Security Metric of the Week #28: Benford's Law

Authors:
Krag Brotby, CISM, CGEIT
Gary Hinson, PhD, MBA, CISSP

Benford's law is a fascinating theorem in number theory with applications in information security, accountancy, engineering, computer audit and other fields.

Benford's law predicts the distribution of initial digits on numbers in numeric data sets generated in an unbiased and unconstrained fashion. In short, roughly a third of such multi-digit numbers start with a 1, whereas only one twentieth start with a 9. If someone (such as a fraudster) or something (such as a rogue or buggy computer application) has been manipulating or fabricating data, the numbers tend not to have leading digits with the predicted frequencies. Turning that on its head, if we compare the actual against predicted distributions of leading digits in a data set, significant discrepancies probably indicate something strange, and possibly something untoward going on: we would have to dig deeper to determine the real cause.

The PRAGMATIC scores for this metric are as follows:

P	R	A	G	M	A	T	I	C	Score
84	30	53	95	11	98	62	98	23	62%

Benford's law is normally used to analyze data sets for fraud, and as such the metric has some merit as a fraud indicator. However, a data set that complies with Benford's law may have been manipulated by a fraudster clever enough to ensure that his fictitious numbers have the predicted frequencies of initial digits. This is not an altogether unrealistic scenario, since successful fraudsters are indeed clever and manipulative by nature.

The need to explain the mathematical basis for the metric to most audiences detracts from its Meaningfulness score. The Timeliness and Cost-effectiveness scores are depressed by the practicalities of obtaining and analyzing sufficient volumes of raw data and exploring the real reasons for any skewed distributions. As far as we know, there are limited applications of Benford's law to information security, hence the low Relevance score. While Benford's law is highly Accurate (if applied correctly) and Independent, it is only Actionable if the reasons for skewed distributions are

understood (for instance identify and fire the fraudster, or diagnose and debug the rogue program).