1 Understanding BGP using table dumps

a) AS 3

b) 144.228.241.81

c) Only 1, not including MIT (namely, it must traverse AS 1239)

d) No, it is not

e) Because MIT would rather send traffic cheaply over Internet2, when it is allowed (which is when the traffic is between two educational institutions, mostly). MIT has a routing entry that directs traffic intended for U of Oregon (and other universities, presumably) to its Internet2 link. The traceroute likely does not match the AS path because the IP address-to-AS number resolution, in the traceroute, is accomplished via queries to a database that map IP addresses to the database owner’s concept of AS. The AS numbers in the AS path are reported by the actual AS, and the database that maps IP addresses to ASes may no longer be aware of which ASes have which IP addresses (IP addresses can be resold). Moreover, there are cases in which the same logical AS uses different AS numbers. In this example, the queries to the database induced by the "-A" option can only return one AS number, so it might be returning one different from what the AS reports as its own. In any case, note that the AS Path in the MIT table is logically the same as the one the traceroute indicates – through various Internet2 links to U. of Oregon.

f) 41

g) The best route is the one with AS PATH=1239 3. There are two routes with the shortest AS path: (1239 3) and (3356 3). Both these routes also seem good enough with respect to the MED, eBGP vs. iBGP and
IGP cost metrics. Hence, some criteria to break the tie (like, router ID or oldest route first) seems to have been used to choose the best route in this case.

h) Sprint (1239), Level3 (3356), Cogent (174), Internet2 (11537) is another AS that provides service to MIT through Harvard (10578)

2 Equal-airtime Wi-Fi

a) An allocation of times \( t_i \) to channel \( i \) is proportionally fair when it maximizes \( \sum_i \log U_i \) where \( U_i = r_i t_i \) is the utility and throughput. We want to maximize \( \sum_i \log(r_i t_i) \) such that \( \sum t_i = 1 \). Each \( t_i \) is the fraction of each second assigned to a given channel \( i \). We can write an objective as \( \max \sum_i \log t_i + R = \max \prod_i t_i + R \) where \( R = \sum r_i \) is constant. We can optimize this objective function, using Lagrange multipliers or AM-GM inequality, and find that the equation is optimized when \( t_i = \frac{1}{n} \) for all \( i \).

b) The min-max throughput is the harmonic mean: \( \frac{n}{\sum(\frac{1}{r_i})} \). The proportional fair throughput is the arithmetic mean: \( \sum \frac{r_i}{N} \).

3 Datacenter Network Design

a) \( c_{i,j} = R \) for \( j = N - i + 1 \) and 0 otherwise. The total cost is \( N \cdot R \).

b) 1. Any TOR may need to send at rate \( R \) to any other TOR, so we need \( c_{i,j} = R \) for all \( i \neq j \). The total cost is \( N \cdot (N - 1) \cdot R \).

2. TOR \( i \to j \) traffic flow is routed via a randomly chosen intermediate TOR. A somewhat subtle point is that it’s important for TOR \( i \) to split traffic to each destination \( j \) evenly across intermediate nodes (e.g., it is not enough to split all outgoing traffic from TOR \( i \) evenly, but have all traffic from a specific destination go to one intermediate node).

To determine the cost, consider the link between a pair of TORs: \( i \to j \). This link must carry \( \frac{1}{N} \) of the traffic sent by TOR \( i \) (to any destination), and \( \frac{1}{N} \) of traffic received by TOR \( j \) (from any source). Since the traffic matrix is admissible, the most this can be is: \( \frac{1}{N} \cdot R + \frac{1}{N} \cdot R = \frac{2R}{N} \).
Therefore, \( c_{i,j} = \frac{2R}{N} \) is sufficient to support any admissible traffic matrix, and the total cost is \( N \cdot (N - 1) \cdot \frac{2R}{N} = 2 \cdot (N - 1) \cdot R. \) Notice that VLB has \( \frac{N}{2} \) less cost than direct routing.

c) 1. A full mesh topology is not possible in this case, because each TOR must dedicate \( 2 \times \) more ports to uplinks (connected to other TORs) than downlinks (connected to servers) in order to support all admissible traffic matrices with VLB (based on the solution to part (b,2)). Therefore each TOR can at most attach to 8 servers, leaving 16 ports for uplinks. But the number of TORs would then be \( \frac{288}{8} = 36 \), so 16 uplink ports is not enough to create a full mesh.

2. This is straightforward with 24 TOR switches, each attached to 12 servers and using the other 12 ports to connect to a layer of spine switches in a Clos topology. In total the topology requires 36 switches: 24 ToRs and 12 Spines.