Applied GIS

ISSN: 1832-5505

Vol-10 Issue-01 Jan 2022

Tolerable False Data Injection Attacks in Smart Grids: Detection through Extended Distributed State Estimation

M GIRIBABU[1] , VANTERU NARAYNA REDDY[2]

Associate professor[1,2]

DEPARTMENT OF ELECTRICAL AND ELECTRONICS ENGINEERING

P.B.R.VISVODAYA INSTITUTE OF TECHNOLOGY & SCIENCE

S.P.S.R NELLORE DIST, A.P , INDIA , KAVALI-524201

Abstract:

One of the deadliest cyber-attacks against smart grids is known as false data injection (FDI), which may result in energy theft from end users, false dispatch in the distribution process, and device malfunction during power production. In this work, we develop a new kind of FDI attack called tolerant false data injection (TFDI). These assaults circumvent the conventional method of detecting incorrect data by taking use of the detector's latitude for dealing with observational mistakes. After that, we present an EDSE-based approach to TFDI detection in smart grids. Using graph partition methods, the smart grid is broken down into its constituent parts. The extended subsystem is the result of each subsystem being expanded to include the neighbouring buses and tie lines. Each expanded subsystem's bogus data is checked using the Chi-squares test. Decomposition makes the bogus data easily distinguishable from common observational mistakes, hence improving the sensitivity of the detection process. IEEE 14-bus, IEEE 39-bus, IEEE 118-bus, and IEEE 300-bus systems are subjected to extensive TFDI attack simulations. The suggested technique greatly decreases the related computing costs, while simulation results demonstrate that the detection accuracy of the EDSE-based method is much greater than that of the previous method. Keywords: smart grids; security; false data injection (FDI); bad data detection; extended distributed state estimation (EDSE)

Introduction

In smart grids, information techniques are applied to provide a desirable infrastructure for real-time Quantification, Dissemination, Determination, and Management. Millions of buildings and streets are outfitted with sensors for this reason. Due to its interconnection with the data infrastructure, the question of how to prevent false data injection (FDI) attacks — those that manipulate data in transmissions or acquire unauthorized access and control over electrical systems — arises. In addition, hackers are drawn to FDI assaults because of the potential financial rewards they provide (hackers, for instance, may alter their energy expenditures by tampering with the readings on their smart meters). The control center might be led astray by the phony information and endanger the smart grid in the process. Since the ground-breaking work of Schweppe in 1970 [2], it has been widely accepted that power system state estimate (SE) is a suitable way to analyse the poor data. Processing the set of redundant measurements, often bus voltage magnitudes and phase angles, in real-time is used in supervisory control and data acquisition (SCADA) systems to decrease

observation errors, identify incorrect data, and predict the electrical states of power systems. It is hypothesized that faulty data detection techniques [3] may safeguard smart grids against FDI assaults, such as the energy conservation test, the Chi-squares test, and the normalized residuals test.

Preliminaries

SE

Power system SE is widely used to ensure the safety and economy of operation of power system. The state variables are related to the measurements as shown in Equation (1):

$$z = h(x) + e$$

where x is the state variables; z is the meter measurements;

$$h(x) = [h_1(x_1, x_2, \ldots, x_n), \ldots, h_m(x_1, x_2, \ldots, x_n)]^T,$$

where $h_i(x_1, x_2, \ldots, x_n)$ is a function of $x_1, x_2, \ldots, x_n$; and $e = [e_1, e_2, \ldots, e_m]^T$

is the measurement error. For a well-proofreading system, these errors can be considered to follow the Gaussian distribution of zero mean [3]. In the SE, measurements are usually the values that can be observed easily, such as the line power flow, bus power injections, bus voltage magnitudes, and line current flow magnitudes, etc. The state variables are usually complex phasor voltages which cannot be measured conveniently. Both the measurements and state variables follow the same constraints, such as power balance theory and the Kirchhoff's Law, etc. When using the polar coordinates for a system containing N buses, the state vector will contain (2N − 1) elements, N bus voltage

ISSN: 1832-5505
Vol-10 Issue-01 Jan 2022

magnitudes and (N − 1) phase angles. In general, measurements are more than state variables (m > n), since there are more lines than buses and more kinds of measurements than state variables. Essentially, power system SE is a process which uses real-time redundant measurements to improve data accuracy and automatically excluded from the error message caused by random interference. The objective is to find an estimate x ^ of x that is the best fit of the measurement z according to Equation (1). The problem is usually solved by the weighted least squares (WLS) algorithm [3]. The SE can be formulated as a quadratic optimization problem:

$$\min_x J(x) = \min_x [z - h(x)]^T R^{-1} [z - h(x)]$$

where R−1 is the measurement inverse covariance matrix. The Newton's method can be applied to solve the quadratic optimization problem. The increment can be calculated by:

$$\Delta x^{(k)} = G(x^{(k)})^{-1} H^T(x^{(k)}) \cdot R^{-1} \cdot [z - h(x^{(k)})]$$

$$\text{where } H(x^{(k)}) = \left. \frac{\partial h(x)}{\partial (x)} \right|_{x=x^{(k)}}$$

is the Jacobi matrix; and

$$G(x^{(k)}) = H^T(x^{(k)})\ R^{-1}\ H(x^{(k)})$$

is the gain matrix. The convergence criterion is the following:

$$\max(|\Delta x^k|) < \varepsilon_x$$

where εx is a predefined threshold.

TFDI

Most researches on the FDI construction follow the same idea: the attackers find an attack vector, a, to be equal to Hc. Then the manipulated measurement za = z + a can pass the bad data detection and identification of direct-current (DC) SE [8,9]. Thus, the measurement residual is:

$$\begin{aligned} \left\| z_a - H\hat{x}_a \right\| &= \left\| z_a - H\left(\left(H^T R^{-1} H \right)^{-1} H^T R^{-1} \right) z_a \right\| \\ &= \left\| z + a - H\left(\hat{x} + c \right) \right\| \\ &= \left\| z - H\hat{x} + (a - Hc) \right\| = \left\| z - H\hat{x} \right\| \\ &\text{when} \quad a = Hc \end{aligned}$$

From the perspective of the attacker, it is almost an unattainable mission to find an attack vector a in the real world. Firstly, the topology of the power system is one of the top secrets of most power companies. It is difficult to obtain the measurement matrix H. Secondly, solving the a = Hc, which in real systems is an ultra-high dimensional equation is difficult. It would be a NP-hard problem, when the attackers want to inject a specific data with limited compromised meters. Moreover, if the system topology is changed, the FDI attack would trigger bad data detection. Subject to the constraints of invisible observation errors and the false alert rate, the tolerance mechanism for measurement errors in SE is necessary. Instead of solving the problem in Equation (6), the attacker can construct a TFDI below the threshold of estimated residuals:

$$z_a = z + a = h(x) + e + a$$

$$s.t.\ J(\hat{x}) = \sum_{i=1}^{m} \frac{(z_i - h_i(\hat{x}))^2}{\sigma_i^2} < \chi^2_{(m-n),p}$$

Moreover, there is a high probability that the false data could not be detected when the attackers manipulate the data on both sides of the same transmission line. There are four power flow measurements per line. In each direction, there is a pair of active powers and reactive powers. Since the active power is related to economic interests, it is more attractive for attackers to falsify. On the transmission line Li,j (between the bus i and j), Pi,j denotes the active power from busi to busj, observed on busi, and Pj,i denotes the active power from busi to busj, observed on busj. The original active power from bus i to j , org Pi j and , org Pj i are changed by same times to be , inj Pi j and , inj Pj i simultaneously to guarantee the balance of line power flow. Injected data levels (IDL) is defended to present the relative injected errors against the measurements:

$$IDL = \frac{P_{i,j}^{inj} - P_{i,j}^{org}}{P_{i,j}^{org}} \times 100\%$$

Comparing with the strict conditions required by the undetectable FDI attack, the TFDI only needs the attacker to manipulate meters on target transmission lines. Moreover, from [8], it can be seen that the probability of finding an attack vector for a target FDI (unconstrained case) in an IEEE 300-bus system is about 20%, even if the attacker can compromise 60% of all smart meters. In experiments, traversal attacks are conducted in IEEE 57- and 300-bus systems. The probabilities to construct a TFDI are shown in Table 1. It can be

ISSN: 1832-5505
Vol-10 Issue-01 Jan 2022

seen that the possibility to construct a TFDI attack is much higher than for an undetectable FDI.

Table 1. Success probability to find a tolerable false data injection (TFDI) attack. IDL: injected data levels; and IEEE: the Institute of Electrical and Electronics Engineers.

System	Success probability with different IDL (%)							
	25%	50%	75%	100%	125%	150%	175%	200%
IEEE 57	67	60	56	52	48	44	37	31
IEEE 300	72	66	61	58	55	52	47	41

In addition, we modify the active power on each bus in IEEE 39-, 57- and 118-bus systems with different IDL. A relative low detection precision is performed by the Chi-squares test, as shown in Table 2. Furthermore, with the scale of the power system grows, the tolerance of measurement errors is accordingly increased. We can see from Table 2 that it is easier for the attackers to bypass the detection in the larger system.

Table 2. Detection precision of the Chi-squares test against TFDI attacks.

IEEE 39-bus system		IEEE 57-bus system		IEEE 118-bus system	
IDL	Detection precision	IDL	Detection precision	IDL	Detection precision
10%	67%	120%	51%	150%	75%
20%	76%	150%	56%	200%	82%
30%	89%	200%	69%	250%	86%
40%	96%	300%	76%	300%	88%
50%	100%	400%	79%	350%	93%
60%	100%	500%	85%	400%	94%

It's important to remember that attackers build the TFDI based on the information and access they have to smart meters. The system's observability is of little concern to them. TFDI assaults conceal themselves amid regular measurement mistakes and take advantage of the detector's tolerance of typical cumulative random noises. It just causes certain smart meter readings to be inaccurate and does not change the system's observability overall. Since the TFDI method is simple to implement and works with both AC and DC models, it is important for power engineers and security experts to be aware of the threat posed by this kind of attack. In this article, we'll go through several ways to defend yourself against this kind of assault.

Possible Dangers and Attack Scenarios

Invasion of Smart Meters

The FDI is based on cyber methods. To get access to invalid activities on smart meters or network communications is the primary goal of cyberattacks. Modbus/TCP and DNP 3.0/TCP are the most used protocols for communicating with smart meters. Modbus/TCP uses port 502, whereas DNP3.0/TCP uses port 20,000. An adversary may begin by scanning the whole network segment for hosts with open ports (either 502 or 20,000). After that, unique hosts are identified and labeled as potentially malicious. The attack may then interact with these gadgets to confirm that they are smart meters and to learn what kind of goods they are. Smart meters may be hacked in two ways: (1) Hacking into a gadget has traditionally included deciphering passwords. Smart meters often demand authentication when their settings need to be changed. Smart meters don't have complicated password procedures because of the limited computing resources and storage. In this simulated attack on smart meters, the password consists of four digits, and it can be cracked in a matter of seconds. (2) Plaintext communication is another weakness that may be used to get access to smart meters. Password protection systems for certain smart meters may be somewhat involved. However, Modbus/TCP or DNP 3.0/TCP is often used as the communication protocol for smart meters since it allows for unencrypted data transmission. Critical activities on smart meters, such as changes to system time, IP addresses, and firmware upgrades, need authentication, and an attacker may discover these events by monitoring the data flow. Attackers may get access to smart meters if they are able to locate the package containing authentication information and capture the password. If an attacker gains access to a system, they may alter the values of any measurements they take. Most smart meters just allow you to read off numbers like active power and reactive power. Some parameters, however, are editable, including time and CT ratio. Alternating electric currents may be measured with the use of a CT. Where I1 and I2 are the main and secondary currents, respectively, the CT ratio K is denoted as K = I1/I2. Changes in K have an equivalent effect on the active and reactive power levels. Altering the CT ratio allows the attacker to skew power consumption readings.

Identifying Erroneous Data Using EDSE

Section 2.2 demonstrates that the Chi-squares test has a threshold that allows for random and unavoidable fluctuations in the data. Attackers may develop sophisticated TFDI assaults by masking their signals with background sounds in measurements. The accumulated normal observation errors from each measurement become

ISSN: 1832-5505
Vol-10 Issue-01 Jan 2022

more problematic for the Chi-square test as the number of measurements increases. False data won't be able to hide amongst the noise of regular measurements if the big system can be properly dissected. To deal with TFDI attacks, an EDSE-based faulty data detection approach is presented.

Separation of the Power Grid

The weighted-undirected graph model of power systems may be built for a smart grid with n buses and m transmission lines by writing G = V, E, where V is a collection of vertices representing load buses or generators and E is a set of edges representing the transmission lines in smart grids. Graph adjacency matrix is represented by A = ai,j, where i,j = 1, 2,..., n. When buses i and j are physically linked, their physical attributes are reflected in the ai,j element, which is non-zero in this case.

The branch's importance in the modeled graph may be calculated in the following ways:

transmission line impedance; line power flow at each sample period; the fundamental architecture of the power system (ai,j = 1 if bus i and bus j are linked). Transmission line impedance (Z = R + jX), which represents the electrical distance between buses, is used as edge weight in this research. The transmission line's reactance, X, is equal to R, its resistance. When compared to X, R's value is negligible. Therefore, the edge weight is decided upon as the absolute value of the line reactance |X|. Using clustering techniques like the L-bounded Graph Partition Method (LGPM) [25], the K-Medoid [26], Chameleon [27], etc., the massive graph is partitioned into multiple smaller subgraphs. In this study, we use the LGPM approach, which is stable and independent of the initial clustering centers, to decompose graphs. Table 3 shows the essential steps in LGPM's method.

Analysis and Experiment

Section 6.1 simulates three assault scenarios on the IEEE 14-bus system to examine the efficacy of the EDSE-based approach. Section 6.2 uses the IEEE-39 bus system to present a statistical comparison of detection performances between the traditional and EDSE-based methods; Section 6.3 discusses some TFDI attacks that are not detected by the EDSE-based method; Section 6.4 demonstrates the evaluation of time complexity; and Section 6.5 addresses the appropriate number of subsystems. Cases of Attack on IEEE 14-Bus Systems, Version 6.1 In Figure 3, we see three different attack scenarios built on the IEEE 14-bus architecture. The LGPM is responsible for breaking down the IEEE 14-bus architecture.

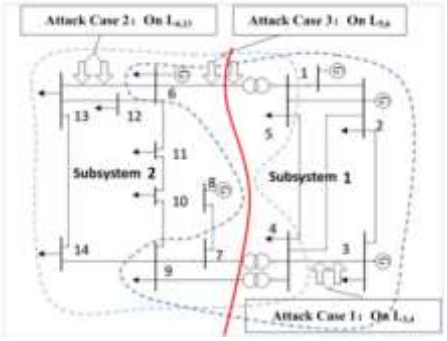

Figure 1. Attack cases on IEEE 14-bus system.

As shown in Table 4, the IEEE 14-bus system is divided into two subsystems, “subsys_1” and “subsys_-2 “

$$L_{5,6},\ L_{4,7},\ L_{4,9}$$

are tie lines. In subsys_1, there are 8 buses including adjacent buses: bus6,bus7 and bus9. The number of state variables n1 is 15 and the number of measurements m1 is 40. The degree of freedom o1 in this subsystem is m1 − n1 = 25. According to the property of distribution, the threshold of bad data suspicion is 43.77. In subsys_2, there are 11 buses, 21 state variables, and 52 measurements, and the threshold is 51.00. This indicates that the local threshold is much lower than the global one.

Table 4. Decomposition of IEEE 14-bus system.

System	Bus	n	m	o	T_{sp}
Subsys_1	1,2,3,4,5,6,7,9	15	40	25	43.77
Subsys_2	4,5,6,7,8,9,10,11,12,13,14	21	52	31	51.00
Global	All	27	80	53	72.15

To test the performance of EDSE-based bad data detection, three attack cases are constructed as shown in Table 5. In Table 5, Li,j denotes the transmission line where the false data are injected. Pi,j denotes the active power from busi to busj, observed on busi. The active power Pi,j and Pj,i are modified at the same time to guarantee the balance of line power flow. The original measurements are simulated by MATPOWER and then the Gaussian noise is added. It should be noted that there is a tiny difference between Pi j , and Pj i, . These two active power measurements are observed at each end of the transmission line. There is some power loss on the transmission line. For an attacker, it is not easy to change the active power to arbitrary values, because active power is usually read-only.

ISSN: 1832-5505
Vol-10 Issue-01 Jan 2022

As explained in Section 4.1, attackers can change the active power through falsifying the CT ratio. In Attack Case 1–3, they increase the CT ratio by 2 times, 3 times and 1.5 times, respectively. In Attack Case 1, false data is only injected into subsys_1. The P4,5 is modified from −61.16 MW to −122.32 MW and P5,4 is modified from 61.67 MW to 122.34 MW. In Attack Case 2, false data is only injected into subsys_2. The P6,13 is modified from 17.75 MW to 53.24 MW, and P13,6 is modified from −17.54 MW to 52.61 MW. In Attack Case 3, the false data is injected into the tie line between subsys_1 and subsys_2. The P5,6 is modified from 66.13 MW to 99.20 MW, and P6,5 is modified from −66.13 MW to 99.20 MW.

Table 5. TFDI attack cases on IEEE 14-bus system.

Attack Case	Modified measurement	P_{ij} (MW)		P_{ji} (MW)	
		Original value	Injected value	Original value	Injected value
Attack Case 1	$L_{4,5}$	−61.16	−122.32	61.67	122.34
Attack Case 2	$L_{6,13}$	17.75	53.24	−17.54	−52.61
Attack Case 3	$L_{5,6}$	66.13	99.20	−66.13	−99.20

As shown in Table 6, global values of J(x ^) are 54.91, 66.04 and 54.73 in three attack cases, respectively. Obviously, they are lower than the threshold To,p(72.15). Thus, the injected false data cannot be detected. When we adopt EDSE-based method to deal with the Attack Case 1, we find that: in subsys_1, the J(x ^) is 51.98, which is higher than the local threshold 1 o p, T (43.77); in subsys_2, the J(x ^) is 25.22, which is below the local threshold 2 o p, T (51.00). It implies that there is false data in subsys_1. Similarly in Attack Case 2, the EDSE-based method can detect the false data in subsys_2. In Attack

Case 3, false data is detected in subsys_2. If the subsystem is not extended to include the adjacent buses, the FDI on tie-line L5,6 will not be found.

Table 6. Detection results on IEEE 14-bus system.

Attack Case	Global		Subsys_1		Subsys_2	
	$T_{o,p}$	$J(x)$	$T_{o,p}$	$J(x)$	$T_{o,p}$	$J(x)$
Attack Case 1	72.15	54.91	43.77	51.98	51.00	25.22
Attack Case 2		66.04		13.11		59.05
Attack Case 3		54.73		19.26		53.48

Conclusions

In order to demonstrate how hackers might manipulate data in smart grids and avoid the conventional bad data detection techniques in power systems, this article presents many TFDI attack examples. These assaults conceal in regular observational flaws, which are within the margin of error for the Chi-square test. Potential losses from energy theft and cracking economic dispatch on the IEEE 14-bus system are calculated, and the implications of such assaults on smart grids are discussed. To address this issue, we offer an EDSE-based approach for identifying TFDI assaults. By breaking down a large system into smaller, more manageable pieces, this technique increases the sensitivity of faulty data detection. Decomposing the power grid into manageable chunks using clustering techniques, expanding each subsystem to encompass neighboring buses, and performing SE and bad data detection inside each chunk are the three main processes that make up the EDSE-based approach. The IEEE 14, 39, 118, and 300-bus systems are simulated through comprehensive TFDI assault scenarios. The results demonstrate a dramatic increase in the detection accuracy of the EDSE-based technique. In addition, the EDSE provides a novel approach to online bad data identification due to its substantially decreased computing complexity and the possibility of further speeding up the detection process through parallel analysis of all extended subsystems. A better response to FDI in smart grids is the cyber-physical fusion method, since this kind of assault produces interactive responses in both the cyber network and the electricity grid. In the future, we want to go further into a detection approach that combines the EDSE with traffic flow anomaly detection. Even if the EDSE misses the bad data, the communication network's alarms will go off if criminals get unauthorized access to smart meters.

References

[1]. Bou-Harb, E.; Fachkha, C.; Pourzandi, M.; Debbabi, M.; Assi, C. Communication security for smart grid distribution networks. IEEE Commun. Mag. 2013, 51, 42–49.

[2]. Schweppe, F.C.; Wildes, J. Power system static-state estimation, Part I: Exact model. IEEE Trans. Power Appar. Syst. 1970, PAS-89, 120–125.

[3]. Abur, A.; Exposito, A.G. Power System State Estimation: Theory and Implementation; CRC Press: Boca Raton, FL, USA, 2004; pp. 44–80.

[4]. Hug, G.; Giampapa, J.A. Vulnerability Assessment of AC state estimation with respect to false data injection cyber-attacks. IEEE Trans. Smart Grid 2012, 3, 1362–1370.

[5]. Kosut, O.; Liyan, J.; Thomas, R.J.; Lang, T. Malicious data attacks on the smart grid. IEEE Trans. Smart Grid 2011, 2, 645–658.

[6]. Giani, A.; Bitar, E.; Garcia, M.; McQueen, M.; Khargonekar, P.; Poolla, K. Smart grid data integrity attacks. IEEE Trans. Smart Grid 2013, 4, 1244–1253.

[7]. Xie, L.; Mo, Y.; Sinopoli, B. Integrity data attacks in power market operations. IEEE Trans. Smart Grid 2011, 2, 659–666.

ISSN: 1832-5505
Vol-10 Issue-01 Jan 2022

[8]. Liu, Y.; Ning, P.; Reiter, M.K. False data injection attacks against state estimation in electric power grids. ACM Trans. Inf. Syst. Secur. (TISSEC) 2011, 14, 21–32.

[9]. Huang, Y.; Esmalifalak, M.; Nguyen, H.; Zheng, R.; Han, Z.; Li, H.; Song, L. Bad data injection in smart grid: Attack and defense mechanisms. IEEE Commun. Mag. 2013, 51, 27–33.

[10]. Yang, Q.; Yang, J.; Yu, W.; An, D.; Zhang, N.; Zhao, W. On false data injection attacks against power system state estimation: modeling and countermeasures. IEEE Trans. Parallel Distrib. Syst. 2013, 25, 717–729.

[11]. Bobba, R.B.; Rogers, K.M.; Wang, Q.; Khurana, H.; Nahrstedt, K.; Overbye, T.J. Detecting False Data Injection Attacks on DC State Estimation. In Proceedings of the Preprints of the First Workshop on Secure Control Systems, CPSWeek, Stockholm, Sweden, 12 April 2010.

[12]. Dán, G.; Sandberg, H. Stealth Attacks and Protection Schemes for State Estimators in Power Systems. In Proceedings of the First IEEE International Conference on Smart Grid Communications (SmartGridComm), Gaithersburg, MD, USA, 4–6 October 2010; pp. 214–219.

[13]. Vukovic, O.; Kin, C.S.; Dan, G.; Sandberg, H. Network-aware mitigation of data integrity attacks on power system state estimation. IEEE J. Sel. Areas Commun. 2012, 30, 1108–1118.

[14]. Kim, T.T.; Poor, H.V. Strategic protection against data injection attacks on power grids. IEEE Trans. Smart Grid 2011, 2, 326–333.

[15]. Pasqualetti, F.; Dorfler, F.; Bullo, F. Cyber-Physical Attacks in Power Networks: Models, Fundamental Limitations and Monitor Design. In Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), Orlando, FL, USA, 12–15 December 2011; pp. 2195–2201.

[16]. Shuguang, C.; Zhu, H.; Kar, S.; Kim, T.T.; Poor, H.V.; Tajer, A. Coordinated data-injection attack and detection in the smart grid: A detailed look at enriching detection solutions. IEEE Signal Process. Mag. 2012, 29, 106–115.

[17]. Zonouz, S.; Rogers, K.M.; Berthier, R.; Bobba, R.B.; Sanders, W.H.; Overbye, T.J. SCPSE: Security-oriented cyber-physical state estimation for power grid critical infrastructures. IEEE Trans. Smart Grid 2012, 3, 1790–1799.

[18]. Choi, D.-H.; Xie, L. Fully Distributed Bad Data Processing for Wide Area State Estimation. In Proceedings of the IEEE International Conference on Smart Grid Communications (SmartGridComm), Brussels, Belgium, 17–20 October 2011; pp. 546–551.

[19]. Xie, L.; Dae-Hyun, C.; Kar, S.; Poor, H.V. Fully distributed state estimation for wide-area monitoring systems. IEEE Trans. Smart Grid 2012, 3, 1154–1169.

[20]. Electric Rates. Availiable online: http://www.pge.com/nots/rates/tariffs/electric.shtml#RESELEC (accessed on 30 October 2013).