COMS 4180 Network Security Assignment 3

Due Tues, April 22nd, 2014, 11:00pm Eastern time. 100 points
The homework is to be done individually.
NO LATE HOMEWORK WILL BE ACCEPTED. The answers will be given in class on the 24th, which is the last lecture before the exam.

What to submit: A text/pdf/word file with your answers. Submit the file via courseworks.

1. 10 points
Look at the default rules in Spamassassin available at
https://spamassassin.apache.org/tests_3_3_x.html

a. (3 points) Think of how the rules can be categorized, for example several of the rules use ”white lists”. List 5 other general types of rules used by Spamassassin (i.e., white list cannot be one of the 5) You may only include one list type in the 5. (i.e. don’t write 5 different types of lists as your answer.) Do not use the values under the ”test name” column from the website as your answer, although the term you use to categorize a subset of the rules may be a term that appears in some of the test names (for example, "whitelist" appears in some of the test names and is a way to categorize those tests).
b. (3 points) Spamassassin, with its default rules and settings, does a poor job of distinguishing phishing emails from legitimate emails. What are some possible reasons for this?
c. (2 points) Spamassin assigns negative points if the email has features indicative of a valid email. List 5 such rules used by Spamassassin.
d. (2 points) In the default settings, the points assigned to a rule indicating spam range from almost zero to 1000. The most weight (1000 points) is given to a ”Generic Test for Unsolicited Bulk Email”, the first rule in the list. Most other rules are assigned a fraction of a point or a few points, but there are a couple given substantially more weight than the others (but less than the ”Generic Test for Unsolicited Bulk Email”). What type(s) of rules are these and what their point values? Hint - they are near each other but not next to each other in the list.

2. 15 points (5 points for each pseudo code/script)
Suppose you had an IDS that could read short scripts for detecting a piece of malware. You are given short segments of communication between a bot and master or commands executed by the bot for three different infections. For each one, write a short segment of pseudo code (or use actual shell script or programming syntax) that will detect the presence of the bot. Assume the packet contents are available to the scripts in the text form shown on three files. The pseudo code or scripts will be read to determine the grade; they will not be executed.

The communications are in the three files:
hw3_prob3_ex1.txt
hw3_prob3_ex2.txt
hw3_prob3_ex3.txt

In the files, messages from the master are labeled "Attacker" and messages from or commands issued by the bot are labeled "Victim".

3. 40 points

Perform static analysis on the five binaries test1,test2,test3, test4 and test5 to determine if any of the executables are similar. In what programming language(s) do you think they are written? Were any of them packed by UPX and how can you tell?

You must decide what tools to use for static analysis. If there is any static analysis you want to use but for which you don't have a tool, you may write your own code but there is no need to submit the code with your answer.

Include in your answer:

(1) a list of what you used to perform the static analysis

(2) for each possible pairing of programs, state whether or not they are similar in any way and explain how you know from your static analysis

(3) for each program state what language you think it is and explain how you know from your static analysis

(4) for each program, state whether or not it was packed by UPX and explain how you know from your static analysis

4. 30 points (20 for part a, 10 for part b)

Scenario: there is a Linux system on which runs a set number of processes performing some task. The server is up 24 x 7. The only user on the system is the administrator, who logs in as root periodically to perform routine checks and maintenance.

a, Using only the data available from the netstat and top commands, describe in detail how an IDS can monitor the Linux system for suspicious connections and usage. Include how often to collect data, what data (from that produced by netstat and top) to use and what statistics to compute. Also discuss the likeliness of false positives and false negatives.

b. Suppose the developers of the software running on the system are given logins to the system. How would this impact the IDS collection and analysis of the data and how may it impact the accuracy of the results?

5. (5 points)
Would the signing of OSPF route updates prevent propagation of routing errors due to configuration errors such as typing errors or entering of erroneous data when provisioning a router? Why or why not?