False Positives And Software Measurement

Computer Measurement Laboratory
info@cmlab.biz

At CML we have developed a new measurement based software process control technology that permits us to measure executing software systems in real time. One of the areas that simply begs for the application of this technology is computer security. To that end, we have designed and developed the CML Hardware-enabled Monitoring of Software (HMS) system. This is fundamentally a measurement approach to the modeling of executing programs. We really want to know what programs look like when they are executing in a normal, or certified, manner.

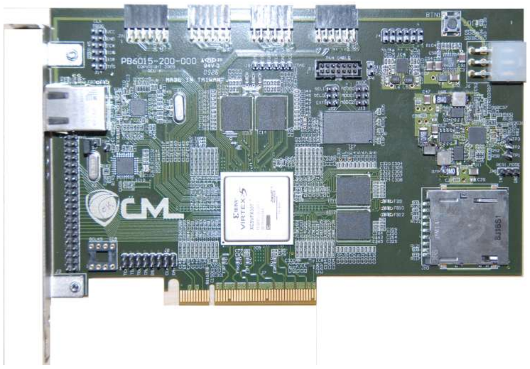

Figure 1. Opus HMS

To do this measurement, we have designed and built a PCI Express card that will permit us to monitor all of the software that is executing in a host Linux kernel environment. We can see a picture of this Opus HMS measurement tool in Figure 1.

In its calibration mode, this measurement tool will build a model of normal or certified behavior for each program that is to run on the host Linux side. When a Opus HMS system is placed into its operation role, it will monitor the executing software and constantly measure the execution behavior of each certified system. (It simply will not permit non-certified programs to execute).

The function of the Opus HMS measurement system is very much like another measurement tool that we learned to use in first grade. This tool is shown in Figure 2. It is called a ruler. The purpose of this tool is to obtain a number for the length of an object in our environment. The marks of this particular ruler will allow us to measure things down to the nearest 1/16 of an inch or 1mm.

Figure 3. A ruler

As we began to introduce our new Opus HMS measurement tool to the computer security community one of the first questions that came up was "what is your rate of false positives with this tool?" The answer to this question is pretty simple. It is precisely the same as the false positives encountered when applying the ruler. It simply doesn't make sense to ask that question about a ruler. Similarly, it simply doesn't make sense to ask that question about the Opus HMS system. It is simply a measurement tool.

In the early stages of the emerging discipline of biology, one of the most influential members of this community was a man by the name of Carl Linnaeus. His great contribution to science was a taxonomy that would permit plants and animals to be classified into various families, genus and species. At the beginning of the new discipline of biology, this classification scheme made a whole lot of sense. It introduced order. Unfortunately, with the recent advances in the analysis of DNA we find that a great deal of this classification effort was really misguided.

Similarly, in the relatively new field of computer security, the notion of classification is a driving force. Vulnerabilities are classified. Exploits are classified. However, this need for classification is driven by ignorance just as was the case in Linnaeus' day. The major problem with the classification approach is that it can only be applied ex post facto. You must have been able to witness a number of events of the Manama type in order to create a new taxonomic class for it. We would like not to have had our software compromised in the first place. There is little or no value in attaching a name to the event.

Let us now return to the notion of the ruler false positives. We can take our ruler and make it serve as a classification device. We could employ our ruler, say, in the manufacture of copy paper. We could take the ruler shown in Figure 2 and use it as a go no-go gauge. We would measure each sheet of paper to insure that it was within some manufacturing tolerance for the length of the paper. We might insist, for example, that each sheet of paper be 11 inches long plus or minus 1/100 of an inch. Unfortunately, our ruler only has a resolving power of 1/16 of an inch. This measurement tool is clearly inadequate to the task it has been assigned to perform. It will cause a lot of sheets of paper to be rejected when they were, in fact, just fine (false positives). It will also permit us to accept a lot of sheets of paper when, in fact, they were either too big or too small (false negatives).

The problem with the ruler in the above example was not in the ruler itself. It was in the inappropriate use of the ruler. It did not have the resolving power to make the discriminations necessary in the application to which it was applied.

The bottom line is that the concept of false positives (or false negatives, for that matter) simply does not apply to the Opus HMS system. HMS simply measures the current activity of a program and compares this activity against a standard model of normal activity established during a calibration phase. It continuously monitors the difference between the measurements of the software as it is currently executing to a standard model for each software system.

Through the User Interface to HMS we can set an upper bound for that difference. In so doing, the HMS system will generate an interrupt on the host computer whenever this threshold is crossed for an executing program. We, in turn, will intercept this interrupt in our Adaptive Engine and take whatever action that has been agreed upon with the Security Administrator in advance. We might choose to kill the process, to slow the process down so that its activities may be observed, to sandbox the process, or possibly ban the user that created the process from the system. If, of course, the difference threshold is set very low, the system will permit virtually no divergence from normal or certified behavior. If, on the other hand, the threshold is set very high, then the HMS system will permit a fairly wide range of behavior that is divergent from the standard model. Thus, false positives or false negatives are an artifact of the user (Security Administrator), not the basic technology.

One of the primary objectives in the design of our current Opus HMS system has been to establish the right level of measurement granularity. If the granularity of measurement is too low, there will be tremendous data bandwidth in the measurement process. If the granularity of measurement is set too high, then we will not be able to resolve the very abnormalities that we seek to understand. Getting the measurement granularity right has been the subject of many years of research and observation.