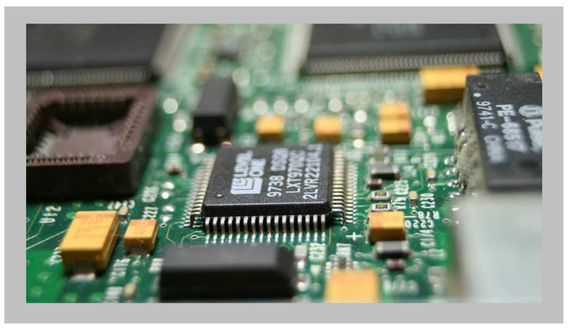
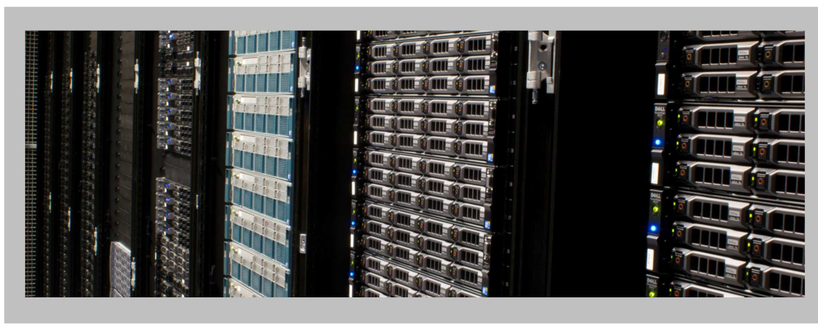
SAND2016-7025D

Exceptional service in the national interest

Sandia National Laboratories





Disk Usage as an Indicator of Vulnerability in Software Products

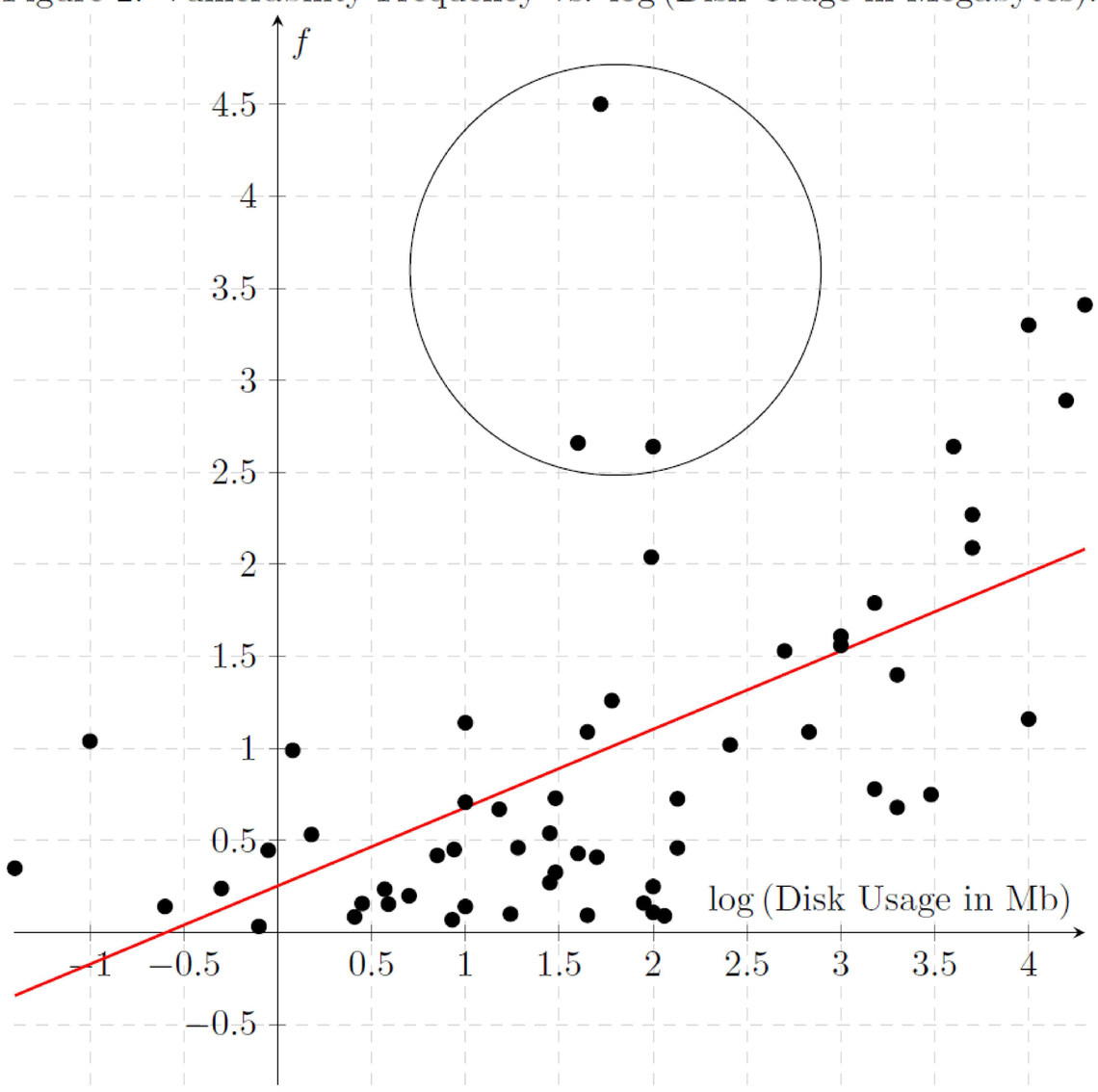
Sandia National Laboratories has developed a methodology for assessing the risk associated with Information Technology (IT) purchases called SARA, the Supply Chain Acquisition Risk Assessment. The current SARA methodology estimates the risk of an IT product acquisition by studying the product's threat, vulnerability, and consequence. For software products, a contributor to vulnerability is the complexity inherent in the product. The current methodology captures this contribution in the Code Length factor. Previous studies have found that the number of lines of code is a good proxy for the complexity of software, and that the complexity of software is related to the number of vulnerabilities present.

However, the Lines of Code (LOC) of a software application is not a commonly documented or freely available statistic. LOC information is only readily available for open-source software products. For closed-source products, the only sources are press releases and news articles, which can be unreliable and sometimes conflict. A better metric is desired.

The vulnerabilities in a piece of software can instead be approximated by the disk space requirements of that software. Disk space requirements are published with the application as part of its hardware requirement specifications. This metric is more accessible to the SARA analyst and provides a vulnerability estimate that is as good as, if not better than the current methodology's.

In this study, ~100 open source and closed source products with vulnerabilities listed on the National Vulnerability Database were studied. The disk space requirements for each product were compared to the rate at which vulnerabilities were reported for that product using the equation $ft = \sqrt{n}$, where $n =$ number of vulnerabilities discovered and $t =$ years the product has been available. This relationship is currently used in another section of the SARA methodology.

Figure 2: Vulnerability Frequency vs. log (Disk Usage in Megabytes).



There does appear to be a relationship between disk usage and product vulnerability modeled by the equation:

$$f = 0.4257 * \log(d) + 0.254$$

$d =$ disk usage in megabytes, $r^2 = 0.33$

The notable outliers were internet applications like Google Chrome, Firefox, and Thunderbird. Since these applications routinely create network connections, the large number of vulnerabilities associated with these products is not unusual. Removing these products from the dataset improves r^2 to 0.49. Since there appears to be a moderate correlation between software size and vulnerability, disk space usage may replace code length as a software vulnerability factor in the SARA methodology using the scoring table depicted here.

Scoring Table Comparisons

Current

Score	Interpretation (LOC)
10	>100,000,000
8.2	10,000,000–100,000,000
6.4	1,000,000–9,999,999
4.6	100,000–999,999
2.8	10,000–99,999
1	<10,000

Proposed

Score	Interpretation
10	>10 Gb
8.2	1 Gb to 10 Gb
6.4	100 Mb to 1 Gb
4.6	10 Mb to 100 Mb
2.8	1 Mb to 10 Mb
1	< 1 Mb

Hankun Zhao
University of California – Berkeley
B.S. Electrical Engineering/Computer Science, est. 2019
Mentor: Margaret Todd
Org: 8112
July 27th, 2016

U.S. DEPARTMENT OF ENERGY
National Nuclear Security Administration

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.
SAND No. 2011-XXXXP