Does it make sense to benchmark security? - CSAM Case Study #2

José Donato, donato@student.dei.uc.pt, 2016225043

I. Introduction

This article aims to study the security benchmarking and whether it is worth doing it or not. I start by talking about the differences between assessment and benchmarking and the different types of benchmarking. In the section after I discuss, in more detail, security benchmarking. To conclude, I give my opinion based on the topics exposed before whether it makes sense or not to benchmark security systems, giving my thoughts about the paper "On the Brittleness of Software and the Infeasibility of Security Metrics" [1].

II. Assessment and Benchmarking

Before we can define assessment, we have to define another term, measuring. Measuring has the goal of discovering a number to define a characteristic of an object or an event, i.e., to get a quantitative value of something using a well known scale/reference.

Measuring is a subset of Assessment. Assessment encompasses a qualitative approach beyond the quantitative. It can be subjective and tries to rank something for what it is worth.

In benchmarking, there is an agreement between the people doing the experience to achieve some comparison between the products being tested, following a specific procedure. It must have a high level of representativeness (the conditions of the experience must simulate what would happen in the "real" world). It must also be useful because otherwise no one will care about the results of the benchmark.

When we are performing a benchmarking, we have three main concepts:

- Workload: type of operations that will be performed in the benchmarks, it influences the representativeness
- Measures (or metrics): What is measured from the experience (for example, the GPU took 5 seconds to load the video)
- Rules/procedure: Specify what needs followed during the benchmark execution. They must be easy to follow because the benchmark need to be easy to repeat in each environment. It influences the reproducibility.

After knowing what a benchmark is, we can define computer benchmarks: standard tools to compare and assess different systems about specific characteristics (for example, the GPU performance).

- Performance benchmarking: focus on a specific domain and compare different systems (databases, OSs, etc). Commonly used by vendors to promote their products (marketing). The problem of this is that during the tests it is assumed that there are no errors and everything always works.
- Dependability benchmarking: see if a system reacts well in case of problems. Includes performance benchmarking and fills its problem. Faults are inserted to see how the system reacts. The faults inserted do not take into account the malicious behaviours that result from the exploitation of some vulnerability in a system.
- Security benchmarks: Includes both performance and dependability benchmarking and is a recent topic. Different than the others because of the existence of zero-day vulnerabilities (every day appear software security flaws that have no patch to fix in place). The goal is to rate and compare systems in terms of security.

III. Security Benchmarking

Security benchmarking is much more recent than the other types of benchmarking. Instead of fault injection as occur in dependability benchmarking, in addition to those injections, we insert malicious attacks. What fails in this type of benchmarking is that we don't know all the vulnerabilities and attacks for a specific system/software so we can't insert an attackload that represent all the attacks that a malicious attacker could perform. As said in the "Benchmarking Untrustworthiness: An Alternative to security Measurement" paper [2], the "key difficulty is that security is usually more influenced by what is unknown about a system than by what is known". We can't get good metrics because we don't know how to represent a good attackload.

Note that we should only do security benchmarking after fixing all known vulnerabilities (using a vulnerability detection tool to detect those and then fix them). Even if our system has zero known vulnerabilities, there are always vulnerabilities that can exist that we don't have knowledge about and only the attacker has. In the other hand, Security benchmarking shows great results for studying vulnerability detection tools for vulnerabilities that we have knowledge (we inject faultloads that contain vulnerabilities and attacks and then compare the results between each tool). Besides this, there are some companies like BitSight that say they are making security benchmarking a reality [3] by using security ratings [4] to rank systems in terms of security.

IV. Opinion

In my opinion, I think the author of the paper "On the Brittleness of Software and the Infeasibility of Security Metrics" [1] has a point. It is hard to establish a comparison between the systems about its security strength because there are no systems impenetrable.

Although, I don't agree that we cannot improve security. For example, a program made by a student just to pass some course from first year of bachelor (developed in an ad hoc manner, not testing the app, etc), will be far more insecure that an application developed carefully following a development methodology, CIS guidelines and relentlessly tested [2]. Although we don't have a metric that justify how insecure each application really is (and even this is becoming wrong because of the rating systems made by companies such as BitSight [3]), since the first application was developed following a far worse process, we are much more suspicious of this one on a security approach.

Based on this last thought, a new metric called trustworthiness appear in the security benchmarking. This way, we can rank systems based on their security trustworthiness (in the example above, the second application has for more trustworthiness than the first one).

References

[1] Steven M. Bellovin. On the Brittleness of Software and the Infeasibility of Security Metrics. https://www.cs.columbia.edu/ smb/papers/01668014.pdf.

[2] Marco Vieira Afonso Neto. Benchmarking Untrustworthiness. https://www.igi-global.com/article/benchmarking-untrustworthiness-alternative-security-measurement/46937.

[3] BitSight. Make security benchmarking a reality. https://www.bitsight.com/blog/make-security-benchmarking-a-reality.

[4] BitSight. Security ratings for benchmarking. https://www.bitsight.com/security-ratings-for-benchmarking.