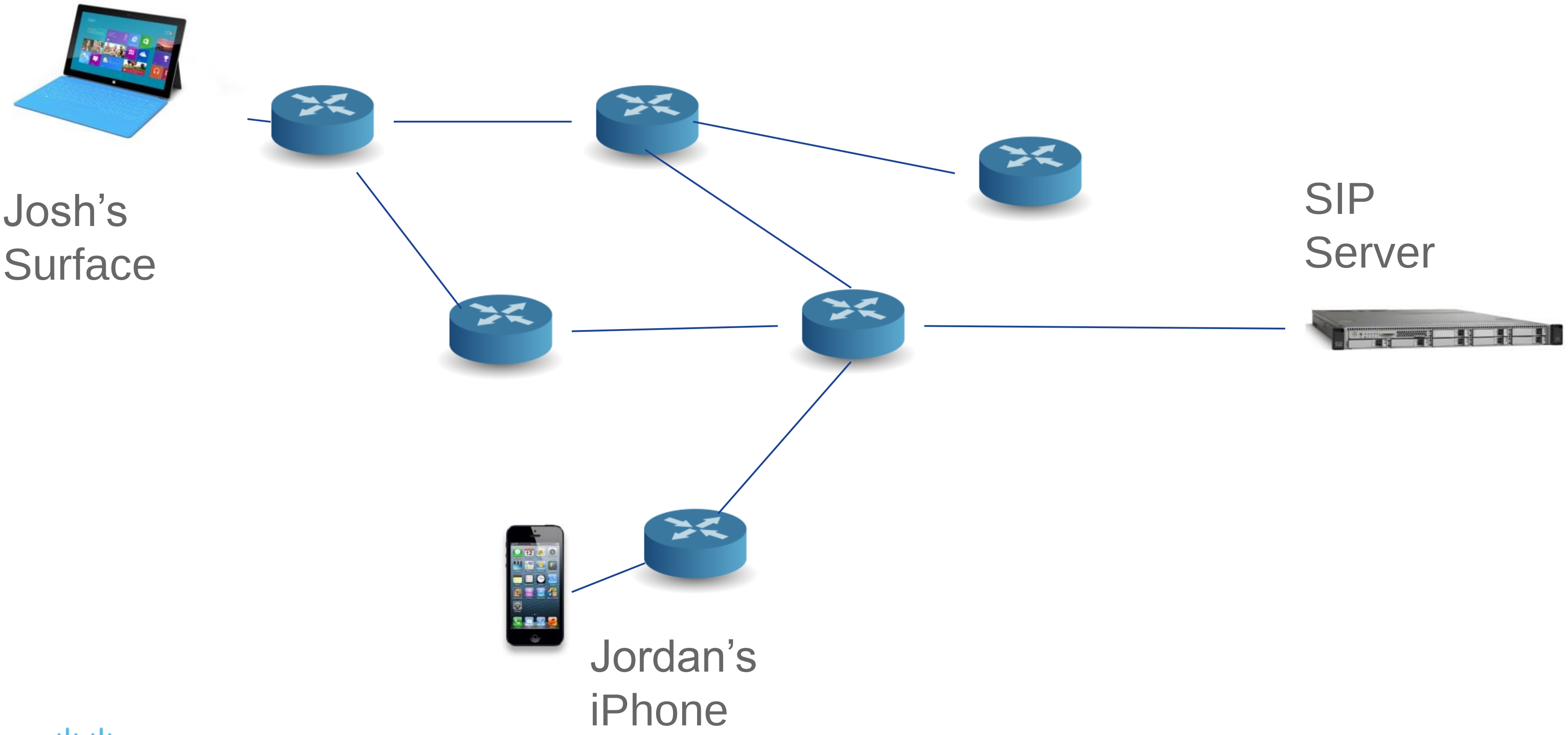
Signaling

In VoIP, Signaling is the process of controlling the flow of voice and video packets.

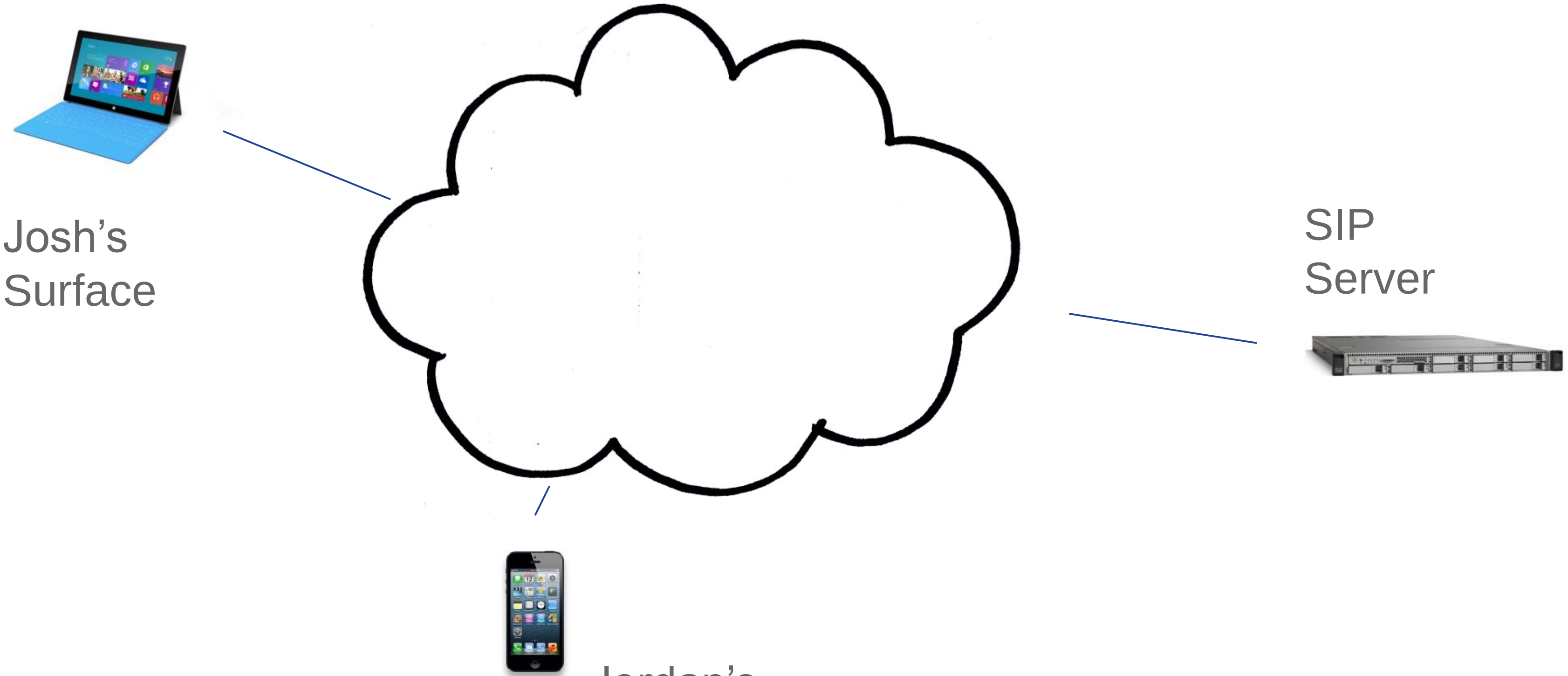
Signaling includes discovery, setting up the call, tearing it down, and related functions.

Signaling in VoIP is done using the Session Initiation Protocol (SIP).

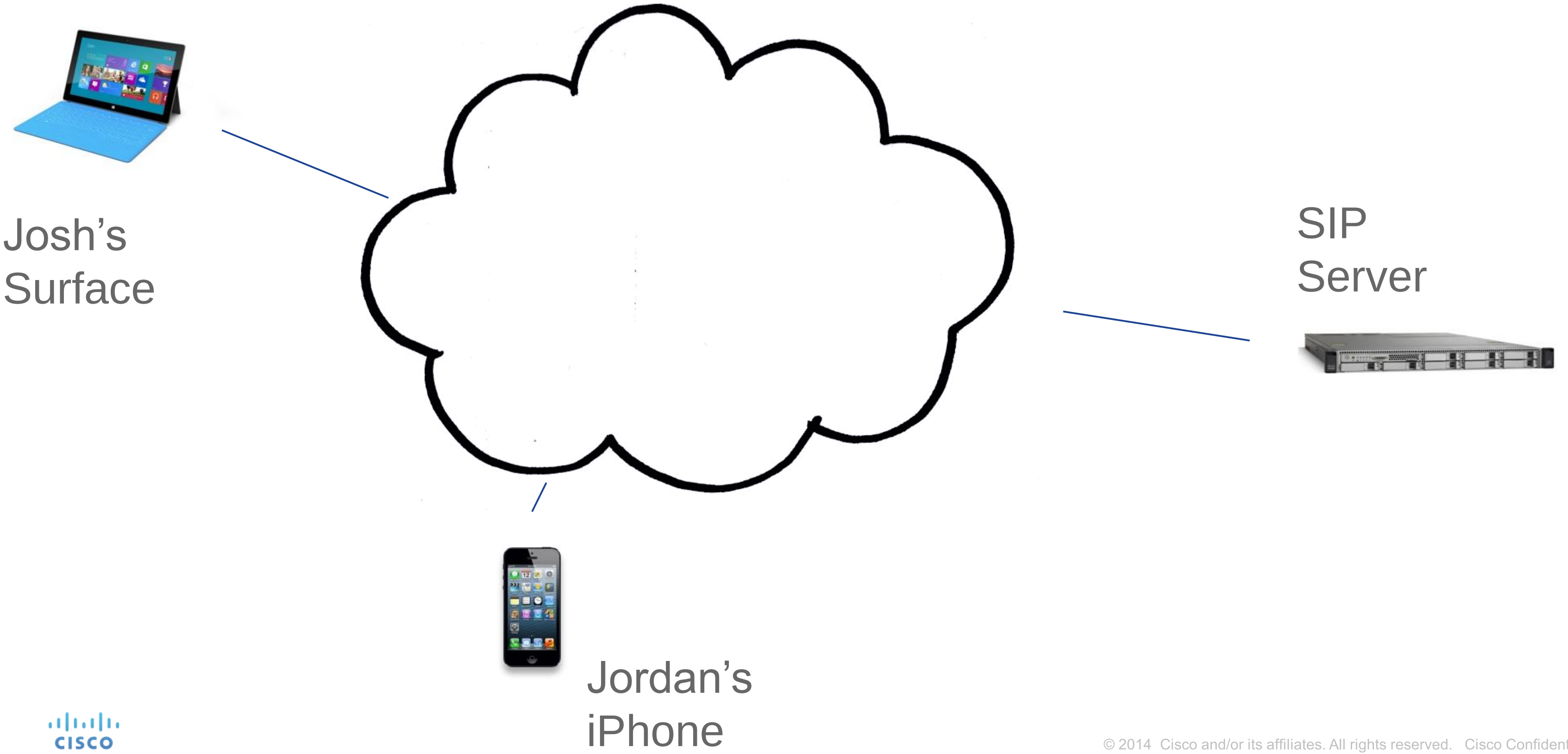
SIP Basics



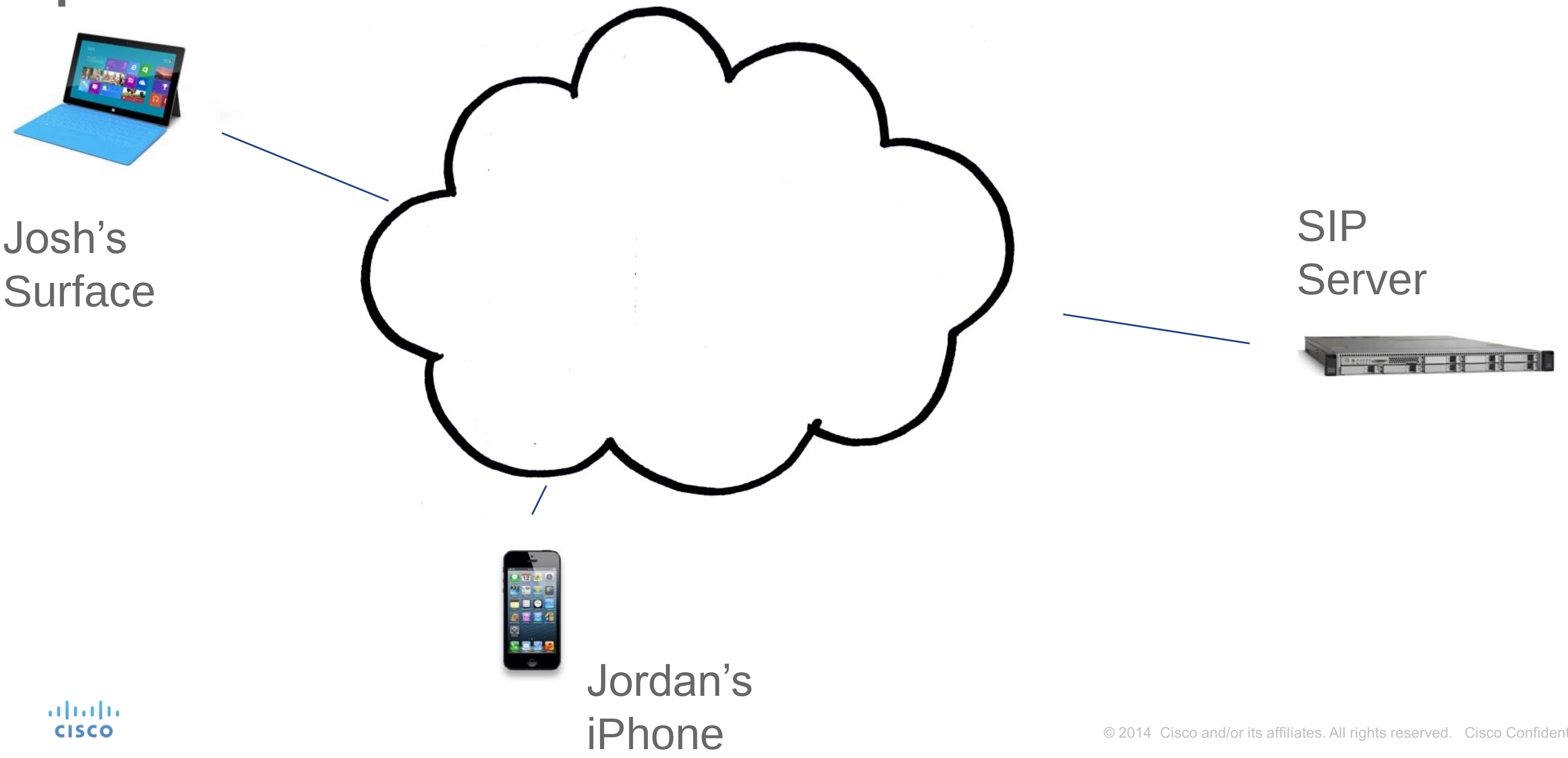
Lets Simplify



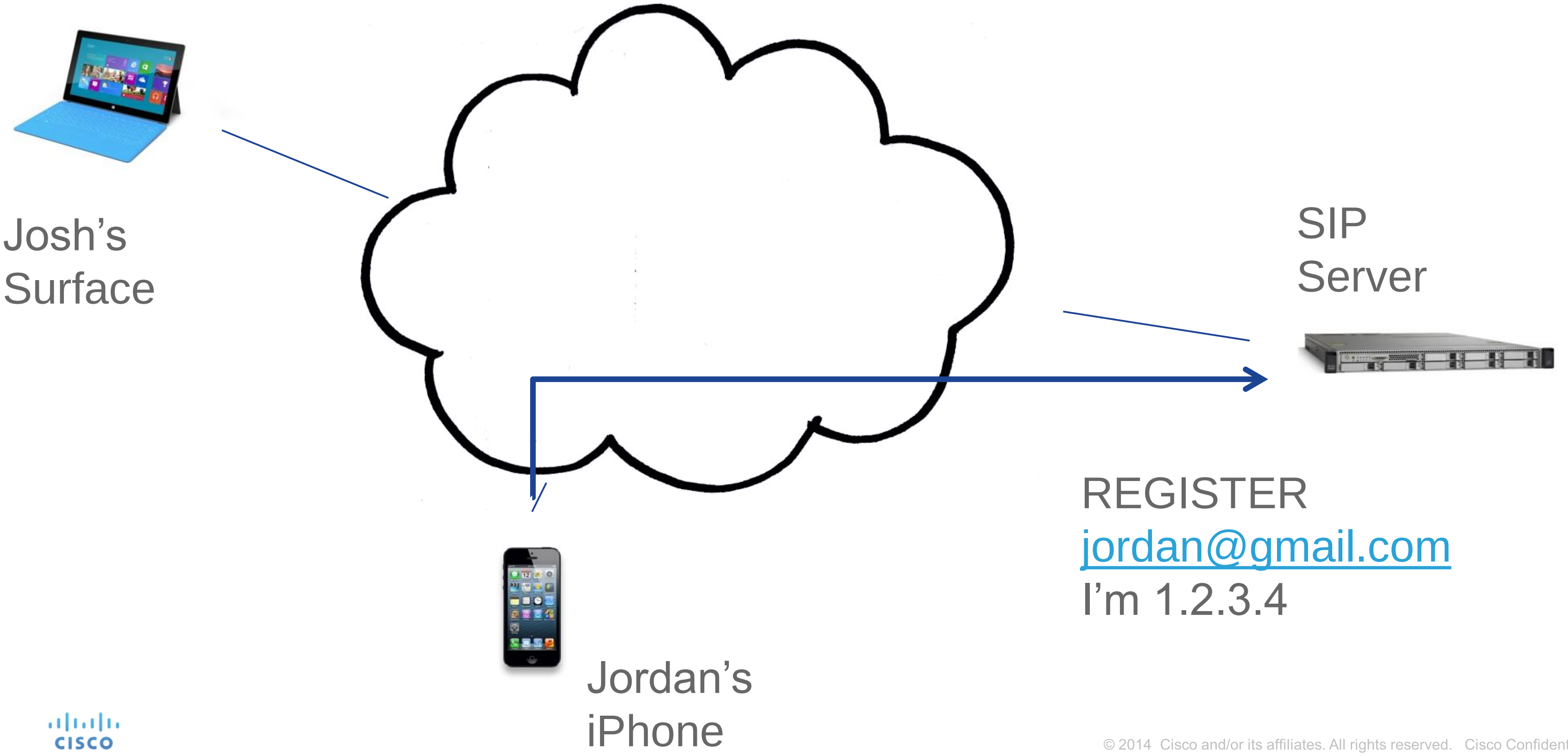
Josh Wants to call jordan@gmail.com



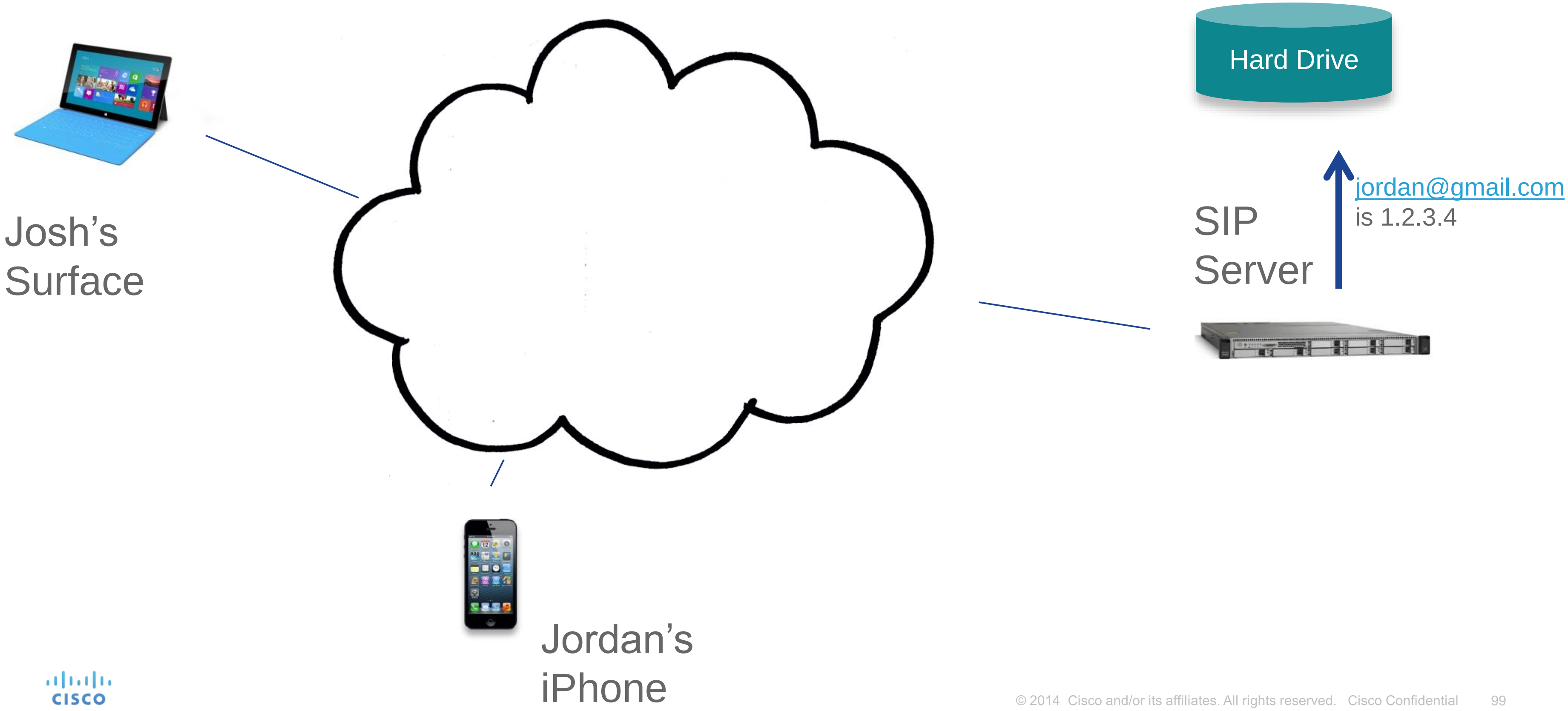
How does Josh’s computer find Jordan’s phone??



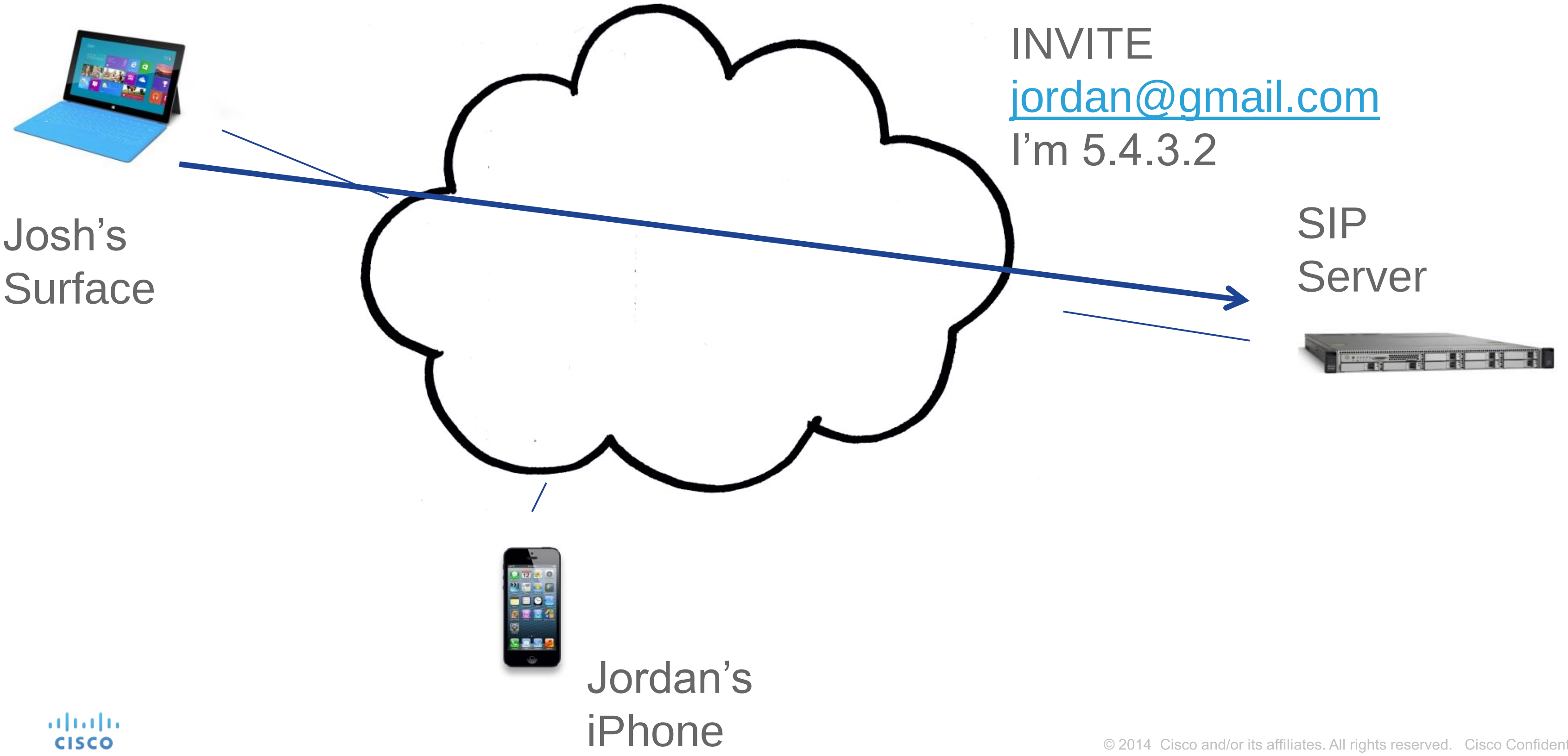
Registration



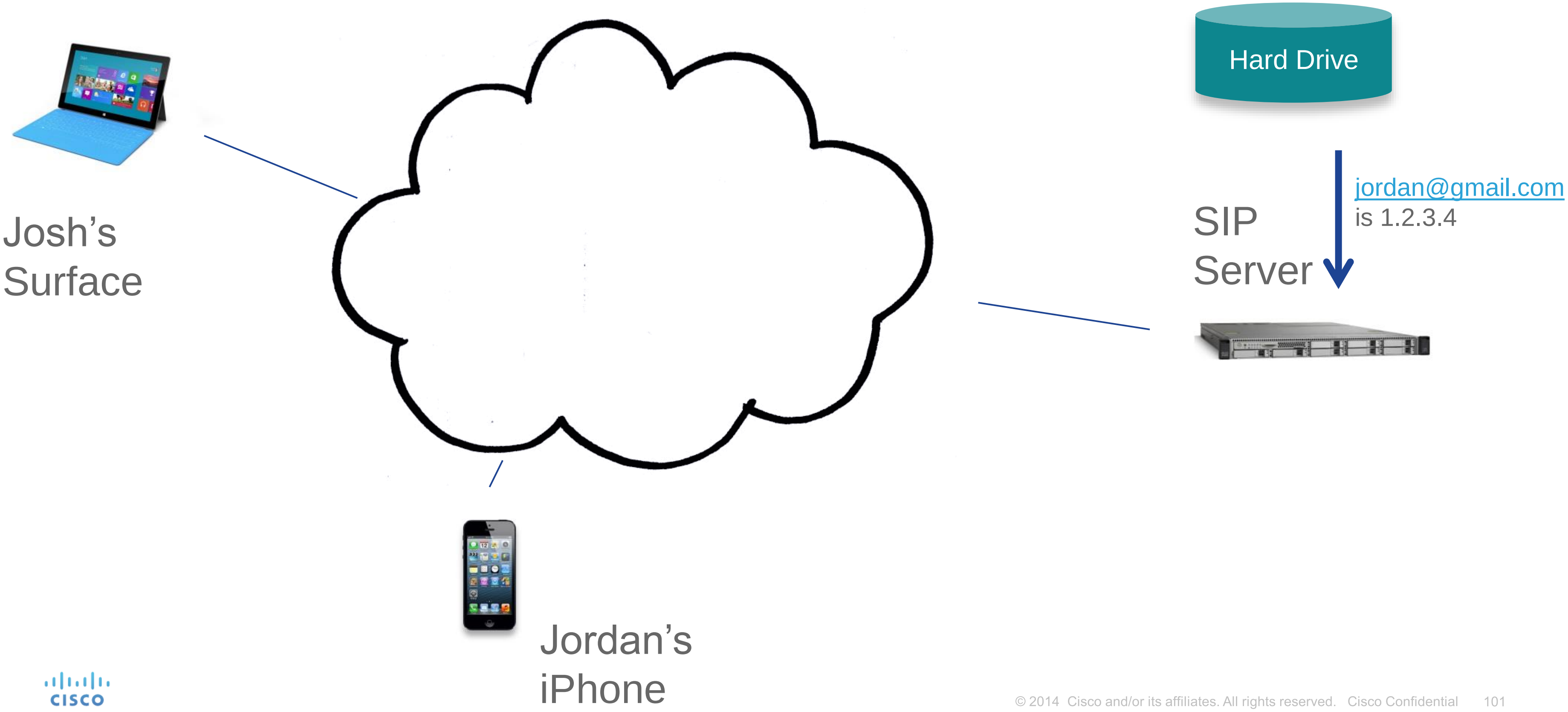
Registration



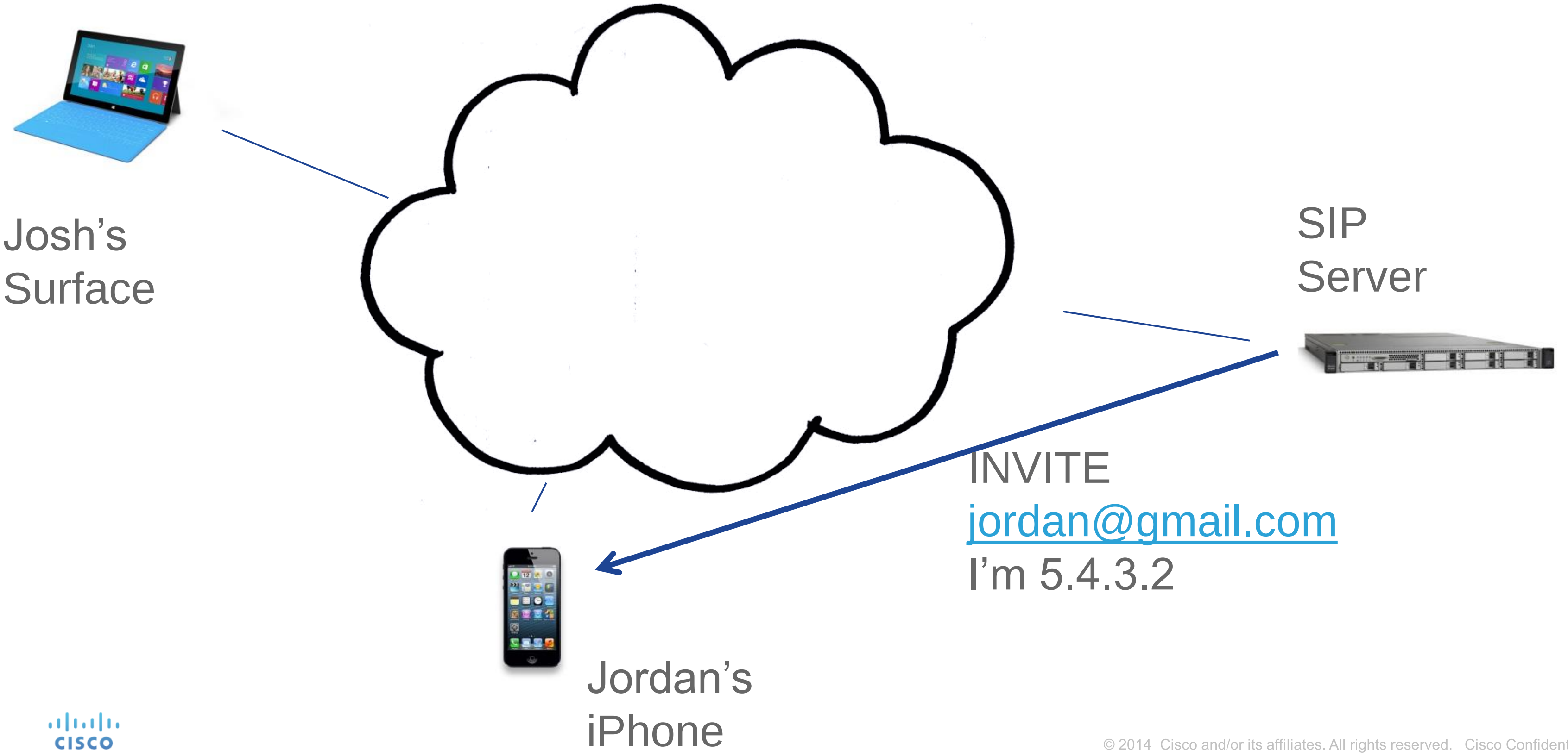
Josh makes the call



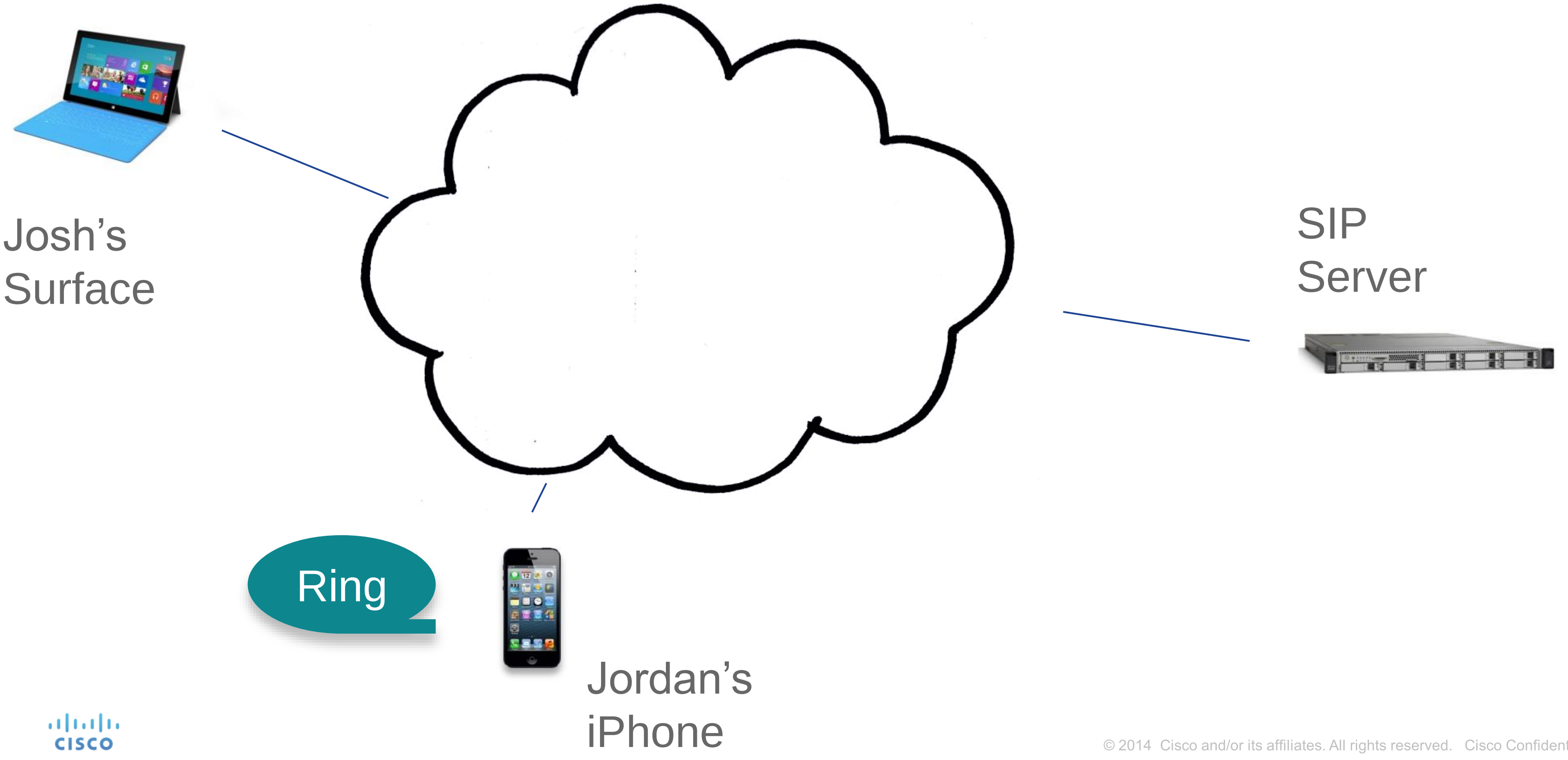
SIP Server Looks up Jordan



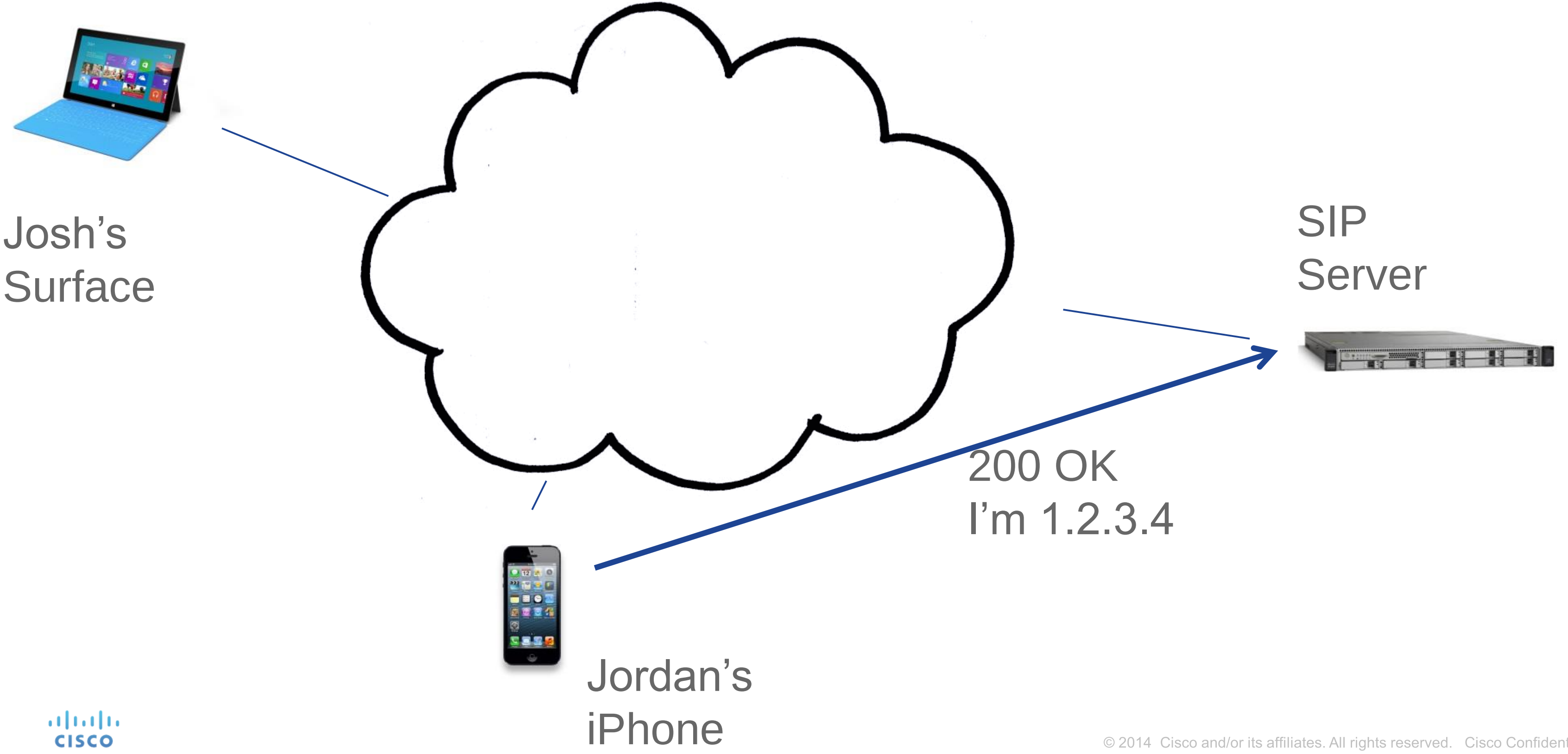
SIP Server Sends to Jordan’s phone



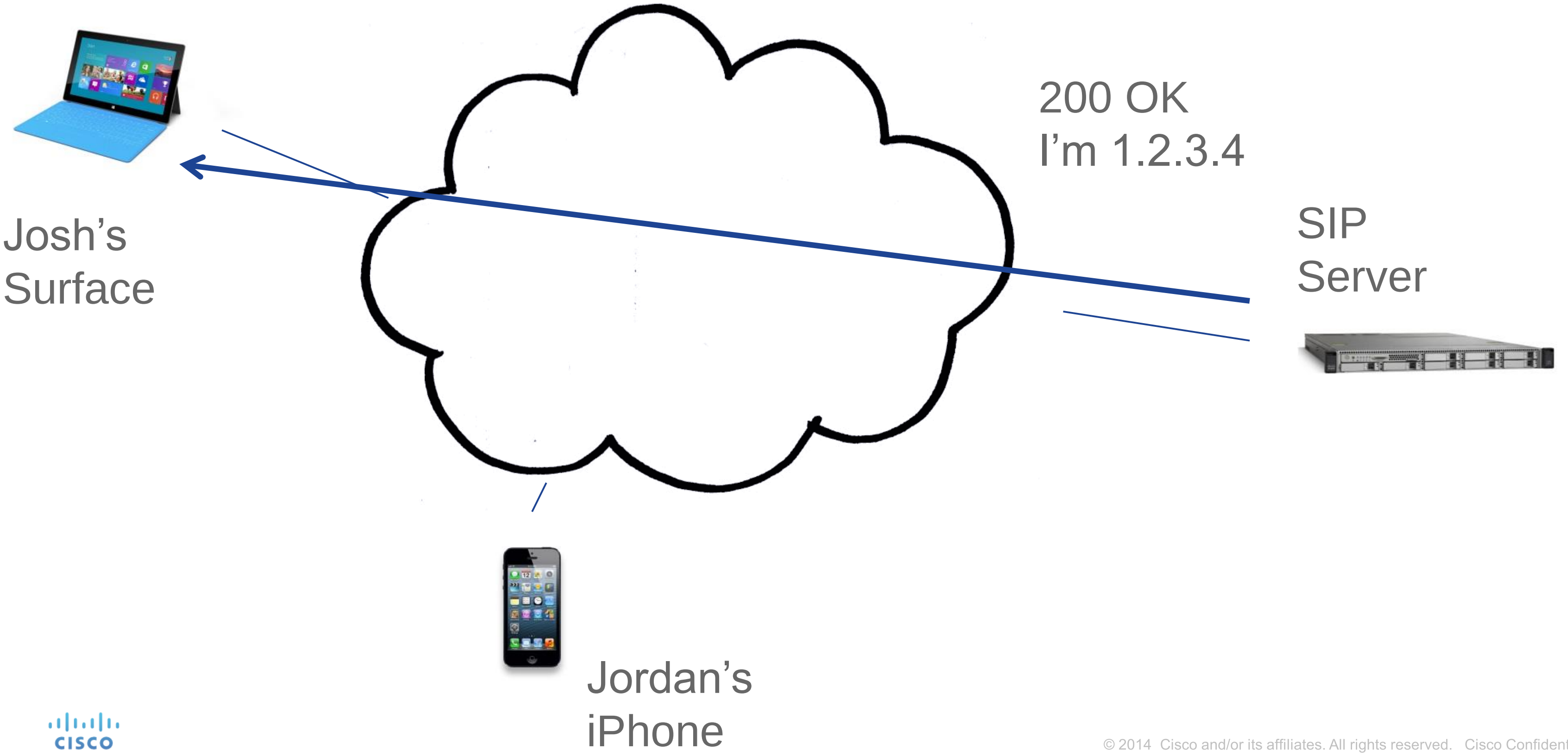
Ring!



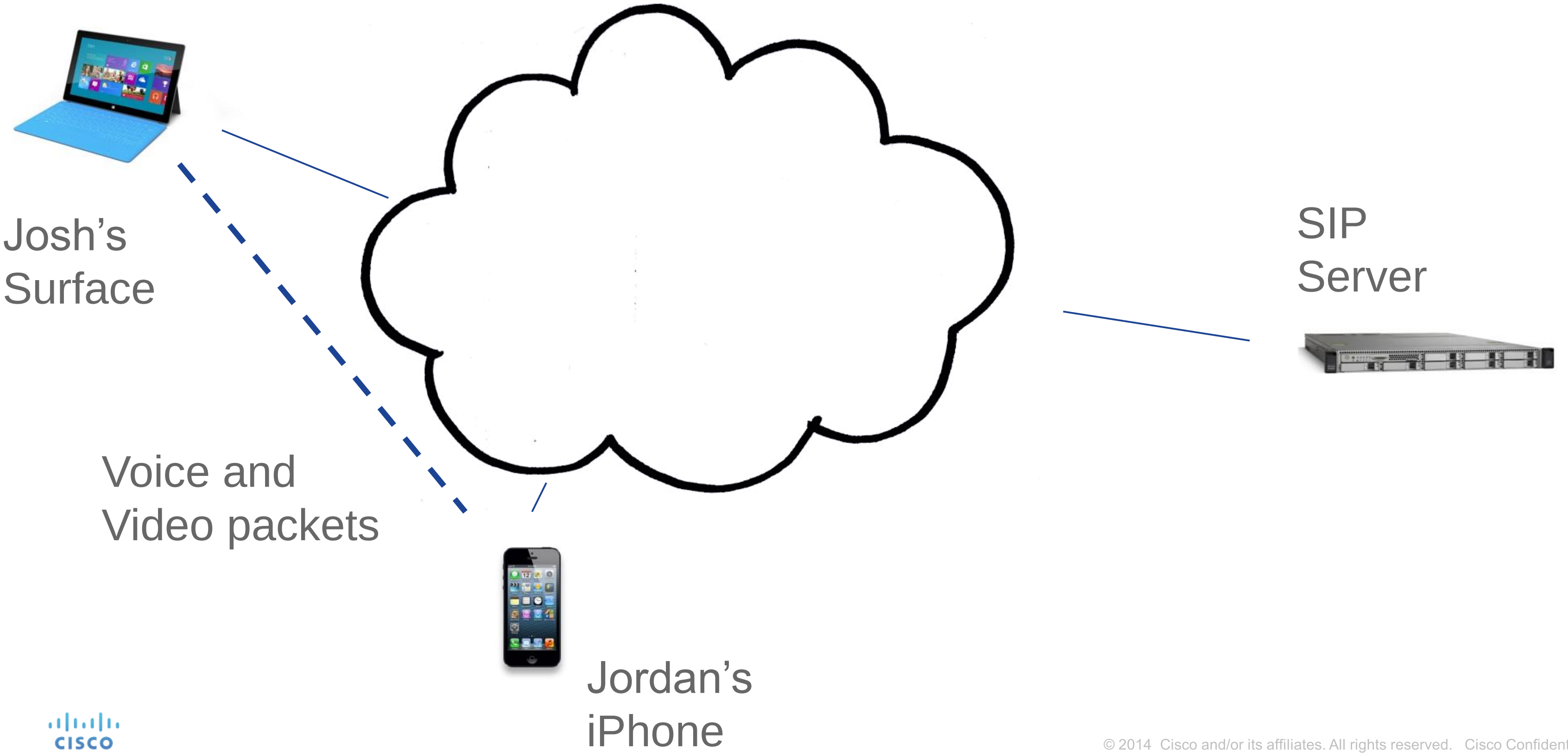
Jordan accepts the call



SIP Server tells Josh



Voice and Video Packets sent



Summary

- Voice over IP (VoIP) is using the Internet to make phone calls
- VoIP has many benefits including cheaper calls, better experiences, mobility, and network unification
- Low delay is a big requirement for VoIP
- Voice and video are sent using codecs, which perform lossy compression
- Setup of a call is done using signaling protocols – notably SIP