Efficient and Practical RDF Streaming with Protocol Buffers

Extra Materials

Piotr Sowiński

June 2022

1 Raw serialization / deserialization throughput

Table 1: Raw Serialization Throughput (kT/s)

	jelly_full	jelly_noprefix	jelly_noprefix_sm	jelly_norepeat	jena-proto	n3	rdf-xml	turtle
aemet-1	631.70	865.96	915.20	472.45	1139.28	2447.45	155.20	153.37
aemet-2	529.71	749.02	791.03	402.54	1208.88	3222.04	77.89	140.62
flickr_10m	515.78	622.08	663.29	390.21	960.98	1830.92	73.24	212.83
identica	584.37	738.04	724.74	504.93	1039.36	998.10	65.34	156.38
migr_reschange	556.70	778.51	815.99	481.80	807.79	4270.32	63.32	138.79
mix	581.46	787.31	754.71	444.18	938.34	1692.79	72.68	150.33
nevada_10m	466.06	689.56	648.60	380.21	1098.52	3246.86	15.66	129.38
petrol	562.16	717.86	702.57	453.00	1058.82	3185.75	71.23	151.74
tour_cap_nuts3	572.14	797.74	826.71	470.52	1096.37	3956.50	62.96	136.85
wikipedia	624.38	783.09	813.74	549.47	1252.36	1792.14	61.95	188.74

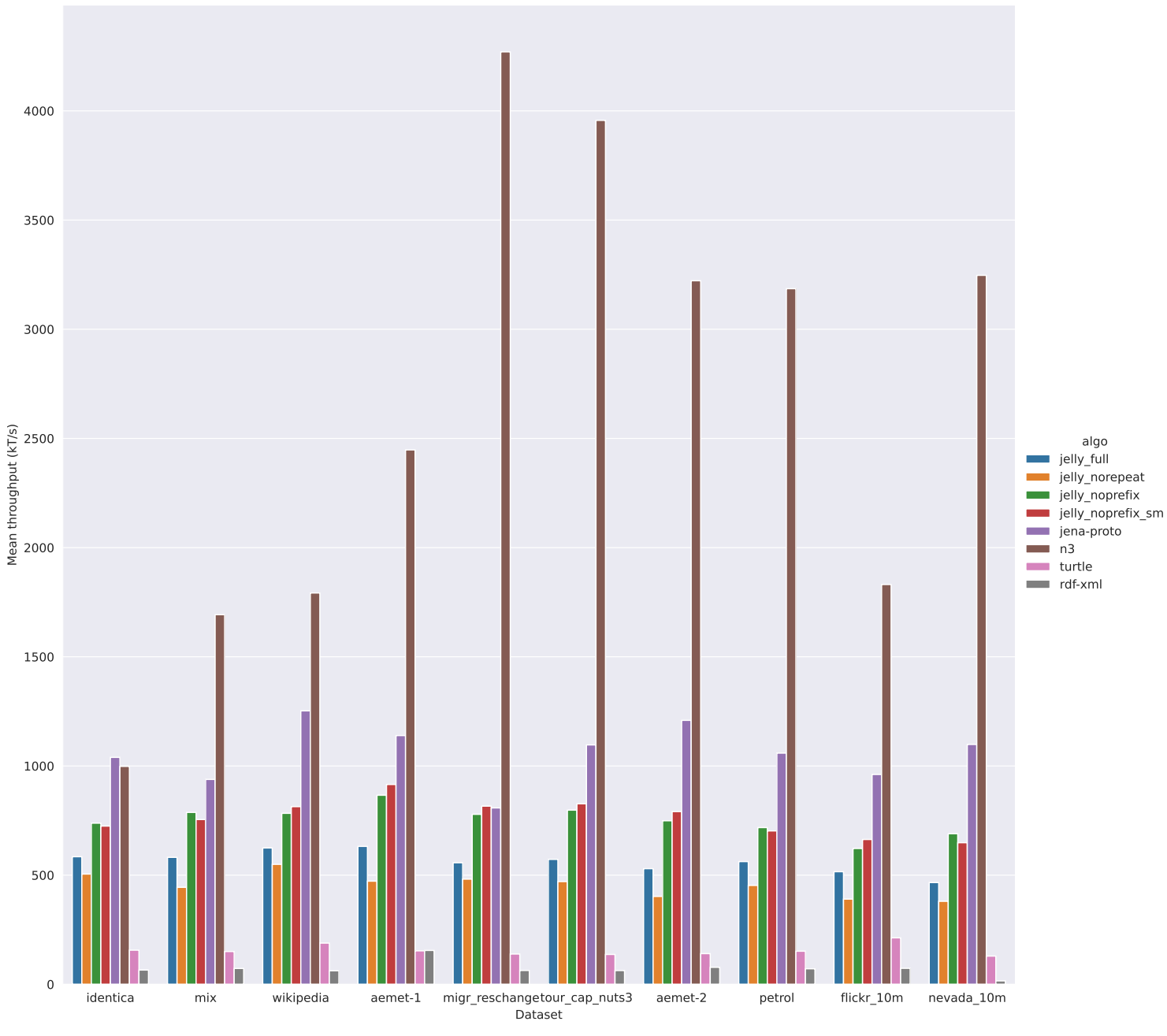

Figure 1: Raw serialization throughput

Table 2: Raw Deserialization Throughput (kT/s)

	jelly_full	jelly_noprefix	jelly_noprefix_sm	jelly_norepeat	jena-proto	n3	rdf-xml	turtle
aemet-1	923.69	989.02	1010.35	795.75	524.55	206.70	123.42	264.39
aemet-2	867.30	963.57	937.07	738.58	515.97	161.43	80.91	174.45
flickr_10m	760.34	850.08	821.77	631.22	488.89	136.91	83.61	179.76
identica	1026.00	1102.15	1151.08	884.71	449.01	145.05	86.06	158.43
migr_reschange	842.02	973.73	980.12	708.11	529.97	164.96	71.44	180.63
mix	728.61	774.07	763.33	684.66	369.81	124.08	73.48	113.20
nevada_10m	772.15	872.97	793.09	659.01	519.47	116.92	66.61	119.80
petrol	214.34	213.49	215.96	203.32	213.44	86.57	38.39	46.49
tour_cap_nuts3	852.47	966.92	977.65	737.74	527.12	165.27	69.94	183.91
wikipedia	1216.56	1193.94	1177.12	982.96	512.13	172.62	98.44	160.84

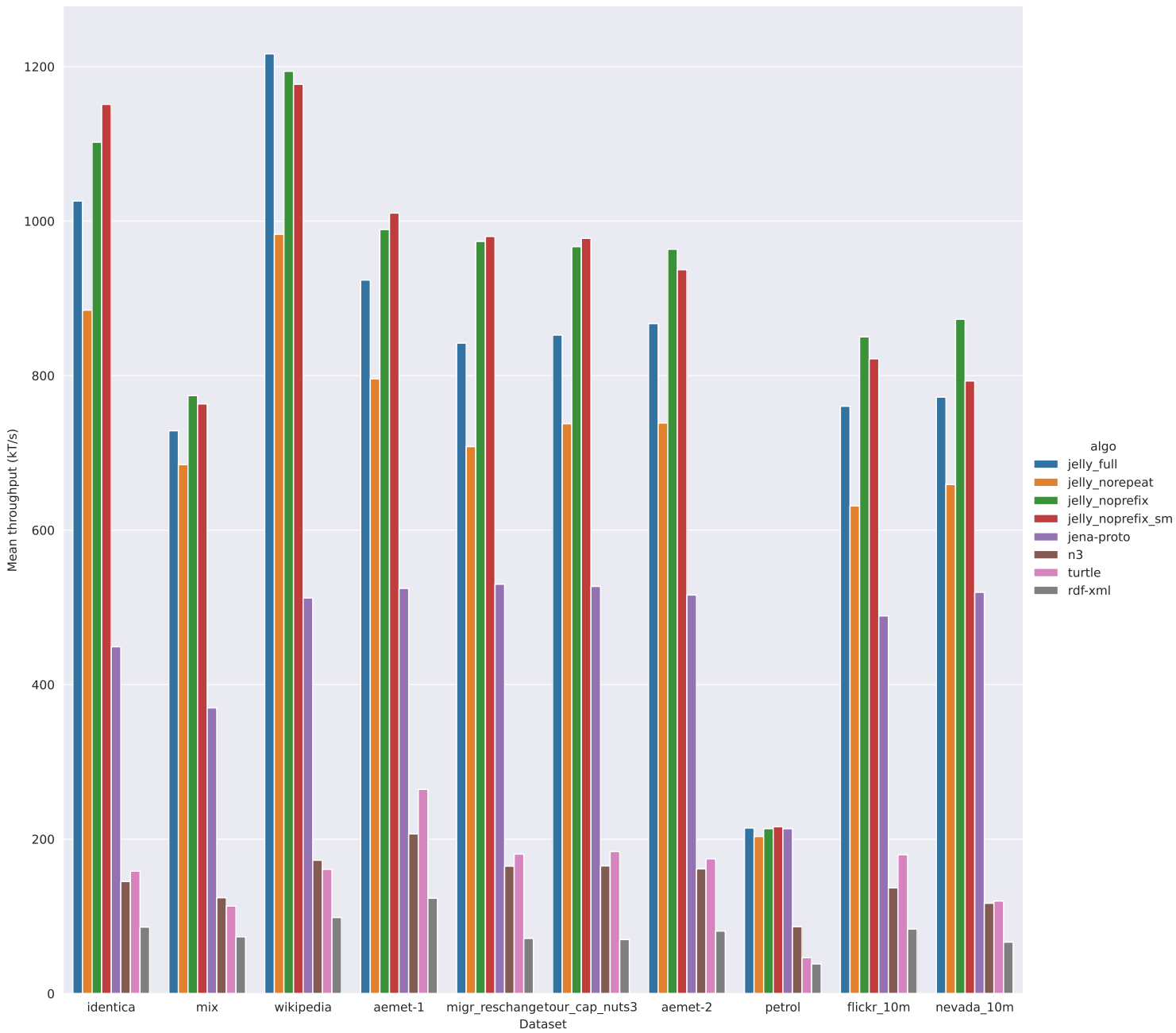

Figure 2: Raw deserialization throughput

2 Serialized size

Table 3: Compression Ratio (%)

dataset **method**	aemet-1	aemet-2	flickr_10m	identica	migr_reschange	mix	nevada_10m	petrol	tour_cap_nuts3	wikipedia
jelly-full	19.40	22.07	31.56	47.28	15.76	30.51	19.17	25.50	15.90	49.56
jelly-full-gzip	1.80	1.59	10.70	13.75	0.75	6.95	2.27	4.51	0.93	10.04
jelly-noprefix	18.95	22.99	32.84	49.67	17.20	33.46	20.44	31.11	17.49	55.87
jelly-noprefix-gzip	1.77	1.58	10.41	13.62	0.78	7.14	2.33	4.70	0.94	10.25
jelly-noprefix-sm	20.56	23.37	33.44	50.28	17.10	37.02	28.66	35.01	17.38	55.54
jelly-noprefix-sm-gzip	2.02	1.75	10.39	13.61	0.80	7.82	2.82	5.28	0.97	10.27
jelly-norepeat	24.31	24.37	35.89	56.36	18.08	34.48	21.27	28.38	18.21	53.04
jelly-norepeat-gzip	3.83	2.93	12.71	15.33	2.04	8.84	3.35	6.80	2.15	12.72
jena-proto	110.15	106.85	108.29	109.91	106.29	109.51	105.70	107.93	106.29	112.64
jena-proto-gzip	5.56	3.84	12.88	18.34	2.60	11.38	4.16	7.04	3.08	13.61
n3	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
n3-gzip	4.39	3.35	10.83	16.15	2.59	9.10	3.19	7.00	2.76	11.47
rdf-xml	28.04	49.66	51.63	60.82	46.12	50.23	50.37	54.53	46.47	76.26
rdf-xml-gzip	2.85	1.28	7.95	8.67	0.75	6.11	1.46	4.12	0.93	8.85
turtle	65.94	68.00	46.32	88.88	69.09	78.51	71.97	78.35	69.41	98.26
turtle-gzip	2.21	1.53	7.07	9.50	0.97	6.42	1.99	4.32	1.02	9.29

3 End-to-end streaming throughput

Table 4: End-to-End Streaming Throughput – unlimited network (kT/s)

dataset algo	aemet-1	aemet-2	flickr_10m	identica	migr_reschange	mix	nevada_10m	petrol	tour_cap_nuts3	wikipedia
grpc_jelly_full	530.11	457.05	397.54	274.00	519.96	358.69	345.66	195.63	496.54	410.32
grpc_jelly_full_gzip	349.10	328.76	221.47	165.40	384.22	223.31	233.49	190.21	372.95	232.21
grpc_jelly_noprefix	677.50	582.76	465.09	318.56	681.20	372.71	412.08	194.49	655.22	465.50
grpc_jelly_noprefix_gzip	426.51	393.35	242.28	182.94	488.42	249.45	255.27	182.79	472.64	255.39
kafka_jelly_full	631.16	577.98	500.87	296.06	608.73	403.76	466.36	179.08	620.93	522.39
kafka_jelly_full_gzip	414.44	383.44	242.86	193.44	460.90	251.94	328.39	176.44	487.53	282.39
kafka_jelly_noprefix	863.66	794.17	515.63	331.84	888.80	396.76	538.59	185.60	846.49	614.02
kafka_jelly_noprefix_gzip	482.44	492.57	263.99	212.03	658.47	288.85	418.40	176.11	628.85	291.96
kafka_jena-proto	367.45	323.43	249.80	201.30	327.36	248.39	245.41	162.20	334.59	363.34
kafka_jena-proto_gzip	325.01	292.01	227.13	157.24	310.64	217.01	257.13	162.88	311.35	310.49
kafka_n3	187.31	136.79	110.66	105.57	138.95	105.12	96.23	76.12	142.56	152.23
kafka_n3_gzip	190.49	140.47	112.38	100.36	143.26	106.82	99.71	74.94	144.49	152.00

Table 5: End-to-End Streaming Throughput – 100 Mbit/s network (kT/s)

dataset algo	aemet-1	aemet-2	flickr_10m	identica	migr_reschange	mix	nevada_10m	petrol	tour_cap_nuts3	wikipedia
grpc_jelly_full	442.97	297.12	250.64	185.78	412.56	261.88	225.15	192.91	408.85	251.25
grpc_jelly_full_gzip	361.06	341.22	220.64	154.17	413.18	219.04	238.76	186.93	373.24	236.31
grpc_jelly_noprefix	447.50	277.82	239.25	176.51	369.07	241.93	174.05	193.10	361.52	216.87
grpc_jelly_noprefix_gzip	427.05	387.02	244.54	166.95	481.62	242.62	251.49	193.14	464.36	255.67
kafka_jelly_full	109.64	49.19	41.62	35.85	75.63	47.04	49.06	52.55	64.45	41.70
kafka_jelly_full_gzip	396.37	381.40	218.88	110.71	451.97	173.79	337.22	174.17	468.40	214.02
kafka_jelly_noprefix	109.28	46.57	41.42	34.70	63.08	41.37	46.61	41.79	62.06	40.57
kafka_jelly_noprefix_gzip	475.85	462.09	220.54	110.97	581.76	174.49	429.42	178.61	578.06	214.05
kafka_jena-proto	20.39	15.72	20.02	20.22	15.23	19.95	14.06	20.01	15.21	26.55
kafka_jena-proto_gzip	319.06	286.35	176.36	98.73	312.19	167.25	235.34	160.78	306.36	210.40
kafka_n3	20.87	17.35	20.43	22.59	16.67	20.71	14.77	20.41	16.64	31.20
kafka_n3_gzip	184.60	140.65	111.19	66.54	141.18	91.44	98.75	76.39	145.14	147.64

Table 6: End-to-End Streaming Throughput – 50 Mbit/s network (kT/s)

dataset algo	aemet-1	aemet-2	flickr_10m	identica	migr_reschange	mix	nevada_10m	petrol	tour_cap_nuts3	wikipedia
grpc_jelly_full	219.83	140.84	117.13	100.59	195.58	137.32	109.41	157.03	200.23	123.28
grpc_jelly_full_gzip	344.39	327.89	215.87	149.20	412.59	222.47	233.20	184.99	361.15	230.46
grpc_jelly_noprefix	220.04	140.96	114.54	94.30	184.73	123.06	87.10	127.28	182.22	110.13
grpc_jelly_noprefix_gzip	427.81	404.71	243.67	160.02	480.51	242.46	252.93	190.83	458.16	256.75
kafka_jelly_full	70.16	31.16	26.36	23.07	46.93	30.02	31.14	33.28	40.52	26.47
kafka_jelly_full_gzip	391.51	376.30	139.56	75.99	468.64	118.84	339.96	189.76	476.46	143.23
kafka_jelly_noprefix	68.17	29.52	26.20	22.31	39.91	26.44	29.37	26.51	39.31	25.63
kafka_jelly_noprefix_gzip	419.63	450.36	140.71	76.08	589.01	118.64	339.26	179.30	592.30	143.30
kafka_jena-proto	12.86	9.96	12.56	12.93	9.67	12.62	8.94	12.55	9.65	16.72
kafka_jena-proto_gzip	226.87	286.42	109.74	68.55	311.12	114.95	157.18	152.23	297.68	141.32
kafka_n3	13.24	10.93	12.88	14.35	10.53	13.17	9.39	12.86	10.51	19.45
kafka_n3_gzip	187.10	144.84	124.03	61.10	149.51	88.14	107.87	79.75	151.17	139.19

4 End-to-end streaming latency

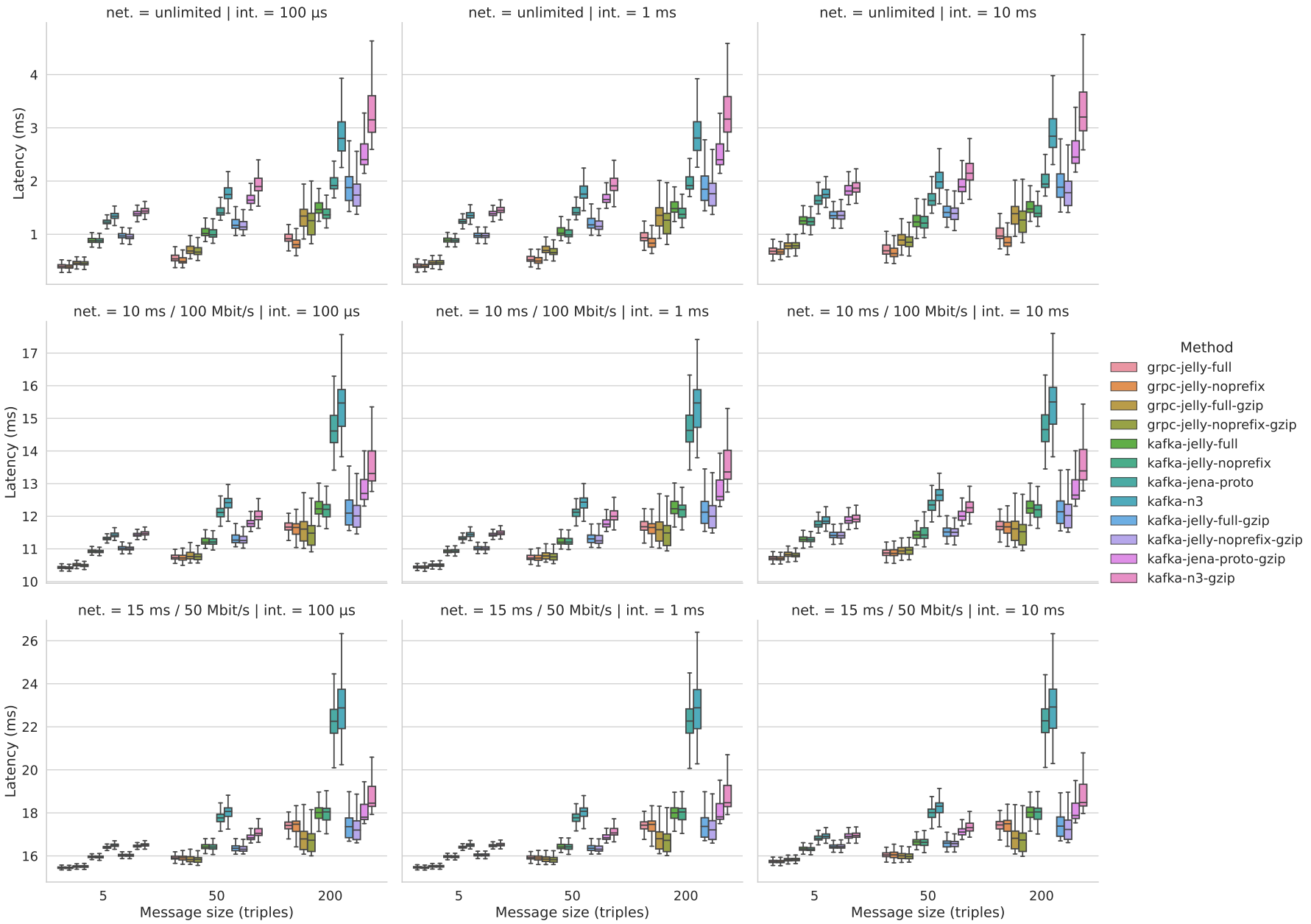

Figure 3: Streaming latency (aggregated over all five datasets)