NETMET 2021-2022
Midterm exam – 8 February 2022
Duration: 1 ½ hours

Allowed: English dictionary or bilingual dictionary (printed books only, with no written notes)

Not allowed: Documents, computers, pocket calculators, mobile phones, etc.

Answer sheet and anonymous ID

The first sheet of paper in this exam is your answer sheet. It bears an anonymous ID number. This number, recorded on the attendance sheet for the exam, provides the connection between your answer sheet and your identity.

Please do not write your name on the answer sheet.

You will submit only the answer sheet at the end of the exam.

Multiple-choice questions numbered M1, M2, ... (1 point each)

The multiple-choice questions each have four possible answers: A, B, C, and D.

A fifth choice X indicates "no answer". Mark your answers by writing A, B, C, D, or X in the box alongside the question number.

For each multiple-choice question, there is only one correct answer.

- Correct answer = 1 point
- Incorrect answer = ½ a point subtracted
- No answer = 0 (no points gained, none lost either)

Because of the penalty for an incorrect answer, it is not in your interest to make a random guess.

If there is ambiguity, and we cannot tell which answer you have provided, we will need to mark the answer as incorrect.

Short answer questions numbered S1, S2, ... (1 point each)

Please use the space provided in which to write your answers to the short answer questions.

Additional remarks

Do not hesitate to raise your hand with any questions you might have. Best wishes for a successful exam!

Multiple choice questions for Answer Sheet 1

M1. Which of the following statements is **False**?

A. "Active measurements" refers to the technique of sending probes and inferring network and/or application properties based upon the responses

B. Passive measurements rely on the observation of existing traffic

C. An advantage of active measurements is that they do not introduce overhead.

D. Monitoring routing messages is one example of passive measurements

M2. What is the most popular method for measuring available bandwidth of residential broadband links and why?

A. Packet pairs with small, back-to-back probes, because the probing overhead is small.
B. Size-delay methods because they can measure per hop bandwidth
C. Flooding over multiple TCP connections because results are reliable in the presence of different types of cross-traffic
D. Flooding over a single TCP connection because it reduces probing overhead

M3. Which of the following statements is **incorrect** about interface counts:

A. Interface counts are useful for customer billing
B. Interface counts allow us to directly measure the Origin-Destination traffic matrix.
C. Interface counts are well suited for detecting volume anomalies
D. None of the above

M4. The traffic matrix is a useful way of representing traffic of a network. How can operators obtain the traffic matrix?

A. It is impossible to measure the traffic matrix directly, so operators must estimate it.
B. Network operators can use flow statistics to measure the traffic matrix
C. Network operators can use packet captures to measure the traffic matrix
D. B and C.

M5. A network operator observes a large spike in the number of bytes in the network. After digging further, she observes that the number of packets also increased, but not the number of flows. What can explain this spike?

A. A worm scanning for vulnerable hosts
B. A large file transfer.
C. A distributed denial of service attack
D. None of the above

M6. Which of the following statements is **True**?

A. NIDS are capable of detecting both known and zero-day attacks
B. Anomaly detection identifies statistical outliers and is suitable for volume attacks.
C. NIDS are proactive rather than reactive
D. All of the above

M7. A network telescope suddenly observes a large volume of TCP acks from a given IP address X. What can we infer about X?

A. Nothing.
B. X is the victim of a denial of service attack
C. X is a host in the same subnet as the network telescope
D. X is infected by a worm

M8. The following table represents an anonymized dataset. What are the privacy properties of this dataset?

Zipcode	Age	Disease
476**	2*	Heart Disease
476**	2*	Heart Disease
476**	2*	Flu
4790*	≥40	Flu
4790*	≥40	Heart Disease
4790*	≥40	Flu

A. 2-anonymous and 3-diverse
B. 3-anonymous and 2-diverse
C. 2-anonymous and 2-diverse
D. None of the above.

M9. A research group wants to study the popularity of websites over time for a population of users from packet traces collected at a large company's network. In which format should the network administrator release the data to help the researchers and yet minimize the risk of leaking any sensitive information?

A. Release the full packet traces as long as IP addresses are anonymized
B. Release packet traces, but only for HTTP(S) traffic and with IP addresses anonymized
C. Process the packet traces to extract only the visited host in HTTP(S) packets and release a trace with timestamps and visited hosts.
D. Any of the above

M10. Select the statement that **correctly** matches the anonymization technique with its privacy property:

A. t-closeness ensures t-diversity for the sensitive attribute in each equivalence class
B. k-anonymity guarantees protection against cross-referencing attacks
C. Differential privacy guarantees that nothing can be learned about an individual while useful information can be learned about a population
D. None of the above

M11. Which of the following statements represent a challenge for large scale censorship measurements?

A. Recruiting geographically distributed vantage points
B. Collecting pervasive and continuous measurements across a wide range of URLs
C. A and B
D. Measuring censorship targeting a specific URL from a specific location for a limited period of time

M12. If a censor is blocking the site www.example.com using DNS blocking at the ISP DNS resolver, what is the best way for users to evade it?

A. Using DNSSEC
B. Change www.example.com IP address
C. Encrypting the traffic
D. None of the above .

M13. Censorship can cause collateral damage (instances where allowed traffic gets blocked together with censored traffic). Select the **True** statement.

A. IP-based blocking with ACL does not cause collateral damage because it only blocks a single IP address
B. Blocking access to a third party DNS server does not lead to collateral damage
C. Both A and B are True
D. None of the above .

M14. Which of the following statements represent a challenge of measuring Online Social Network (OSN) graphs?

A. OSNs limit the number of requests
B. Some nodes are isolated
C. OSN graphs are large and highly dynamic
D. All of the above.

M15. A number of companies today explore data publicly available in OSNs. What are the challenges to precisely identify the different accounts of a given user in multiple OSNs?

A. OSNs have millions of users so it is likely to have multiple users with the same attributes (e.g., the same name)
B. Users may hide or simply not provide some information
C. Users may provide inconsistent information in different OSNs
D. All of the above.

M16. Which of the following statements is **False** about sampling methods of OSN graphs?

A. An effective random sampling approach is to explore all of the neighbors of a node before visiting the nodes at the next depth level
B. Graph sampling techniques that randomly select the next node to visit are inherently biased to high-degree nodes .
C. Sampling techniques that eliminate bias to high degree nodes provide the closest estimate to random sampling
D. Breadth-First Search is not an effective graph sampling method as it only covers a region of the OSN graph

Short answer questions for Answer Sheet 1

S1. Encore is a system developed by researchers at Georgia Tech to measure large-scale censorship measurements. Briefly describe how Encore achieves this, and discuss one ethical implication of Encore and one solution to reduce Encore's risks.

S2. K-anonymity has been criticized for its weak privacy preserving guarantees. Provide two examples of privacy attacks that can be performed on a k-anonymous dataset.

S3. There are many ways to perform IP-based blocking; one such method is known as BGP hijack. Briefly explain how BGP hijack is performed, how it can be detected, and suggest one solution to prevent it.

S4. What makes security-related measurements different from other network measurement tasks? Justify your answer.

Multiple choice questions for Answer Sheet 2

M17. Which of the following will allow a route tracing tool to trace an accurate path through a typical load-balanced network topology?

A. keeping the TTL constant
B. keeping the source and destination port numbers constant ,
C. varying the destination address
D. varying the protocol field

M18. Approximately how many autonomous systems (ASes) are announcing prefixes (either IPv4 or IPv6 or both) into the internet? (Choose the closest value.)

A. 1,000
B. 10,000
C. 100,000
D. 1 million

M19. What variant of BGP will a border router of one AS use to communicate information to the border router of a neighboring AS?

A. eBGP
B. iBGP
C. BGPv1
D. None of the above

M20. In latency measurements, what is the propagation delay for packets travelling up to a sattelite in geostationary orbit and back?

A) 240 nanoseconds
B) 240 microseconds
C) 240 milliseconds
D) 240 seconds

M21. Which is **true** of GeoPing geolocation?

A. A target X is geolocated at the location of one of the agents that is pinging it.
B. A target X is geolocated at the location of a landmark that is not an agent.
C. A target X is geolocated at the center of overlapping disks.
D. A target X is geolocated at the location of a landmark with which it shares an extensive traceroute path.

M22. Which is **true** of so-called Street-Level geolocation?

A. A target X is geolocated at the location of one of the agents that is pinging it.
B. A target X is geolocated at the location of a landmark that is not an agent.
C. A target X is geolocated at the center of overlapping disks.
D. A target X is geolocated at the location of a landmark with which it shares an extensive traceroute path.

M23. Which is **true** of Shortest Ping geolocation?

A. A target X is geolocated at the location of one of the agents that is pinging it.
B. A target X is geolocated at the location of a landmark that is not an agent.
C. A target X is geolocated at the center of overlapping disks.
D. A target X is geolocated at the location of a landmark with which it shares an extensive traceroute path.

M24. Which is **true** of CBG geolocation?

A. A target X is geolocated at the location of one of the agents that is pinging it.
B. A target X is geolocated at the location of a landmark that is not an agent.
C. A target X is geolocated at the center of overlapping disks.
D. A target X is geolocated at the location of a landmark with which it shares an extensive traceroute path.

M25. Which distance calculations will be accurate enough for use in global scale geolocation of internet addresses?

A. Euclidean and Haversine
B. Vincenty and Euclidean
C. Haversine and Vincenty
D. All three: Euclidean, Haversine, and Vincenty

M26. How would you obtain a round trip time that can best be used to estimate the distance between two hosts on the internet?

A. Take the minimum of a series of RTT measurements
B. Take the mean of a series of RTT measurements
C. Take the median of a series of RTT measurements
D. Take the maximum of a series of RTT measurements

M27. Which of the following might make it difficult to estimate the distance between two hosts on the internet based on round trip time measurements?

A. asymmetric routing paths between hosts
B. signals do not always travel at the speed of light in a vacuum
C. packets do not necessarily travel in straight lines from one host to the next
D. all of the above

M28. What is the **most reasonable guess** if one sees a three letter airport code in the name of a router interface?

A. the router is located at that airport
B. the router was shipped by its manufacturer to the ISP via that airport
C. the router is located in or near the city served by that airport
D. the resemblance to a three letter airport code is just a coincidence

Short answer questions for Answer Sheet 2

S5. List four components of network latency, and give, for each one, a rough range of values that one might measure for that type of delay.

S6. Show, via a diagram, how geolocation via trilateration works.

S7. Briefly describe how Topology-Based Geolocation works.