CSC 495 — Assignment 4 — Due Tuesday, April 28

1 CSC 495 — Assignment 4 — Due Tuesday, April 28

Note: April 18 is the last day of classes for the semester, but that day will actually follow the Friday class schedule which means we won't meet. You can drop your completed assignment off in my office on the 28th, or you can turn it in early during our last class meeting (April 23), or you can email me your assignment (I prefer to receive hardcopy, but will accept emailed submissions for this assignment).

1. The goal of this question is for you to learn a little about the most cutting-edge static analysis tools for PHP, and to get a little more experience with security research. To start, you should read the following paper (handed out in class, but also available from the link below):

   > Johannes Dahse and Thorsten Holz. "Simulation of Built-in PHP Features for Precise Static Code Analysis," *Proceedings of the Network and Distributed System Symposium (NDSS)*, 2014. Available at http://www.internetsociety.org/doc/simulation-built-php-features-precise-static-code-analysis

   As you start reading this paper, focus on high-level ideas and don't worry about all the details. You can go back and fill in details later, once you see the "big picture."

   a. After reading the paper, set it aside and write a one paragraph summary of the paper in your own words. Make sure you use the term "taint analysis" and describe what this term means.

   b. The authors listed several PHP language features that make static analysis difficult. Pick one such language feature that you found surprising or interesting, describe it, and explain why it causes difficulties for static analysis.

   c. Do you consider this research to be a success? Use the results as described in the Introduction and more fully in Section IV to justify your answer.

   d. Listings 15 and 16 (on pages 11 and 12) show how an XSS vulnerability is exploited in the `osCommerce` software. In particular, if the exploit URL in Listing 16 is used, then the `echo` statement in Listing 15 will output HTML that contains a cross-site scripting exploit. Work through this code and give the exact HTML output that is produced in this exploit.

   e. Published research papers contain comparisons and references to related work that was published prior the paper, so someone reading the paper has some context for the work, and can find prior papers to see how the ideas have evolved. A natural question when reading a paper turns this around to ask "What related work has been done *after* this was published?" Obviously, the paper can't include references to work that hadn't been done before the paper was published, so people use citation search engines to find papers that cite this one. Google Scholar, at http://scholar.google.com is a good example of this. For this part, go to Google Scholar and type in the first part of the title of this paper to find it. Next, look for the “Cited by ..” link with this paper, and click on it. This will provide a list of subsequent papers that have cited this paper as related prior work. For this paper in particular, compare the papers that come up in this search with the description of “Future work” in the final paragraph of the paper. Find a paper that reports on a future work item — give a reference to this paper, with a brief description of the problem that it addresses.

2. For this question, read the short essay:

   > Shriram Krighnamurthi and Jan Vitek. “The Real Software Crisis: Repeatability as a Core Value,” *Communications of the ACM*, March 2015, pp. 34—36. Available at http://cacm.acm.org/magazines/2015/3/183593-the-real-software-crisis/fulltext

   Think about how this relates to Section IV.F. of the paper from the first question, where the authors try to compare their system to previous systems, but had to make somewhat indirect comparisons since the code for previous systems wasn’t available to make head-to-head comparisons. The essay gives several reasons why software (“artifacts”) might not be published, and there are also clear benefits to making the code public. What is your opinion? Take a position either for or against requiring publishing research software (no being wishy-washy!), and write a short argument for your position. Justify your position in the context of the essay, making sure you address arguments both for and against publishing that are outlined in that essay.

3. Consider the following PHP code for checking a user’s login credentials (username and password). Assume that nothing has been done elsewhere in the code to sanitize user inputs.

```
$query = "SELECT login FROM users WHERE login=’". $_POST[’userid’] .
            "’ AND password=’" . $_POST[’pass’] . "’";

$resp = $dbconn->query($query);
if ($resp->num_rows > 0) {
    $_SESSION[’loggedin’] = 1;
    $_SESSION[’username’] = $_POST[’userid’];
} else {
    $_SESSION[’loggedin’] = 0;
}
```

   a. Describe this vulnerability in the same style used on your midterm exam. Specifically, give the following four pieces of information: (a) a description of the *vulnerability*, (b) a brief description of how an attacker could *exploit* the vulnerability, (c) what the *consequences* of the attack are (revealing secrets, taking over the machine, etc.), and (d) at least one way to *correct the problem*. For each answer, clearly label the parts of your answer using the emphasized words above.

   b. For this part, describe an exploit in detail. In other words, give a specific input that an attacker could use to log in as user “admin”, even without knowing the password. Also give the resulting `$query` string for your input.