euronf
NETWORK OF EXCELLENCE

cost
EUROPEAN COOPERATION
IN SCIENCE AND TECHNOLOGY

Exam Info

The assignments must be sent directly to the professors before March 7th, 2010.

They will evaluate each assignment and send the outcomes to the organizers.

When all the grades are available, organizers will send the outcome of the final result to the students.

We expect this to be done before the end of March.

For further logistical information, please email the organizers at tma.school@gmail.com.

For any other questions related to each proposed exam, please ask directly the related professor.

Christian Callegari

Perform the review of one of the following paper:

a) W. Lee, D. Xiang, Information theoretic measures for anomaly detection, in: Proceedings of the 2001 IEEE Symposium on Security and Privacy, Washington, DC, USA, 2001, pp. 130-143.
b) W. Wang, R. Battiti, Identifying intrusions in computer networks with principal component analysis, in: The First International Conference on Availability, Reliability and Security, Vienna, Austria, 2006, pp. 270-279
c) W. Wang, X. Guan, X. Zhang, A novel intrusion detection method based on principle component analysis in computer security, in: Proceedings of the International Symposium on Neural Networks, Dalian, China, 2004, pp. 657-662
d) N. Ye, Y.Z.C.M. Borror, Robustness of the Markov-chain model for cyber-attack detection, IEEE Transactions on Reliability 53 (2004) 116-123
e) M. Ramadas, S.O.B. Tjaden, Detecting anomalous network traffic with self-organizing maps, in: Proceedings of the 6th International Symposium on Recent Advances in Intrusion Detection, Pittsburgh, PA, USA, 2003, pp. 36-54
f) S.T. Sarasamma, Q.A. Zhu, J. Huff, Hierarchical Kohonenen net for anomaly detection in network security, IEEE Transactions on Systems, Man and Cybernetics-PART B: Cybernetics 35 (2005) 302-312.
g) L. Portnoy, E. Eskin, S.J. Stolfo, Intrusion detection with unlabeled data using clustering, in: Proceedings of the ACM Workshop on Data Mining Applied to Security, Philadelphia, PA, 2001
h) K. Sequeira, M. Zaki, ADMIT: Anomaly-based data mining for intrusions, in: Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Edmonton, Alberta, Canada, 2002, pp. 386-395.
i) V. Hautamaki, I. Karkkainen, P. Franti, Outlier detection using k-nearest neighbour graph, in: Proceedings of the 17th International Conference on Pattern Recognition Los Alamitos, CA, USA, 2004, pp. 430-433
j) J. Zhang, M. Zulkernine, A hybrid network intrusion detection technique using random forests, in: Proceedings of the First International Conference on Availability, Reliability and Security, Vienna University of Technology, 2006, pp. 262-269

The review must be sent directly to Cristian Callegari, christian.callegari@iet.unipi.it.

Maurizio Molina

NetFlow and Anomaly detection – Review questions

1) In a DDoS syn flood analysed with the use of NetFlow v5 sampled data
   a) the absence of return traffic means that the target was taken down
   b) the presence of return traffic means that the target did not suffer major performance problems
   c) the target replies with and ack/rst if the source IP of the packet is spoofed, with syn/ack otherwise
   d) the presence of limited return traffic may indicate the presence of a syn rate limiting firewall or the fact that the attack disrupted server performances

2) The “source AS” information in NetFlow
   a) was introduced after NetFlow v5
   b) always reports the AS number of the host originating the IP flow
   c) reports the AS number of the host originating the IP flow or of the previous AS in the path with respect to the interface where NetFlow is collected
   d) is impossible to find if the origin host is behind a NAT

3) The Active timeout
   a) is useful for reporting long lasting flows
   b) must be always bigger than the inactive timeout
   c) must be set taking into account the sampling rate
   d) is not needed if NetFlow is exporting flows using TCP as a transport protocol

4) NetFlow v5
   a) has a fixed format for exporting per flow information
   b) is more reliable than NetFlow v9
   c) is the basis for the IPFIX protocol
   d) will not be supported any more in one or two years

5) IP flow monitoring
   a) gives the same information as packet level monitoring
   b) gives the same information as packet level monitoring if sampling rate is 1:1
   c) gives information that could be extracted by packet level monitoring as well
   d) treats the whole flow as a single aggregate packet

6) IPFIX
   a) is still in a draft status, will be approved in 2010
   b) can use TCP, UDP, STCP
   c) can use TCP, UDP, SNMP
   d) extend NetFlowv9 on Cisco implementations but not on other routers

7) Sampled NetFlow
   a) is functionally equivalent to sFlow
   b) is by default 1/100

c) samples one packet every N
d) samples one flow every N

8) The Inactive Timeout

a) depends on the NetFlow sampling rate
b) is a safeguard against the flow cache growing too much in size
c) must be set on the basis of the average flow size in packets
d) must be set on the basis of the average flow duration

9) In a DoS UDP flood analysed with the use of NetFlow v5 sampled data

a) there are usually a lot of packets involved, but just a few flows
b) there are a lot of flows because the sender continuously changes the destination port
c) the destination port is always one in this set: 22, 53, 80, 443, 445, 139, 135, 1443
d) senders are normally more than one, and they exchange control traffic on TCP between them

10) By collecting NetFlow on all the peering interfaces of a network, in the incoming direction, one can estimate

a) all traffic exchanged by peers
b) all traffic exchanged by peers, multiplying by a factor of two
c) all traffic exchanged by peers, but collecting on transit interfaces too the estimate is more precise
d) all traffic exchanged by peers, but collecting in the outgoing direction too the estimate can be more precise

11) NetFlow v9

a) is not suitable for security related applications
b) is template based
c) has a fixed number of aggregation schema
d) can run on the top of TCP, and on UDP only if the router supports also NetFlow v5

12) IP traffic feature entropies

a) are useful for detecting scans and DDoS syn floods, but not DoS UDP floods
b) are useful for detecting scans, but not DDoS syn floods
c) are useful for detecting scans, DDoS syn floods and DoS udp floods
d) are useful for detecting low volume anomalies only

13) NetFlow defines the format of the information

a) inside the flow cache
b) inside the collector
c) between the exporter and the collector
d) between the flow cache and the exporter

14) In a network scan analysed with the use of NetFlow v5 sampled data there are a lot of flow because

a) the scanner continuously changes its source port
b) the scan induces a lot of ack/rst responses by the targets
c) the scanner sends a lot of packets continuously changing the destination IP address
d) the scanner changes the destination port

15) Multiplying by 1/S the number of packets in a flow (where S=sampling rate)

a) is an unbiased estimator of the flow size (in packets)
b) is an unbiased estimator of the flow size (in packets) in deterministic sampling, but not in random sampling
c) underestimates the flow size (in packets) in case of small flows
d) overestimates the flow size (in packets) in case of small flows

16) In a port scan analysed with the use of NetFlow v5 sampled data
a) there are usually more flows than in a network scan
b) there are a lot of flows because packets have a small size (in bytes)
c) there are a lot of flows because the scanner sends a lot of packets continuously changing the destination address and the destination ports
d) one may or may not see return traffic from the target to the scanner

NetFlow and Anomaly detection – Answer sheet

Name:
Affiliation:
e-mail:

Pass is 12/16

no.	Answer (choose **only one** among a,b,c,d)	**Tutor use only keys**	**Tutor use only Correct (Y/N)**	Comments (if you think question is ill posed
1				
2				
3				
4				
5				
6				
7				
8				
9				
10				
11				
12				
13				
14				
15				
16				

Result (tutor use only)

Correct	wrong	percentage	Pass/Fail

Please fill this page and send it to Maurizio Molina (Maurizio.Molina@dante.net) **in Word format** using, as a subject, the string "**netflow_exam_polito_name**" where "name" is the family name of the student.

Kave Salamatian

The exam will be extremely simple. Please redo lab3 completely and Lab 5 (Kalman filtering and anomaly detection) and send me a detailed report at kave.salamatian@univ-savoie.fr.

More advanced project : Do Lab 4 and try to combine a PCA model with a Kalman Filter filter step described in Lab 5.

The main issue is to find the State space model for PCA. This can be done by calibrating a model on the data before PCA and applying the formulas given in this technical report[1] in it section III.C to get the model after PCA and inject it into the Kalman Filter. If somebody put time to do this, this results in a nice paper!!!!

SO THERE IS A STRONG MOTIVATION THERE !

Possible updates on this project can be found on the following web page:

http://web.me.com/kave.salamatian/Site/Blog/Entr%C3%A9es/2010/2/15_Exam_or_better_said_project.html

Please feel free to contact me for any questions at kave.salamatian@univ-savoie.fr

[1] http://web.me.com/kave.salamatian/Site/Blog/Entr%C3%A9es/2010/2/15_Exam_or_better_said_project_files/infocom2009.pdf

Patrick Thiran and Katerina Argyraki

Question 1

Consider a network that consists of end-hosts, routers, and links (like the networks we encountered during the network-tomography lecture). The operation of a router in this network is very simple: when a packet arrives through an ingress link, the router looks up the packet's destination in a routing table and determines through which egress link to send the packet (as in classic IP routing).

Suppose we want to keep track of the loss rate incurred by each of the links in this network. To this end, we use network tomography, i.e., we periodically measure the loss rate incurred by each end-to-end path, then use this information to infer the loss rate incurred by each link.

Now suppose that, to prevent bad end-hosts from flooding the network with unwanted traffic, we change the way routers operate as follows.

Each router maintains two pieces of state:

1. A routing table (as before)
2. A statistics table. This table has two columns: The first column specifies all the destinations toward which the router has routed packets within the last minute. The second column specifies how many packets toward each destination the router has routed within the last minute. E.g., the statistics table for a router that has routed packets only toward two destinations within the last minute could look like this:

Destination	Number of packets forwarded within the last minute
1.2.3.4 12.34.56.78	1,234,567 20,345,6788

When a packet arrives through an ingress link, the router performs the following operations:

1. It looks up the packet's destination in the statistics table and reads the corresponding packet counter.
2. If this counter is equal to some pre-configured threshold X, the router drops the packet.
3. Otherwise, the router increments the counter by one, looks up the packet's destination in the routing table, and determines through which egress link to send it.

I.e., each router tries to protect each destination from being overwhelmed by packets, by rate-limiting the number of packets that it routes toward that destination.

Assume that the performance cost of looking up a destination and reading/updating a counter in the statistics table is zero.

The question is, with this anti-flooding mechanism in place, can we use network tomography to infer

the loss rate of each link?
Think about which are the assumptions on which network tomography relies and whether the anti-flooding mechanism causes any of them to be violated.

(The anti-flooding mechanism described here is an (over)simplified version of the Pushback mechanism: http://research.microsoft.com/en-us/um/people/ratul/papers/ccr2002-pushback.pdf)

Question 2

Read and review the following paper, which appeared in the Network Systems Design and Implementation (NSDI) conference in 2008:
http://www.cs.washington.edu/homes/ckd/phalanx.pdf

In your review, provide a short summary of the paper and list its strong and weak points.

As part of your review, consider (at least) the following questions:

1. What type(s) of attack is the proposed mechanism trying to handle?
2. Which is/are the targeted resource(s)?
3. What is the state employed by the proposed mechanism and where is it stored?
4. Does the proposed mechanism require any changes in the current network infrastructure?
5. Does the proposed mechanism violate the end-to-end principle?
6. How would you try to circumvent the mechanism if you were a bot master?

Send the answers to katerina.argyraki@epfl.ch.