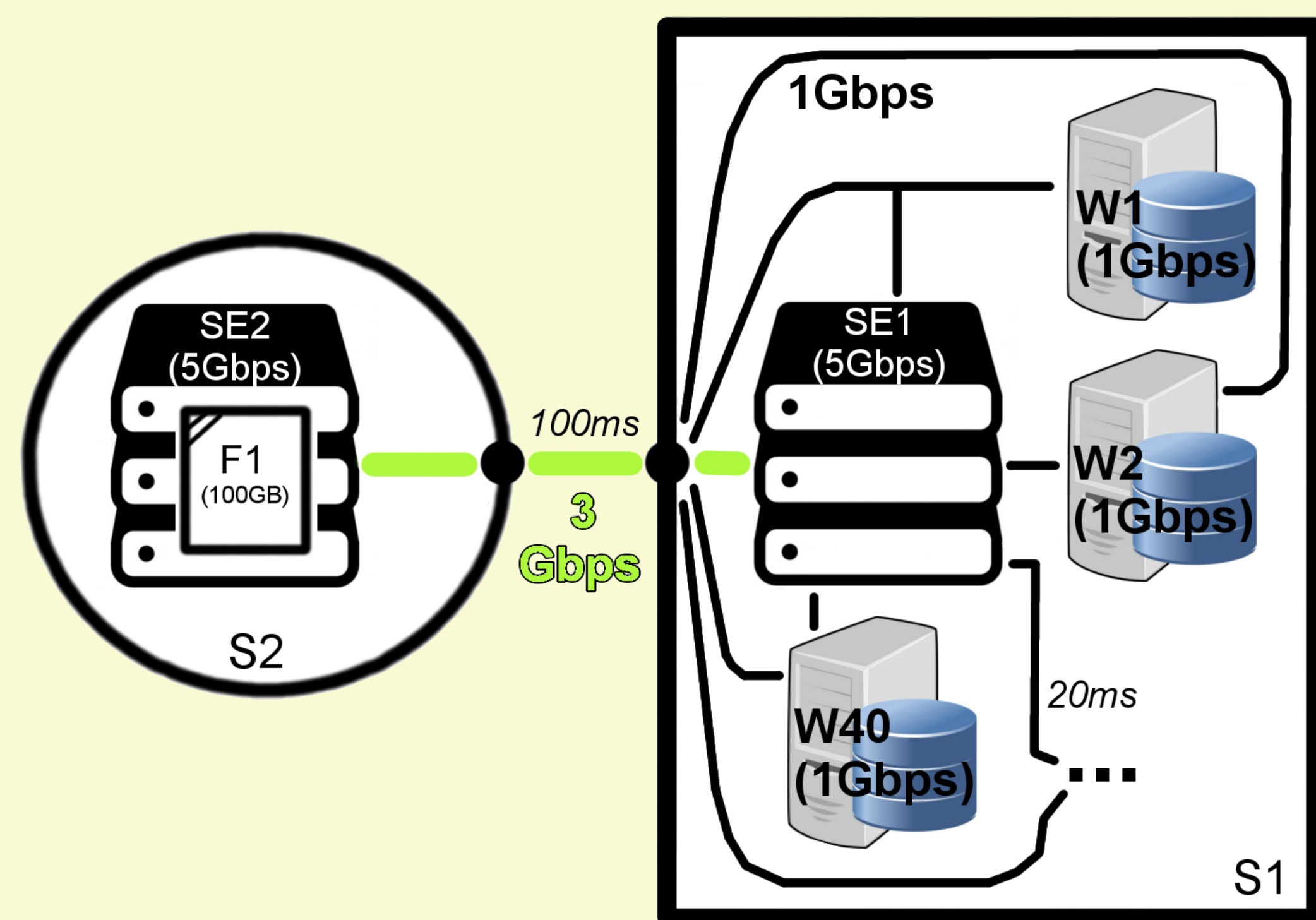


Remote vs Local Data Access: Goodput Break-Even Point



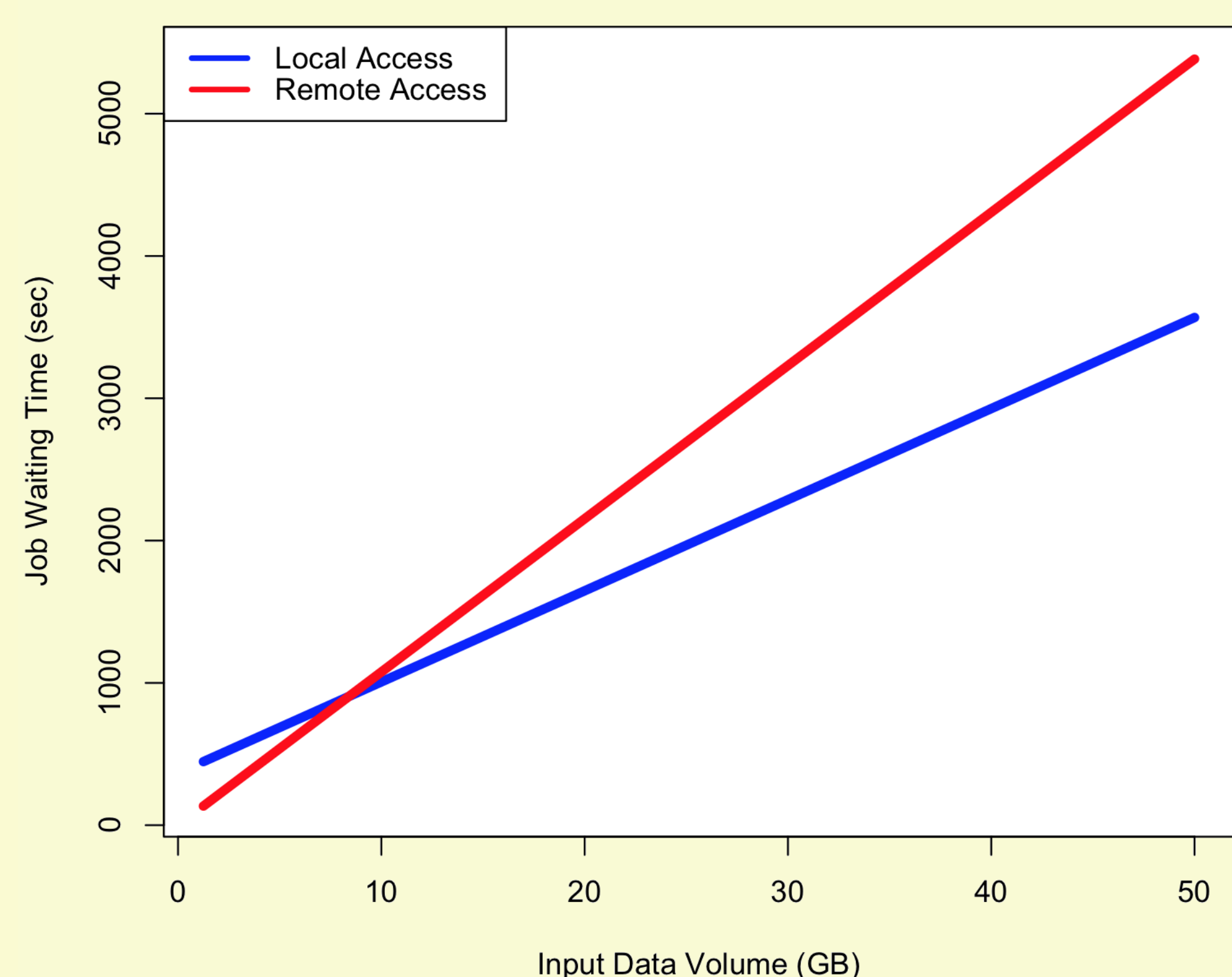
Example I

Each worker node $w_1, w_2, \dots, w_{40}$ is assigned a single computational job $j_1, j_2, \dots, j_{40}$ respectively. Each job $j_1, j_2, \dots, j_{40}$ requires a distinct 1-50GB fraction of the file F_1 , which is located at the remote storage element SE_2 as input.

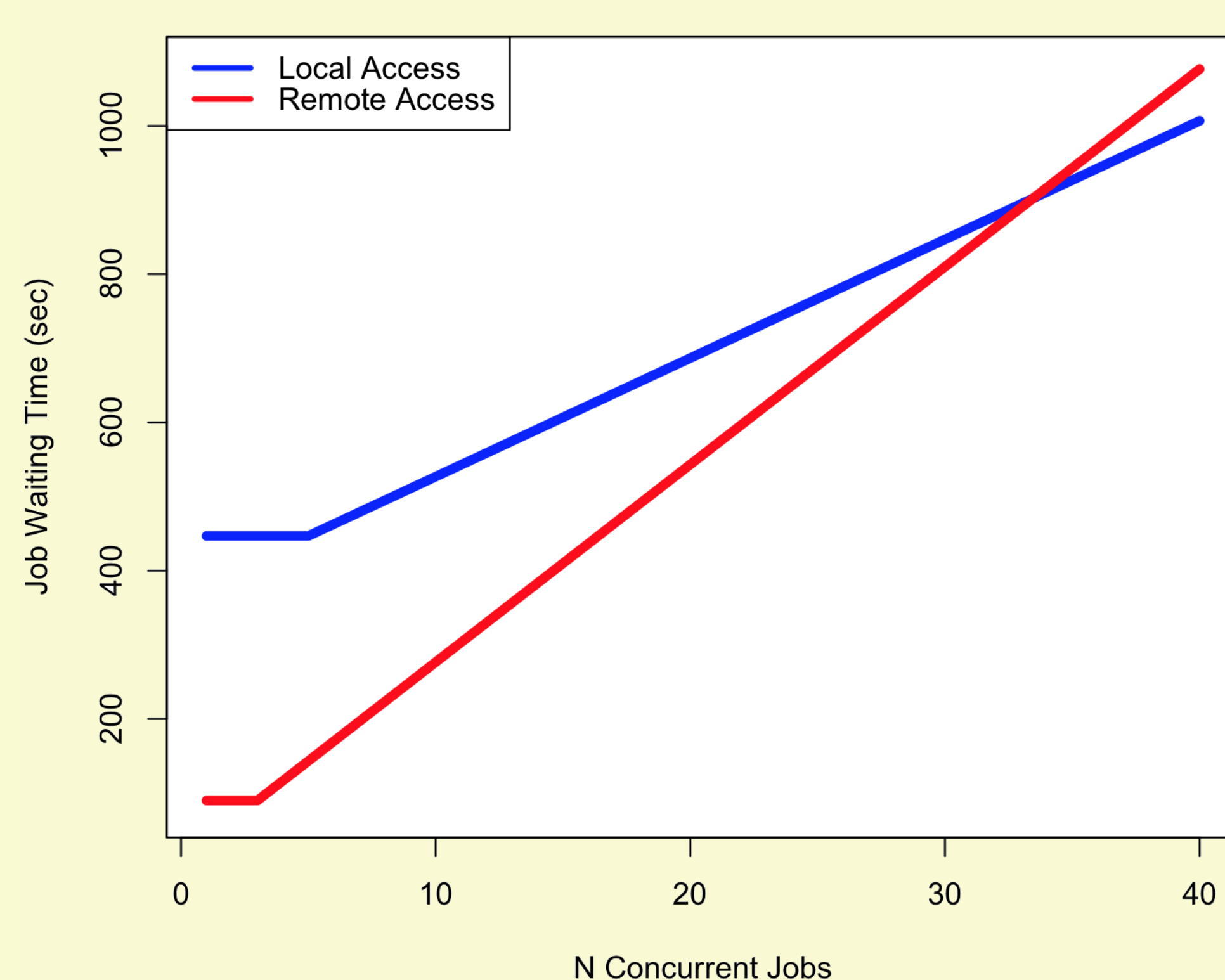
Example II

1-40 worker nodes are assigned a single computational job. Each job needs a distinct 10GB fraction of the file F_1 as input.

Example I



Example II



- Data caching by means of data-placement for minimization of limiting effects posed by large amounts of parallel data flows on throughput of network links or storage elements
- Remote data access for avoidance of increased cumulative latency due to transfer of redundant file fractions

Benefits

In contrast to data-placement, remote data access:

- Transfers data asynchronously with respect to the job execution
- Maximizes utilization of otherwise idle resources
- Bypasses limited disk space and quotas imposed by virtual organizations
- Streams only relevant fractions of input data
- Has smaller coordination overhead
- Enables novel data access policies

Throughput Bottlenecks

Determined by capacities of:

- Worker node NIC
- Storage element NIC
- WAN and LAN links

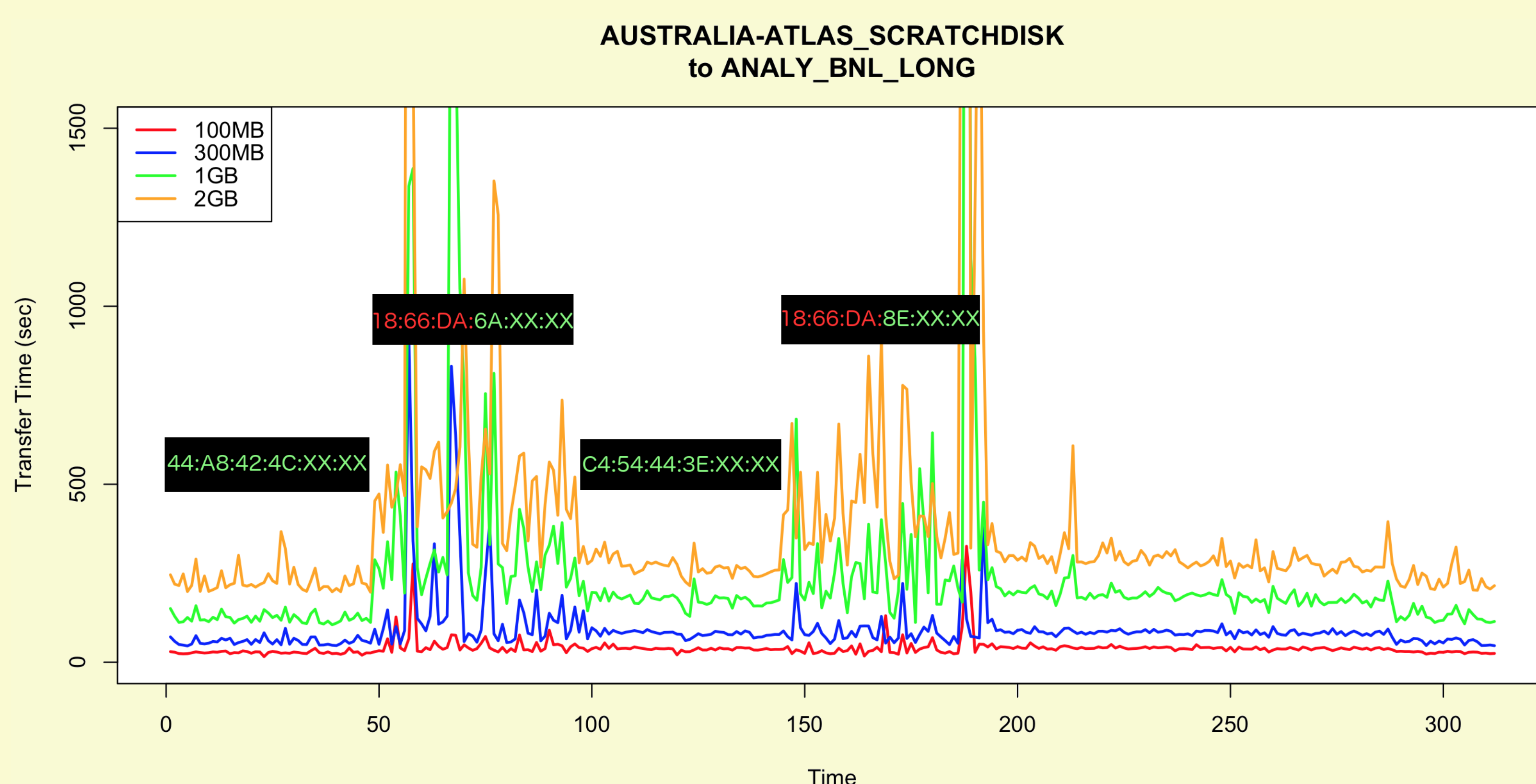
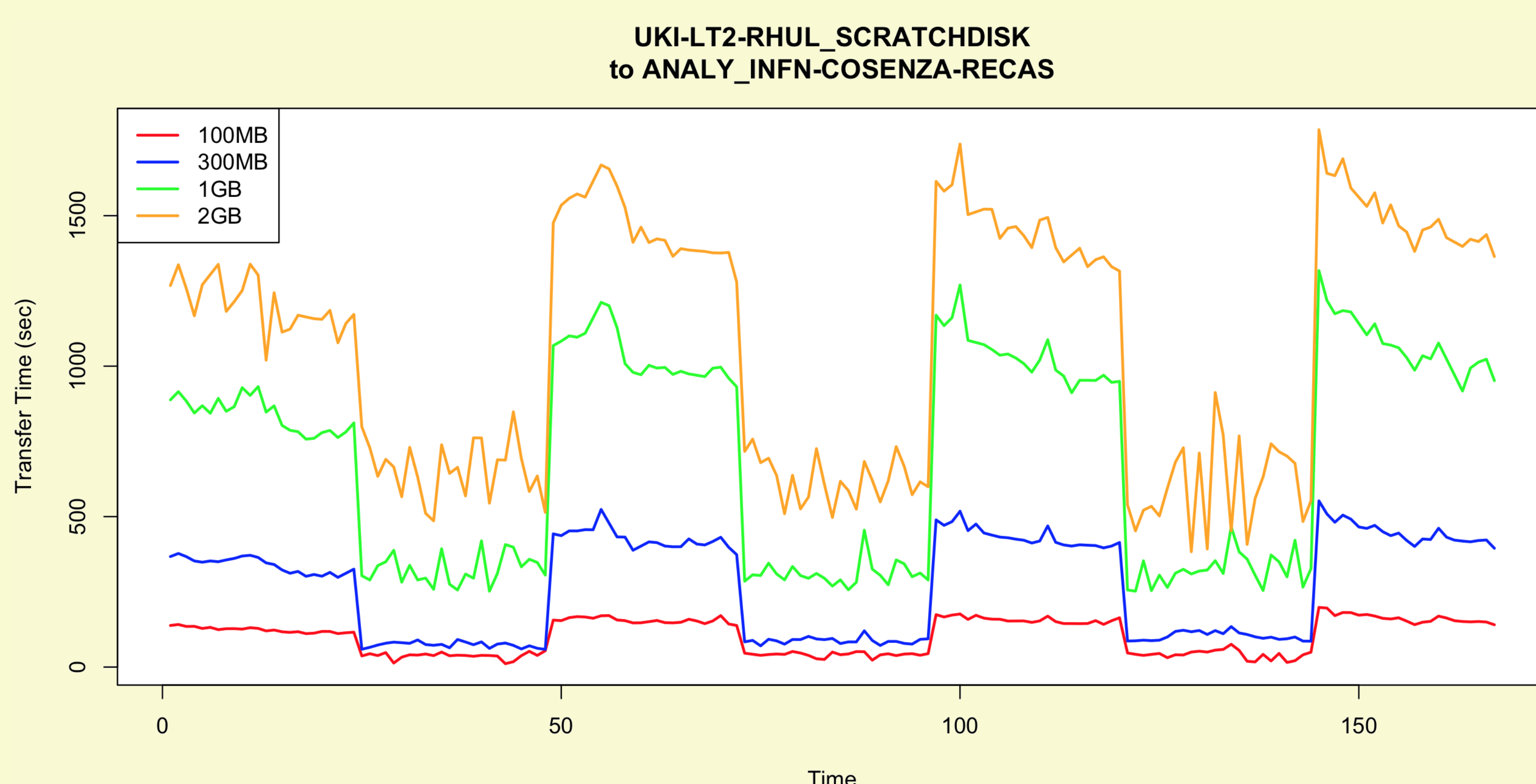
And following types of concurrent traffic:

- Traffic across various jobs
- Traffic across various file accesses per job
- Internet cross-traffic

Deployed Protocols

HTTP/WebDAV, TCP based

Observations in Production



Creation of various network loads among sampling phases demonstrates a bottleneck on the worker node side

Volume (GB)	Phase	Concurrent Volume (GB)	Transfer Time (sec)	σ^2
0.1	2	0.4	40.3	128
0.1	1	1.99	149.7	365
0.3	2	0.67	88	274
0.3	1	5.18	412	2935
1	2	1.44	319	2286
1	1	10.2	995	14944
2	2	1.4	626.8	11609
2	1	11.3	1407.4	24154

Transfers re-assigned to a new worker node once per sampling phase

Different network interface controllers exhibit dramatically varying data throughput in terms of stability and speed

E.g., the MAC address prefix **18:66:DA** is assigned to DELL

Otherwise, the throughput rates are pretty constant in the long term