Detection of Cyber-Attacks of Power Systems through Benford's Law

Federico Milano, *Fellow, IEEE*, and Antonio Gómez-Expósito, *Fellow, IEEE*

***Abstract*—This letter proposes an application of the Benford's law for the detection of cyber attacks in power system state estimators. Benford's law, also known as 1st-digit law, states an unexpected property of the distribution of the first digit of certain sets of data, and has been found to apply to a surprisingly wide range of data domains. The first novel contribution of the letter is to show that the Benford's law applies to power system data as well. A relevant property of this law is its high sensitivity to manipulations and, in fact, it is often utilized to detect frauds. Based on this feature, the second contribution of the letter is to utilize the Benford's law to detect malicious data introduced by hackers in the supervisory control and data acquisition (SCADA) system of a transmission network. Tests based on power system models ranging from 9 to 21,177 buses show promising results.**

***Index Terms*—Benford's law, state estimation, cyber attack, bad data.**

I. Introduction

The state estimation of power systems is a critical analysis carried out by system operators and heavily depends on metering, communication networks and computers to elaborate the data. A failure of the state estimator can lead to severe consequences, e.g. the Northeastern blackout occurred on August 2003 that affected about 55 million people across US and Canada. Future integrated cyber-physical systems such as the smart grid are expected to depend even more than conventional power systems on sensing and communication networks for the control, operation and billing of the devices that are connected to the grid [1].

This letter focuses on the malicious injection of bad data in power systems, often referred to as "cyber attacks". The first studies on cyber attacks of power systems date back to 2009 [2], i.e. when the first case of infiltration of "spies" in an US grid was reported [3]. The interest on this topic has significantly grown since the 2015 Ukraine blackout [4], which is one of the major real-world cyber attacks to date. The interest in this topic is also reflected by a number of review papers on cyber attacks, e.g. [5] and, more recently, [6].

The objectives of this letter are twofold: (i) to demonstrate that the set of measurements of a power system available to system operators through the SCADA system follows the Benford's law; and (ii) to show that the distribution of the first digits of the measurements is highly sensitive to malicious manipulations caused by hackers. Various metrics are proposed to detect the presence of bad data in the set of measurements.

II. Benford's Law

The Benford's law, also known as the *law of anomalous numbers* or *1st-digit law*, describes the frequency distribution of the leading digits of a set of data. Its general expression is:

$$\mathcal{B}_b(i) = \log_b\left(1 + \frac{1}{i}\right), \quad i = 1, 2, \ldots, 9\,, \tag{1}$$

where b is the base of the number and i is digit of interest. The values of $\mathcal{B}$ for $b = 10$ are given in Table I. Note that the the digit 0 is not taken into account in (1). The law can be generalized to digits beyond the first. Considering for simplicity $b = 10$, the frequency distribution for the j-th digit, with $j > 1$, is given by:

$$\mathcal{B}^j_{10}(i) = \sum_{k=10^{(j-2)}}^{10^{(j-1)}-1} \log_{10}\left(1 + \frac{1}{10k+i}\right), \quad i = 0, 1, \ldots, 9\,. \tag{2}$$

For $j > 1$, the frequency of the digits tends to a uniform distribution (see Table I). For this reason, the most relevant applications of the Benford's law generally consider the first digit only. Moreover, the second and following digits of the measurements affect the results of the state estimation much less than the first digit. Accordingly, exclusively the first-digit version of the Benford's law is utilized in this letter.

TABLE I: Benford's law for the first 3 digits of numbers in base 10.

Digit	0	1	2	3	4
1st	–	0.301	0.176	0.125	0.097
2nd	0.120	0.114	0.109	0.104	0.100
3rd	0.102	0.101	0.101	0.101	0.100
Digit	5	6	7	8	9
1st	0.079	0.067	0.058	0.051	0.046
2nd	0.097	0.093	0.090	0.088	0.085
3rd	0.100	0.099	0.099	0.099	0.098

The Benford's law fits quite well a surprisingly large variety of very diverse naturally-occurring sets of data. For example the frequency of the first significant digit of physical constants; the distances of stars from the Earth; and the leading digit of the series of numbers obtained calculating 2^n for $n = 1, 2, \ldots, \infty$. There are also several practical applications of the Benford's law, ranging from price digit analysis and genome data to fraud detection of socio-economic as well as scientific data. The latter application is particularly relevant for the matter of this letter.

There are several interpretations of the Benford's law, e.g. entropy related interpretations, multiplicative fluctuations of certain series of data, e.g., stock prices. An interesting observation is that the Benford's law is strictly correlated to the distribution of the measured data [7]–[9]. In particular, numbers following continuous probability distributions that are common in engineering, e.g., Normal, Weibull, Gamma, etc., tend to have their first digits distributed according to the Benford's law. On the other hand, the first relevant digits of numbers following the uniform distribution do not follow the Benford's law. This property is exploited in the letter (see Section IV-B of the case study).

III. Application of Benford's Law to Power Systems

We assume that a set of m measurements at given times t_k, $k = 1, 2, \ldots, p$, are available to the system operator. These measurements,

F. Milano is with School of Electrical and Electronic Engineering, University College Dublin, Belfield Campus, Dublin 4, Ireland. E-mail: federico.milano@ucd.ie

A. Gómez-Expósito is with Electrical Engineering Department, University of Seville, Seville, Spain. E-mail: age@us.es

This work is supported by the Science Foundation Ireland, by funding F. Milano under Grant No. SFI/15/IA/3074; and by the Spanish CDTI, by funding A. Gómez-Expósito under Cervera Grant No. CER-20191019.

say $\boldsymbol{z}^{(k)} = \boldsymbol{z}(t_k) \in \mathbb{R}^m$, are functions of the states $\boldsymbol{x}^{(k)} = \boldsymbol{x}(t_k) \in \mathbb{R}^n$ of the system, as follows:

$$\boldsymbol{z}^{(k)} = \boldsymbol{h}(\boldsymbol{x}^{(k)}) + \boldsymbol{\epsilon}^{(k)} + \boldsymbol{\eta}^{(k)}, \qquad k = 1, 2, \ldots, p. \tag{3}$$

where $\boldsymbol{\epsilon}^{(k)}$ are the measurement errors, which are assumed to be normally distributed and with zero-average [10]; and $\boldsymbol{\eta}^{(k)}$ is the vector of malicious data introduced by the hacker. In normal operation, thus, $\boldsymbol{\eta}^{(k)} = \mathbf{0}$.

Most papers on state estimation and cyber attacks pay a huge attention to the actual kind of measurements, i.e., on equations $\boldsymbol{h}$, on the redundancy of the measurements, distinction between topology and measurement errors, as well as on the kind of information and data available to the hacker. On the other hand, for the application of the technique discussed in this letter, no particular hypothesis is required except for the fact that $\boldsymbol{\eta}^{(k)}$ is assumed not to be distributed as $\boldsymbol{\epsilon}^{(k)}$. In particular, we assume that the cyber attack consists of bad data that are obtained by either substituting the original measurements with uniformly-distributed random values or swapping measurements. The assumption that malicious data introduced by hackers distribute uniformly is justified by empirical observations in other fields where the Benford's law has been successfully applied, e.g. tax frauds. False data that do not satisfy this assumption might not be detected with the proposed technique. On the other hand, since normally-distributed zero-mean errors can be filtered relatively well by a robust state estimator, one can argue that a cyber attack that introduces malicious data in the form of $\boldsymbol{\epsilon}^{(k)}$ is not actually a threat for the system [11].

In this letter, we propose to check the quality of the measurements $\boldsymbol{z}$ through the Benford's law. However, instead of using the measurement values as they are, we consider a set of normalized variations with respect to the reference value of each measurement. This operation avoids that certain numbers are artificially more common than others. For example, in a transmission system where nominal bus voltages are mostly 220 or 380 kV, the digits 2 and 3 would be very common if no normalization is done on the measurements.

Thus, we consider the distribution of the first relevant digit of the elements of a vector $\boldsymbol{u} \in \mathbb{R}^{m \times p}$, whose elements are defined as::

$$u_{h,k} = \left| \frac{\tilde{z}_h^{(k)} - z_{r,h}}{z_{o,h}} \right|, \quad h = 1, \ldots, m, \quad k = 1, \ldots, p. \tag{4}$$

where $z_{r,h}$ is the reference value of the measured quantities and $z_{o,h}$ is the base value utilized to obtain per-unit quantities.

Equation (4) is adapted depending on the measured quantity. For example, for voltage magnitudes $z_{r,h} = z_{o,h} = V_N$, where V_N is the nominal voltage of the bus where the measurement is taken. Similarly, for frequency measurements, $z_{r,h} = z_{o,h} = \Omega_o$, where Ω_o is the synchronous reference angular frequency of the systems. For active and reactive power injections, $z_{r,h} = 0$ and $z_{o,h} = S_{\text{base}}$, where S_{base} is the power base utilized in power flow analysis. Voltage phase angles are assumed in radians and, hence $z_{r,h} = \theta_{\text{ref}}$ and $z_{o,h} = 1$, where θ_{ref} is the angle reference of the slack bus of the system.

For example, if $\boldsymbol{u} = [0.1, 0.729, 0, 0.384, 1.2, 0.0022]$, the first relevant digits have frequencies 40% for digit 1, 20% for digits 2, 3 and 7, and 0 for digits 4-6, 7 and 9. Note that, if the ith measurement is $|u_i| < 1$, then $|u_i|$ is repeatedly multiplied by 10 until the first digit is not null. Note also that null measurements are omitted in the calculation of the digit frequencies, as (1) is not defined for $i = 0$.

It is important to note that utilizing (4) is not strictly necessary for the purpose of the identification of cyber attacks. If no normalization of the measurement is used, each network has its own "fingerprint" of first digits, which is perturbed by bad data. However, with (4), the distribution of the first digits approximates the distribution predicted by the Benford's law for all networks, or at least all networks that we have tested, and significantly simplifies the detection of bad data. This is illustrated in the last example of Section IV-B.

IV. Case Study

The objectives of the case study are twofold. First, we show that the Benford's law effectively applies to power systems. While we do not have a conclusive explanation on why this happens, yet we have found that this is actually the case for all tested networks. In Section IV-A, we show a small randomly-picked selection of the several networks that we have considered. Then, Section IV-B discusses the effect of malicious data in the distribution of the first digits of the measurements. Three metrics are defined to evaluate the impact of cyber attacks.

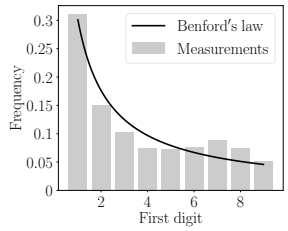
A. Examples

Figure 1 shows the results for four power systems with different sizes, topologies, devices and controllers. The measurements used to obtain the histograms comprise both the standard ones in state estimation, e.g. bus voltage magnitudes, power injections at buses and power flows in transmission lines; and those provided by phasor measurement units, e.g. bus voltage phase angles and bus frequencies. Note that null power injections are not considered in the calculation of the distribution of the first digits. In any case, any modification of zero-power injections can be easily detected and filtered by the system operator. A variable number of measurement snapshots, p, is utilized per each network, ranging from $p = 10$ to $p = 10,000$. Figure 1 has been obtained assuming $\boldsymbol{\epsilon}^{(k)}$ normally distributed, with zero mean and with a standard deviation 1% of the actual values of the measurements.

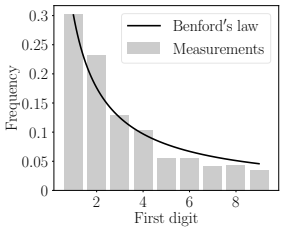
Results indicate that the Benford's law matches with a very good accuracy. In fact, the relative square deviation index – see (7) – returns values lower than 0.05 in all tested cases. As a general rule, the higher the number of variables and measurements of the system, the more accurate the Benford's law is, although, in turn, this very much depends on the normalization of the measurements. Another observation is that bus voltage phase angle measurements are those that mostly cause deviations with respects to the Benford's law. It is important to note that, for every network, set of measurements and choice of the parameters for the normalization function (4), the histogram has a quite fixed shape. This in turn means that the distribution of the first digits comes with a sort of "fingerprint" that is characteristic of that specific network. This property, even more than the accuracy with which the distribution approximates the Benford's law, allows identifying malicious data. The introduction of bad data, in fact, will inevitably create a distortion of the fingerprint of the first digits. The main advantage of using the normalization of the measurements is that the operator does not need to know *a priori* this fingerprint. Reciprocally, the operator can adopt alternative *ad hoc* normalizations (sort of encryption keys), which could even change in time, e.g. daily. In absence of bad data, each normalization would give rise to a specific "fingerprint," unknown to the hacker, that will be perturbed to a certain extent by potential cyber-attacks.

B. All-Island Irish Transmission System

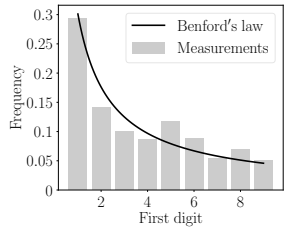
In this section we consider a real-world model of the all-island Irish power system. The topology and the steady-state operation data of the system have been provided by the Irish transmission system operator, EirGrid Group, whereas the dynamic data have been defined based on our knowledge about the technology of the generators and the controllers. The system consists of 1,479 buses, 796 lines, 1,055 transformers, 245 loads, 22 synchronous machines, with automatic



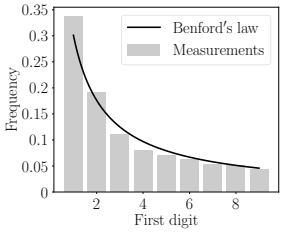
(a) WSCC 9-bus system



(b) IEEE 14-bus system



(c) New England 39-bus system



(d) 21,177-bus ENTSO-E system

Fig. 1: Benford's law applied to power systems of various sizes.

voltage regulators and turbine governors, 6 power system stabilizers and 176 wind power plants. Simulations are solved using Dome [12].

In the tests discussed below, we assume that the system operator has available about $m \approx 10,000$ measurements per snapshot of the system, and $p = 1,000$ snapshots, including bus voltage magnitudes, active and reactive power injections at buses and flows in transmission lines and transformers, and bus frequency measurements. Figure 2 shows the distribution of the first digits assuming $\boldsymbol{\eta}^{(k)} = \mathbf{0}$ and $\boldsymbol{\epsilon}^{(k)}$ normally distributed, with zero mean and with a standard deviation 1% of the actual values of the measurements.

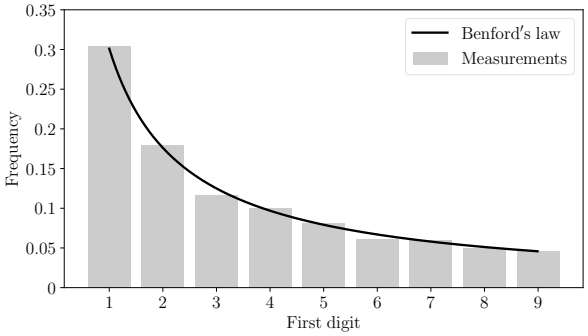


Fig. 2: Distribution of first-digits of normalized measurements for the all-island Irish system.

We define the following linear deviation index (LDI) to evaluate the impact of malicious data:

$$\text{LDI} = \sum_{i=1}^{9} |c_i - \mathcal{B}_{10}(i)|\,, \tag{5}$$

where c_i is the quota of the measurements whose first relevant digit is i normalized with respect to the total number $(m \times p)$ of measurements; a digit-1 index (D1I):

$$\text{D1I} = c_1 - \mathcal{B}_{10}(1)\,; \tag{6}$$

and, finally, a relative square deviation index (RSDI), as follows:

$$\text{RSDI} = \sum_{i=1}^{9} \frac{[c_i - \mathcal{B}_{10}(i)]^2}{\mathcal{B}_{10}(i)}\,. \tag{7}$$

The values of the indices above obtained for the Irish system are given in Table II for various levels of bad data. All indices are very sensitive to the bad data level, returning values that differ 120% (D1I), 33% (LDI) and 150% (RSDI) for a bad data level introduced by a cyber attack as low as 1%.

TABLE II: Indices for different percentages of bad data introduced by a cyber attack.

Bad data [%]	D1I	LDI	RSDI
0	0.0037	0.0304	0.0014
1	−0.0433	0.1314	0.0227
2	−0.0725	0.2114	0.0601
3	−0.0915	0.2664	0.0960
4	−0.1061	0.3057	0.1270
5	−0.1168	0.3340	0.1530
10	−0.1469	0.4184	0.2414
15	−0.1602	0.4551	0.2867

It is interesting to note that the normalization obtained by applying (4) prevents that a cyber attacker can cheat the Benford's law by "swapping" a set of measurements. Say, for example, that the attack consists in systematically swapping voltage magnitude measurements with active power injections at the same bus. If the raw measurements were utilized, the pattern of first digits would remain the same and no issue could be detected by checking their distribution. On the other hand, using (4), a swap of the measurements is immediately visible from the distribution of the first digit. Table III shows the values of the indices in (5) to (7). Again all indices are quite sensitive even to low levels of bad data.

TABLE III: Indices for different percentages of swapped data introduced by a cyber attack.

Swapped data [%]	D1I	LDI	RSDI
0	0.0037	0.0304	0.0014
1	0.0088	0.0423	0.0040
2	0.0132	0.0634	0.0109
3	0.0178	0.0895	0.0211
4	0.0219	0.1143	0.0349
5	0.0262	0.1375	0.0502
10	0.0433	0.2423	0.1621
15	0.0580	0.3303	0.3045

As a last example, we illustrate the concept of "fingerprint" introduced in Section III. The normalization obtained with (4) has been introduced to obtain a distribution of the first digits that fits the classical Benford's law. A hacker that knows that measurement data are double-checked through the Benford's law, might be able to introduce attacks that pass undetected. To avoid this possibility, the normalization step can be utilized to make the shape of the distribution of the first digits unique, and thus serve as an "encryption" that cannot be unravelled by hackers.

Figure 3 shows the distribution of the first digits assuming that active power measurements are normalized with the following formula:

$$p_{h,k} = \left| \frac{\tilde{P}_h^{(k)} - 0.5 S_{\text{base}}}{S_{\text{base}}} \right|, \tag{8}$$

The coefficient 0.5 introduced in (8) leads to high bins for the digits 4 and 5. Table IV shows that a cyber attack can be still be easily detected, even for small percentages of malicious data, assuming that the system operator has a good knowledge of the distribution of the first digits for a situation where no cyber attack contaminates the measurements. In Table IV the indexes D1I, LDI and RSDI have been calculated considering the deviations of the data with respect to

the values of the digits shown in Figure (3). For this reason, the first row of Table IV is null.

Finally, the following remark on measurement errors is relevant. We have assumed so far that these errors are normally distributed. In practice, they can have other distributions and can, thus, impact on the shape of the distribution of the first digits. Independently from their distribution and properties, however, the errors due to the instrumentation show a substantial difference with respect to the false data introduced by cyber attackers. That is, genuine errors are an intrinsic part of any set of measurements and are expected to be always present, to a certain extent. Cyber attacks, on the other hand, constitute extraordinary events, artificially introducing additional errors, which are not distributed in the same way as "natural" errors. The impact of intrinsic measurement errors, thus, is part of the "fingerprint" of the distribution of the first digits of a set of noisy measurements. Hence, these errors not only are not an issue for the detection of cyber attacks but, in case they give rise to a peculiar distribution of the first digits, they can actually also help to detect the cyber attacks.

TABLE IV: Indices for different percentages of bad data introduced by a cyber attack with normalization of the active power measurements based on (8).

Bad data [%]	D1I	LDI	RSDI
0	0	0	0
1	−0.0217	0.1375	0.0383
2	−0.0355	0.2257	0.1026
3	−0.0440	0.2860	0.1649
4	−0.0511	0.3278	0.2173
5	−0.0562	0.3616	0.2633
10	−0.0707	0.4549	0.4166
15	−0.0770	0.4962	0.4961

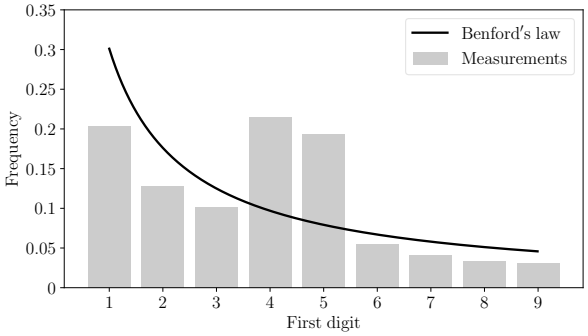


Fig. 3: Distribution of first-digits of measurements for the all-island Irish system with normalization of the active power measurements based on (8).

V. Conclusions

The letter shows that the Benford's law applies to a set of typical measurements utilized for the state estimation of power systems if these measurements are properly normalized. No particular assumption is made on the information available to the hacker, which can be either partial or complete. The only assumption made is that the hacker does not introduce variations with zero mean and Gaussian distribution on the measurements. Results allow concluding that the frequency of digit 1 and a relative square deviation index are particularly sensitive to the injection of bad data and/or data swapping. In fact, bad/swapped data levels as low as 1% can be easily detected. Custom normalizations of the data can also be utilized as an encryption to protect against hackers that are aware of the utilization of data sanity checks based on the Benford's law. The most relevant advantages of the proposed technique are its negligible computational burden and full independence from conventional state estimation tools. However, it does not provide information on which data have been hacked. For this reason, the proposed technique can be utilized to raise a "red flag" on suspicious sets of measurements. Then, conventional bad-data detection techniques can be utilized to carry out further analyses and identify which data are false. Given its high sensitivity to malicious data, the proposed pre-screening technique can be particularly efficient and reduce the possibility of "false positive" detections.

References

[1] O. Kosut, L. Jia, R. J. Thomas, and L. Tong, "Malicious data attacks on the smart grid," *IEEE Transactions on Smart Grid*, vol. 2, no. 4, pp. 645–658, 2011.

[2] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proceedings of the 16th ACM Conference on Computer and Communications Security*. New York, NY, USA: Association for Computing Machinery, 2009, p. 21–32.

[3] S. Gorman, "Electricity grid in U.S. penetrated by spies." [Online]. Available: http://online.wsj.com/article/SB123914805204099085.html

[4] G. Liang, S. R. Weller, J. Zhao, F. Luo, and Z. Y. Dong, "The 2015 Ukraine blackout: Implications for false data injection attacks," *IEEE Transactions on Power Systems*, vol. 32, no. 4, pp. 3317–3318, 2017.

[5] R. Deng, G. Xiao, R. Lu, H. Liang, and A. V. Vasilakos, "False data injection on state estimation in power systems—attacks, impacts, and defense: A survey," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 411–423, 2017.

[6] A. S. Musleh, G. Chen, and Z. Y. Dong, "A survey on the detection algorithms for false data injection attacks in smart grids," *IEEE Transactions on Smart Grid*, vol. 11, no. 3, pp. 2218–2234, 2020.

[7] R. M. Fewster, "A simple explanation of Benford's law," *The American Statistician*, vol. 63, no. 1, pp. 26–32, 2009.

[8] A. K. Formann, "The Newcomb-Benford law in its relation to some common distributions," *PLoS ONE*, vol. 5, no. 5, 2010.

[9] S. W. Smith, *The Scientist and Engineer's Guide to Digital Signal Processing*. USA: California Technical Publishing, 1997.

[10] A. Abur and A. Gómez Expósito, *Power System State Estimation: Theory and Applications*. New York, NY: CRC Press, 2004.

[11] A. de la Villa Jaén, J. B. Martínez, A. Gómez-Expósito, and F. G. Vázquez, "Tuning of measurement weights in state estimation: Theoretical analysis and case study," *IEEE Transactions on Power Systems*, vol. 33, no. 4, pp. 4583–4592, 2018.

[12] F. Milano, "A Python-based software tool for power system analysis," in *IEEE PES General Meeting*. IEEE, 2013, pp. 1–5.