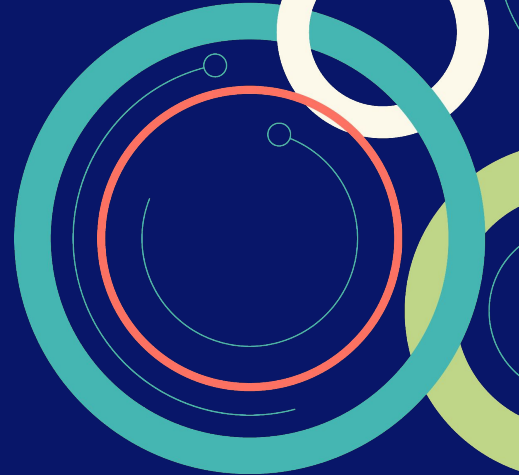




When the cloud fails: Debugging the "undocumented"

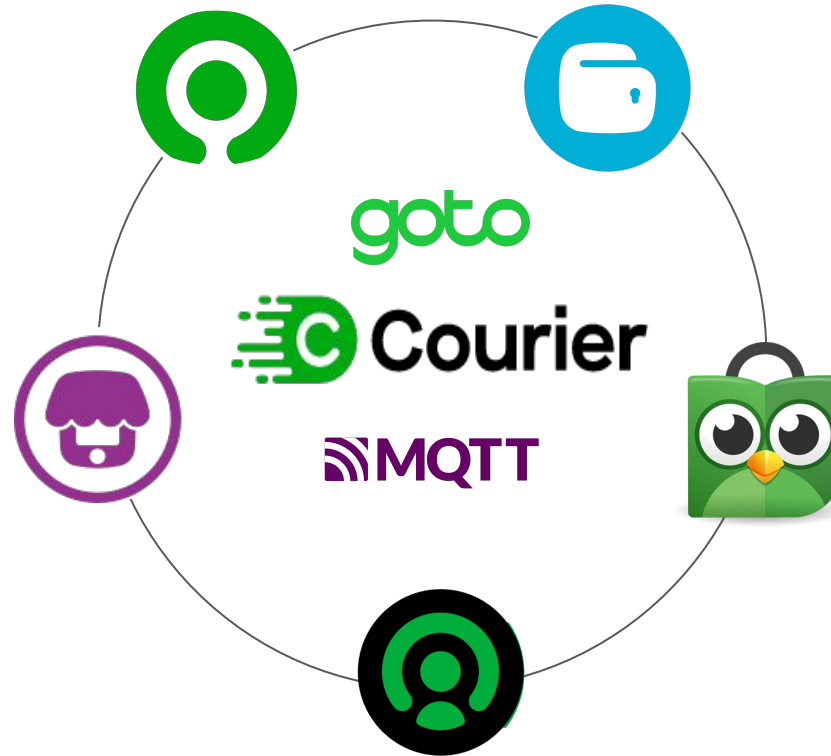
Dhruv Jain - GoTo Group (Gojek), Indonesia



MQTT Infrastructure Operational Scale

2
Million
Concurrent
Connections

3.2
Million
rpm
Throughput



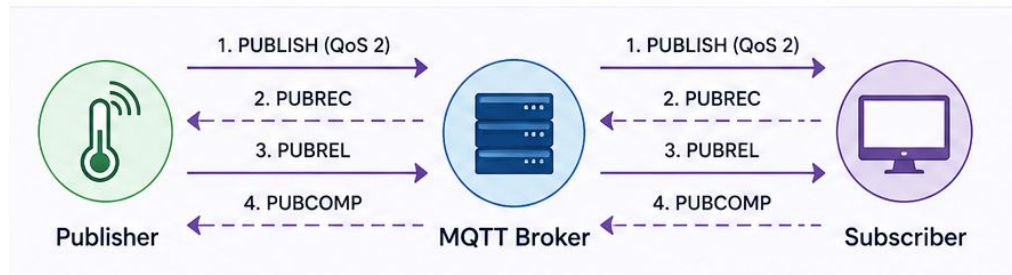
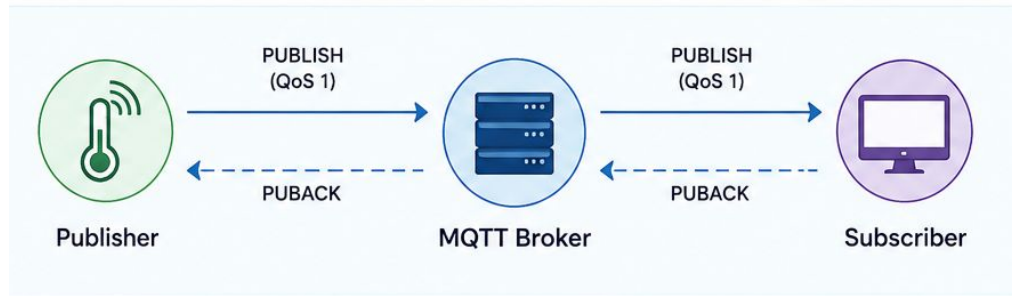
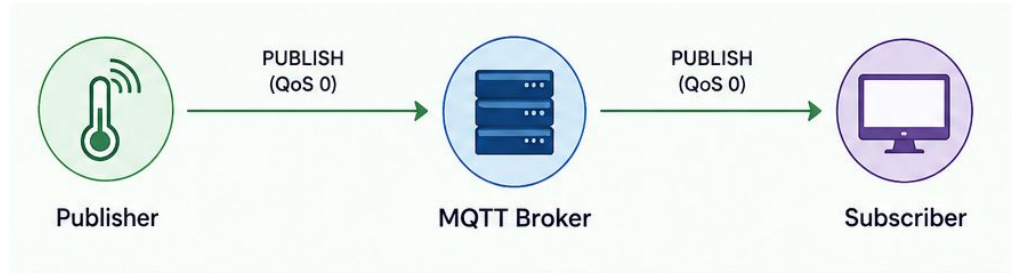
15
Million
Daily Active
Users

2.2 %
Indonesia's
GDP

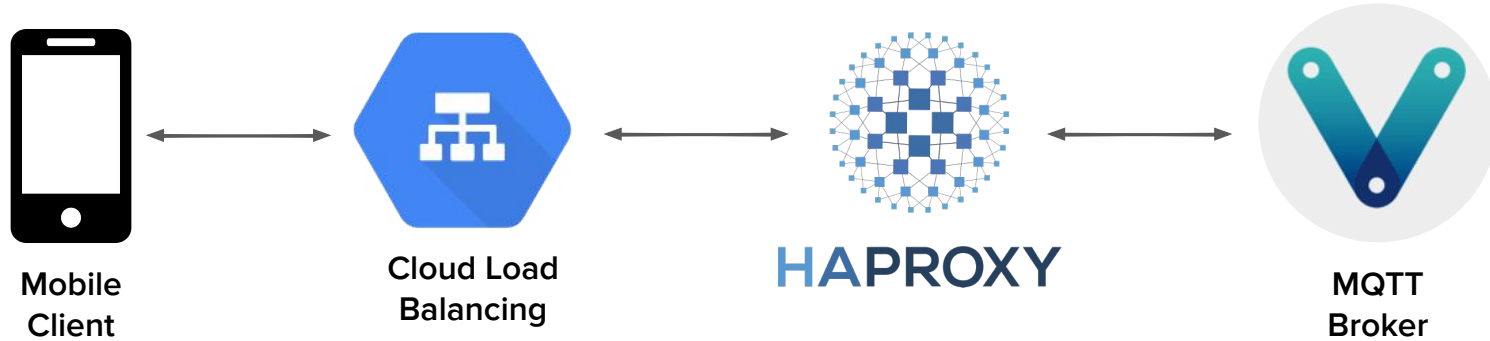


Background: MQTT Basics

1. Lightweight L7 protocol over tcp
2. 3 QoS levels
3. Bi-directional & full duplex long running connection
4. Pub-sub model with topic based routing



Background: MQTT Setup



Customers
150K concurrent
connections



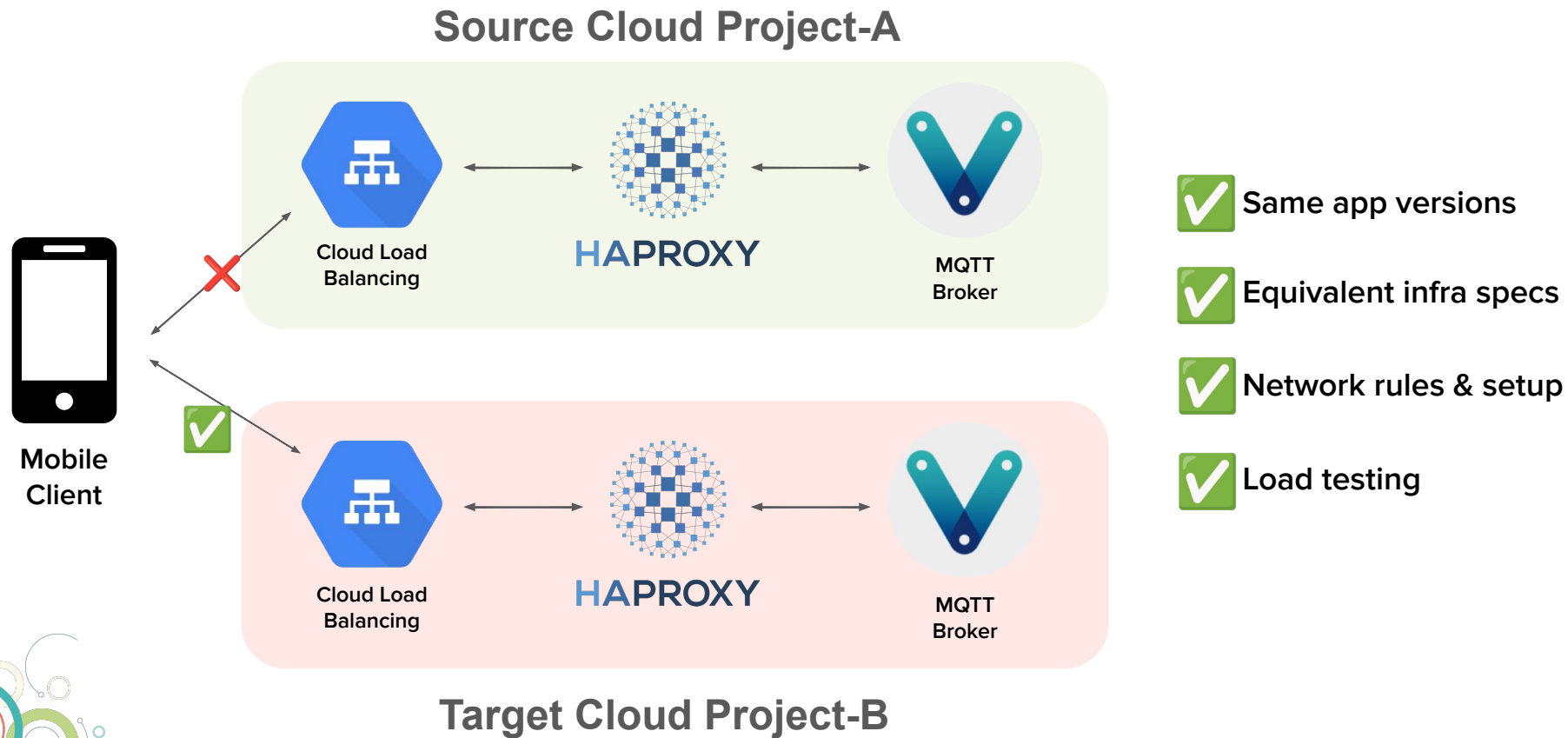
Driver Partners
600K concurrent
connections



Merchant Partners
450K concurrent
connections



The Trigger: MQTT Infrastructure Migration



SYSTEM STATUS OVERVIEW – “Everything Looked Green”

MQTT Broker Health

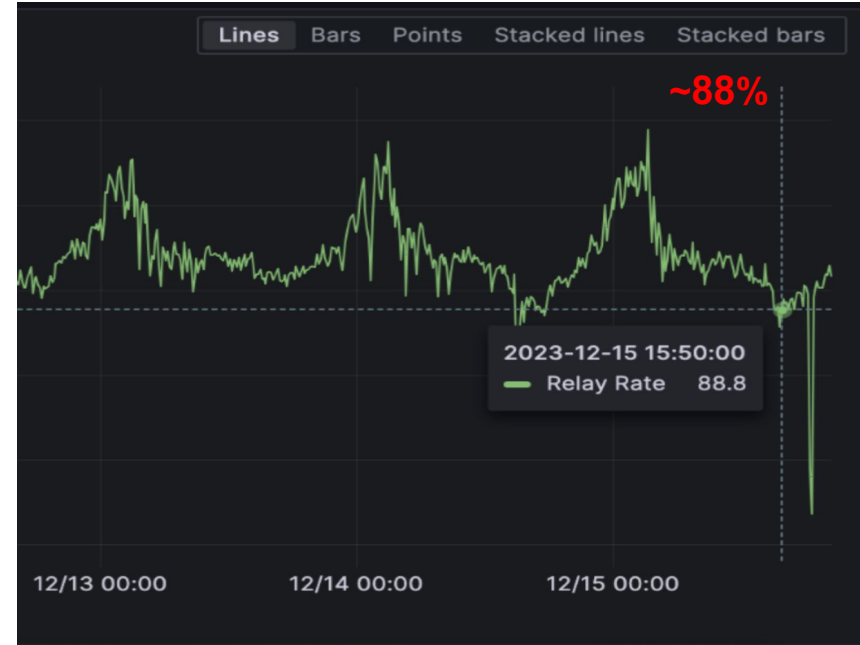
Connections:	✓ Stable
Publish Rate:	✓ Normal
Receive Rate:	✓ Normal
Broker CPU / Memory:	✓ Healthy
Same-Network Latency:	✓ Low latency
Bi-directional Channel:	✓ Working



The Problem: Drop in Delivery Rate for Merchants



Pre-cutover



Post-cutover



~5% Drop Observed in Push Path



The Problem: Drop in Delivery Rate for Drivers



Pre-cutover



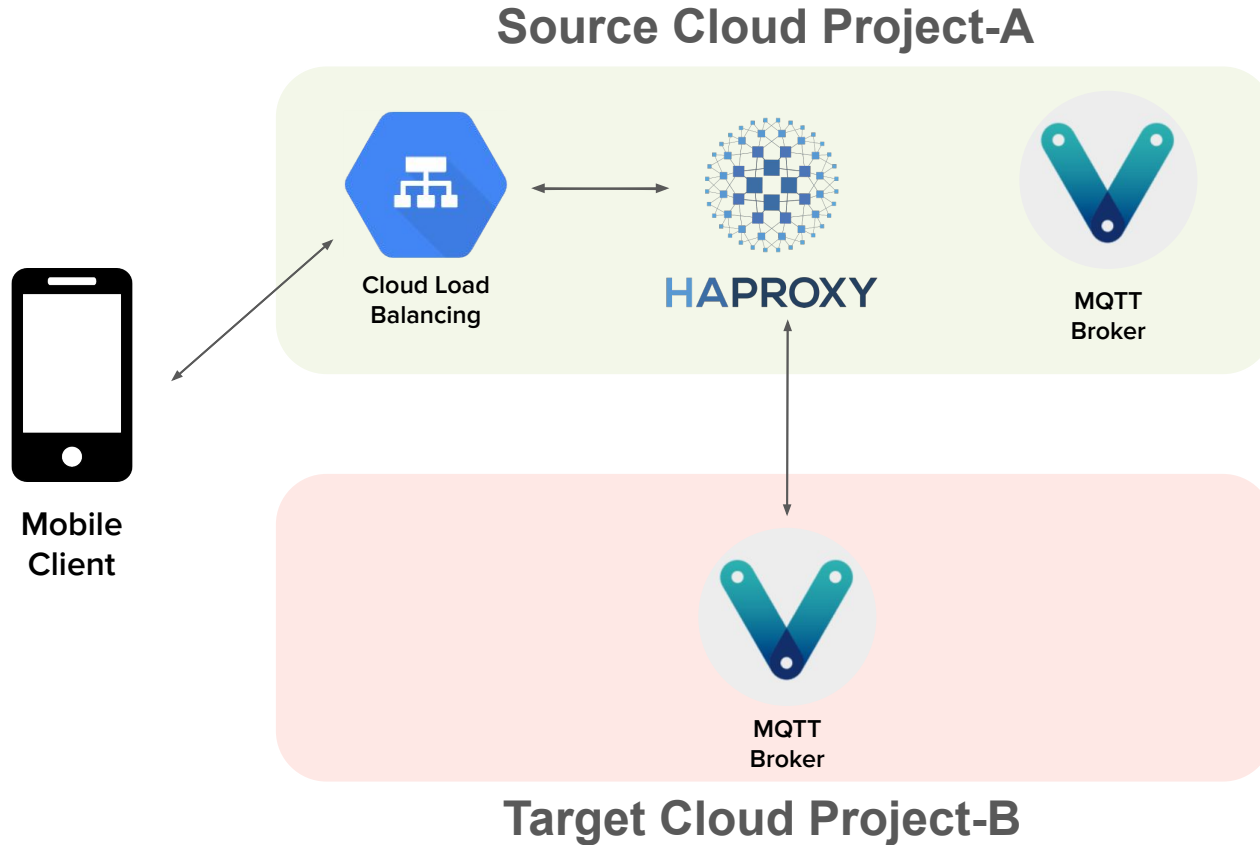
Post-cutover



~4% Drop Observed in Push Path

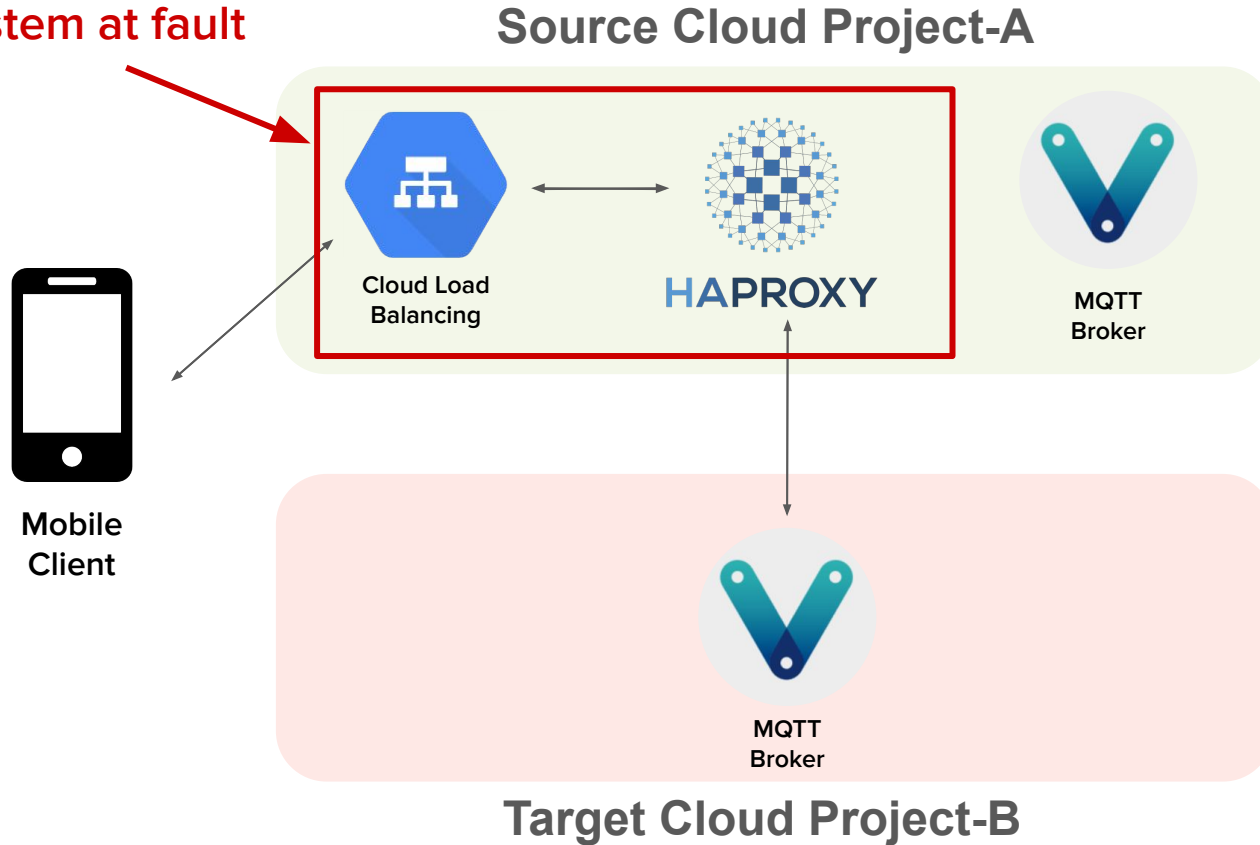


Investigation: Reduce Surface Area

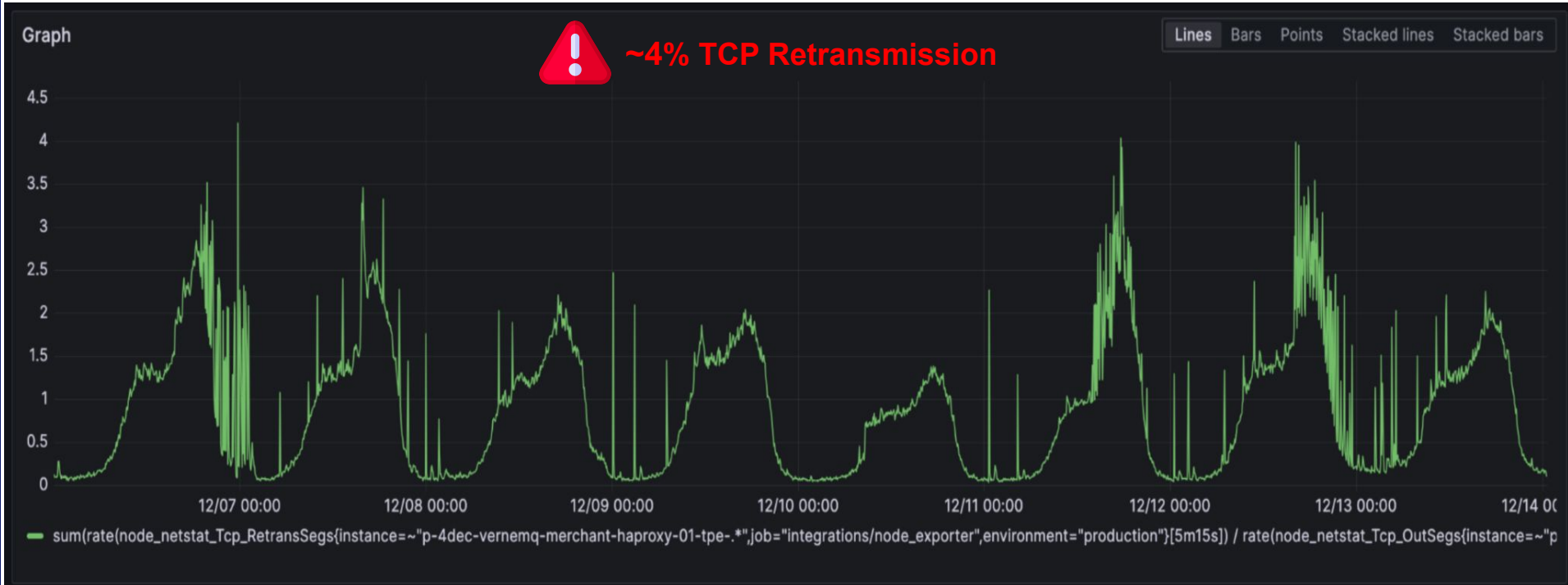


Investigation: Reduce Surface Area

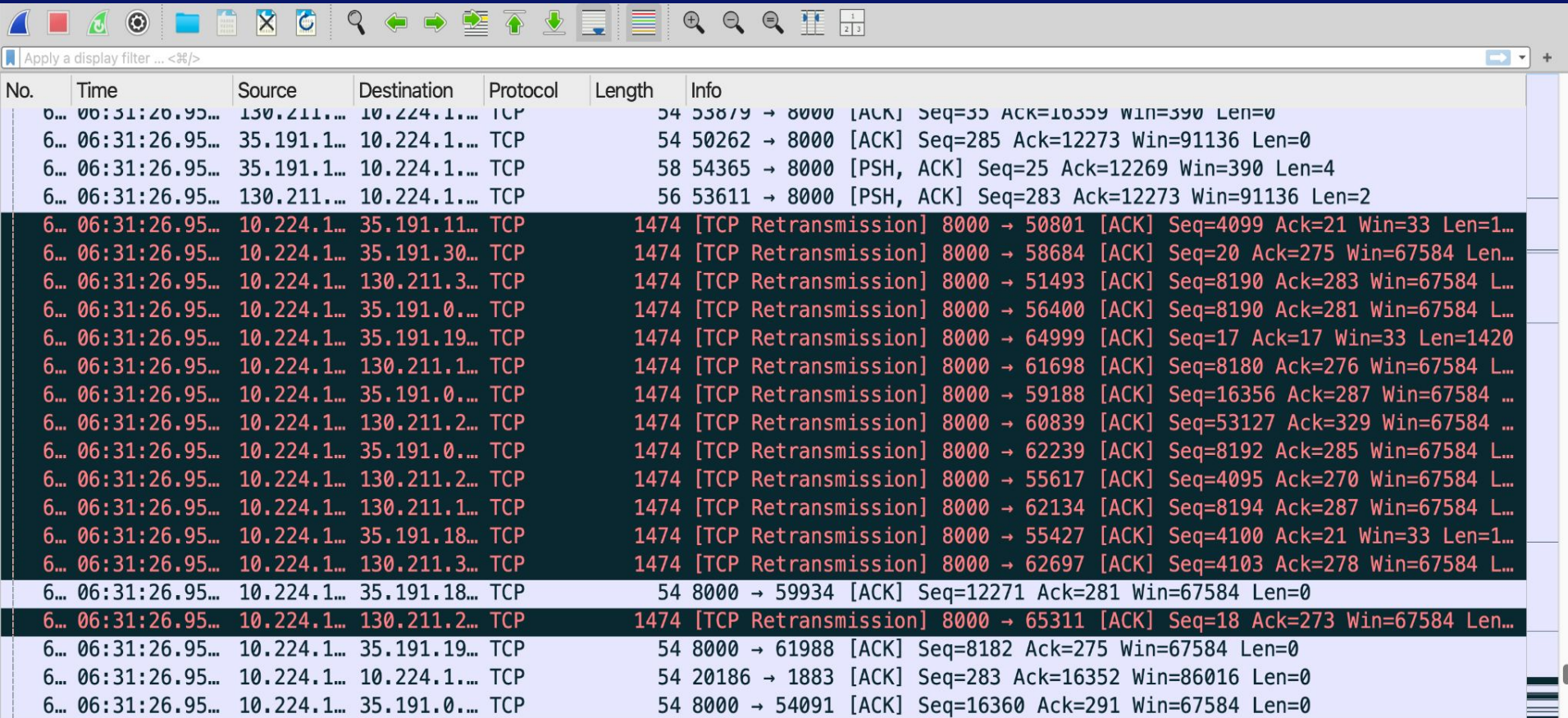
System at fault



Investigation: VM Network Metrics



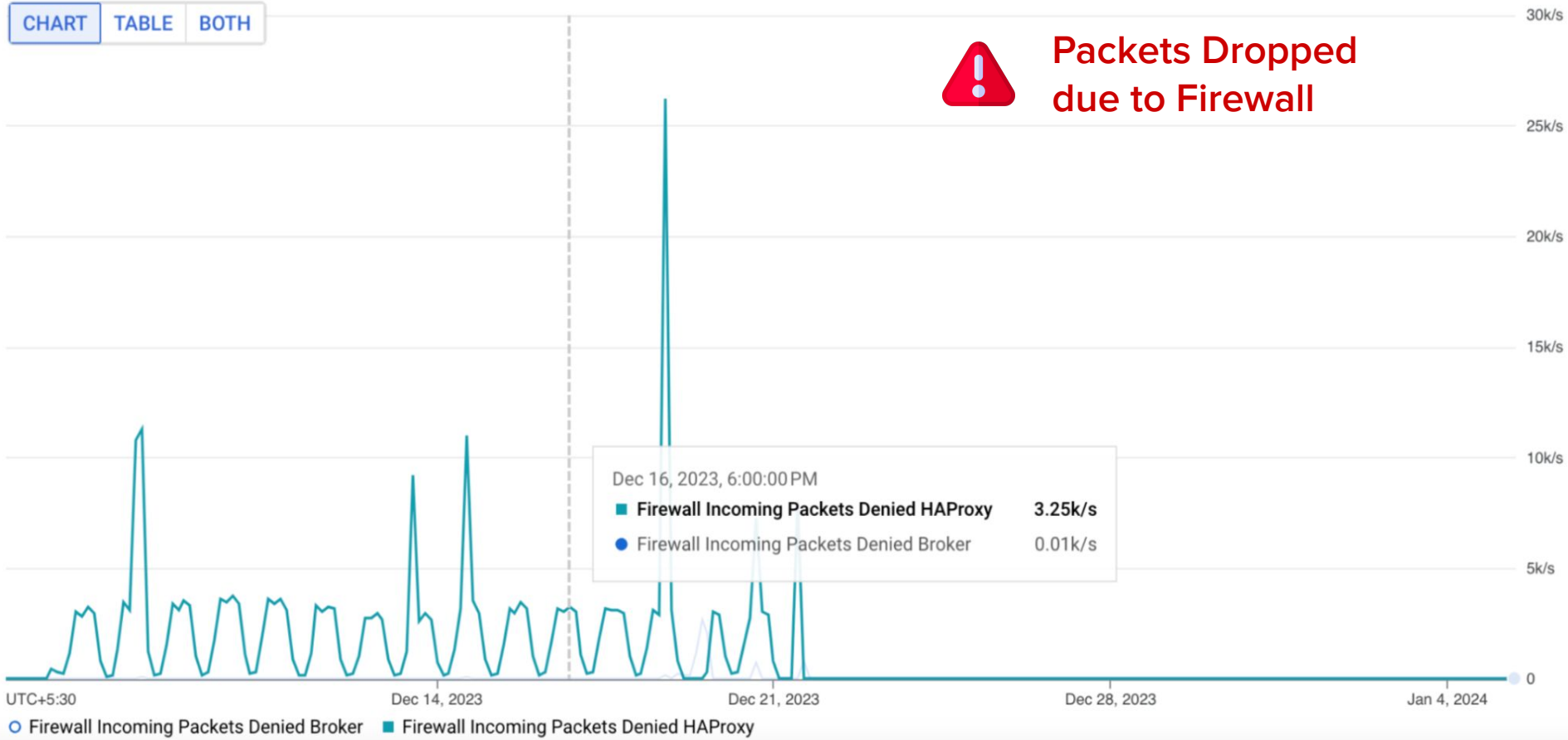
Investigation: Load Test & Wireshark Analysis



No.	Time	Source	Destination	Protocol	Length	Info
0...	00:31:20.95...	130.211.1...	10.224.1.1...	TCP	54	53879 → 8000 [ACK] Seq=35 Ack=16359 Win=390 Len=0
6...	06:31:26.95...	35.191.1.1...	10.224.1.1...	TCP	54	50262 → 8000 [ACK] Seq=285 Ack=12273 Win=91136 Len=0
6...	06:31:26.95...	35.191.1.1...	10.224.1.1...	TCP	58	54365 → 8000 [PSH, ACK] Seq=25 Ack=12269 Win=390 Len=4
6...	06:31:26.95...	130.211.1...	10.224.1.1...	TCP	56	53611 → 8000 [PSH, ACK] Seq=283 Ack=12273 Win=91136 Len=2
6...	06:31:26.95...	10.224.1.1...	35.191.11...	TCP	1474	[TCP Retransmission] 8000 → 50801 [ACK] Seq=4099 Ack=21 Win=33 Len=1...
6...	06:31:26.95...	10.224.1.1...	35.191.30...	TCP	1474	[TCP Retransmission] 8000 → 58684 [ACK] Seq=20 Ack=275 Win=67584 Len...
6...	06:31:26.95...	10.224.1.1...	130.211.3...	TCP	1474	[TCP Retransmission] 8000 → 51493 [ACK] Seq=8190 Ack=283 Win=67584 L...
6...	06:31:26.95...	10.224.1.1...	35.191.0.1...	TCP	1474	[TCP Retransmission] 8000 → 56400 [ACK] Seq=8190 Ack=281 Win=67584 L...
6...	06:31:26.95...	10.224.1.1...	35.191.19...	TCP	1474	[TCP Retransmission] 8000 → 64999 [ACK] Seq=17 Ack=17 Win=33 Len=1420
6...	06:31:26.95...	10.224.1.1...	130.211.1...	TCP	1474	[TCP Retransmission] 8000 → 61698 [ACK] Seq=8180 Ack=276 Win=67584 L...
6...	06:31:26.95...	10.224.1.1...	35.191.0.1...	TCP	1474	[TCP Retransmission] 8000 → 59188 [ACK] Seq=16356 Ack=287 Win=67584 ...
6...	06:31:26.95...	10.224.1.1...	130.211.2...	TCP	1474	[TCP Retransmission] 8000 → 60839 [ACK] Seq=53127 Ack=329 Win=67584 ...
6...	06:31:26.95...	10.224.1.1...	35.191.0.1...	TCP	1474	[TCP Retransmission] 8000 → 62239 [ACK] Seq=8192 Ack=285 Win=67584 L...
6...	06:31:26.95...	10.224.1.1...	130.211.2...	TCP	1474	[TCP Retransmission] 8000 → 55617 [ACK] Seq=4095 Ack=270 Win=67584 L...
6...	06:31:26.95...	10.224.1.1...	130.211.1...	TCP	1474	[TCP Retransmission] 8000 → 62134 [ACK] Seq=8194 Ack=287 Win=67584 L...
6...	06:31:26.95...	10.224.1.1...	35.191.18...	TCP	1474	[TCP Retransmission] 8000 → 55427 [ACK] Seq=4100 Ack=21 Win=33 Len=1...
6...	06:31:26.95...	10.224.1.1...	130.211.3...	TCP	1474	[TCP Retransmission] 8000 → 62697 [ACK] Seq=4103 Ack=278 Win=67584 L...
6...	06:31:26.95...	10.224.1.1...	35.191.18...	TCP	54	8000 → 59934 [ACK] Seq=12271 Ack=281 Win=67584 Len=0
6...	06:31:26.95...	10.224.1.1...	130.211.2...	TCP	1474	[TCP Retransmission] 8000 → 65311 [ACK] Seq=18 Ack=273 Win=67584 Len...
6...	06:31:26.95...	10.224.1.1...	35.191.19...	TCP	54	8000 → 61988 [ACK] Seq=8182 Ack=275 Win=67584 Len=0
6...	06:31:26.95...	10.224.1.1...	10.224.1.1...	TCP	54	20186 → 1883 [ACK] Seq=283 Ack=16352 Win=86016 Len=0
6...	06:31:26.95...	10.224.1.1...	35.191.0.1...	TCP	54	8000 → 54091 [ACK] Seq=16360 Ack=291 Win=67584 Len=0

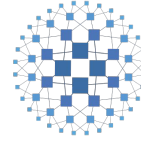
> Frame 5813745: Packet, 58 bytes on wire (464 bits), 58 bytes captured on interface...
Ethernet II, Src: 42:01:0a:e0:01:d2 (42:01:0a:e0:01:d2), Dst: 00:0c:2c:56:14:40 (00:0c:2c:56:14:40), Len: 54, Ethertype: 0800 (802.3), Protocol: 0100 (6), Length: 54, Encapsulated: 0100 (6), Raw: 0000 42 01 0a e0 00 01 42 01 0a e0 01 d2 08 00 45 00 B...B
0010 00 2c 56 14 40 00 40 06 54 22 0a e0 01 d2 82 d3 .V.G.

Investigation: Cloud Provider VM Metrics



Investigation Results & Next Steps

1. No impact on MQTT setup for customer user type
2. Impact during peak hours & push path only for merchant & driver partners
3. High tcp restrans on HAProxy resulting in increased latency in push path
4. Despite firewall correctly configured on HAProxy VM, incoming packets getting randomly dropped with the reason as "denied"

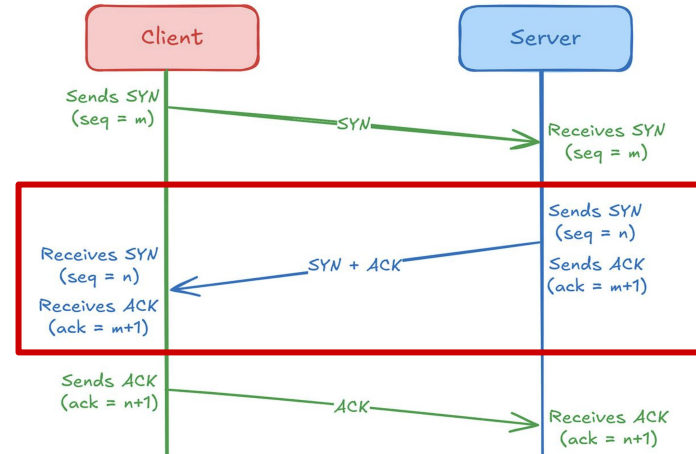


HAProxy

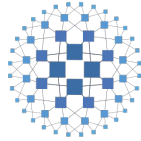
Allow outbound tcp traffic to VerneMQ's port 1883



Allow inbound tcp traffic from HAProxy



Workaround: Additionally Whitelist Reverse Rules



HAPROXY

Allow outbound tcp
traffic to VerneMQ's
port 1883



Allow inbound tcp
traffic from HAProxy

Allow inbound tcp
traffic from
VerneMQ's port 1883

Allow outbound tcp
traffic to HAProxy

New Rules



Workaround: Result



Recovered





Root Cause



HAProxy VM conntrack table hits the limit.

- Updated documentation
- Exported connection tracking table metrics graph



Lessons Learned

1. Observability is the key
 - a. Export cloud provider's metrics
 - b. Always plot tcp retrans in %
2. Load test using exact same setup as production
3. Tools like TCPdump & wireshark for packet level analysis are quite effective.



Questions?



gojek.github.io/courier

Dhruv Jain - GoTo Group
(Gojek), Indonesia

www.linkedin.com/in/dhruvjain99

If you plan to use speaker notes, please follow the steps below to ensure your laptop is set up correctly. This will allow you to view your notes while the audience sees only your slides.

Before Connecting

- Close all unnecessary apps & tabs
- Turn off notifications
- Disable Night Mode / TrueTone
- Don't enter full-screen presentation mode yet

At the podium

- Plug in the HDMI cable
- If needed, use your USB-C to HDMI adapter
- On Mac, click "Allow" if prompted

USE "EXTENDED DISPLAY" MODE (NOT MIRRORED)

Using MAC

- Go to System Settings > Displays
- Click the "Arrangement" tab
- Make sure "Mirror Displays" is unchecked

Using PC

- Press \square Win + P, then select "Extend"
- Or: Right-click on the Desktop > Display Settings
- Under Multiple Displays, select "Extend these displays"



When Using PowerPoint

- Start slideshow: ⌘Cmd + Enter (Mac) or F5 (PC)
- Or: Click Slide Show > Play from Start
- If notes show on the wrong screen, click “Swap Displays”
- <https://www.youtube.com/watch?v=gQ3D4m-5pww>

Google Slides

- Use Google Chrome as your browser
- Click the dropdown next to ‘Slideshow’ > select Presentation Display Options
- Check “Presenter View” and “Full Screen”
- If prompted, select "Allow"
- Choose the external display (may have a strange name like PanasonicXYZ)
- Click Start Slideshow
- <https://www.youtube.com/watch?v=-GT7WCvPcys>

