49th IEEE Conference on Decision and Control
December 15-17, 2010
Hilton Atlanta Hotel, Atlanta, GA, USA

Cyber Security Analysis of State Estimators in Electric Power Systems

André Teixeira*, Saurabh Amin†, Henrik Sandberg*, Karl H. Johansson*, and Shankar S. Sastry†

Abstract— **In this paper, we analyze the cyber security of state estimators in Supervisory Control and Data Acquisition (SCADA) systems operating in power grids. Safe and reliable operation of these critical infrastructure systems is a major concern in our society. In current state estimation algorithms there are bad data detection (BDD) schemes to detect random outliers in the measurement data. Such schemes are based on high measurement redundancy. Although such methods may detect a set of very basic cyber attacks, they may fail in the presence of a more intelligent attacker. We explore the latter by considering scenarios where deception attacks are performed, sending false information to the control center. Similar attacks have been studied before for linear state estimators, assuming the attacker has perfect model knowledge. Here we instead assume the attacker only possesses a perturbed model. Such a model may correspond to a partial model of the true system, or even an out-dated model. We characterize the attacker by a set of objectives, and propose policies to synthesize stealthy deceptions attacks, both in the case of linear and nonlinear estimators. We show that the more accurate model the attacker has access to, the larger deception attack he can perform undetected. Specifically, we quantify trade-offs between model accuracy and possible attack impact for different BDD schemes. The developed tools can be used to further strengthen and protect the critical state-estimation component in SCADA systems.**

I. Introduction

Several infrastructures are of major importance to our society. Examples include the power grid, telecommunication network, and water supply, and due to how essential they are in our daily life they are referred to as critical infrastructures. These systems are operated by means of complex distributed software systems, which transmit information through wide and local area networks. Because of this fact, critical infrastructures are vulnerable to cyber attacks [1], [2]. These are performed on the information residing and flowing in the IT system.

Power networks, for instance, are operated through SCADA systems complemented by a set of application specific software, usually called energy management systems (EMS). Modern EMS provide information support for a variety of applications related to power network monitoring and control. The power system state estimator (PSSE) is an on-line application which uses redundant measurements and a network model to provide the EMS with an accurate state estimate at all times. The PSSE has become an integral tool for EMS, for instance for contingency-constrained optimal power flow. The PSSE also provides important information to pricing algorithms. SCADA systems collect data from remote terminal units (RTUs) installed in various substations, and relay aggregated measurements to the central master station located at the control center. Several cyber attacks on SCADA systems operating power networks have been reported [3], [4], and major blackouts, as the August 2003 Northeast blackout, are worsened by the misuse of the SCADA systems [5]. The 2003 blackout also highlighted the need of robust state estimators that converge accurately and rapidly in such extreme situations, so that necessary preventive actions can be taken in a timely manner. As discussed in [1], there are several vulnerabilities in the SCADA system architecture, including the direct tampering of RTUs, communication links from RTUs to the control center, and the IT software and databases in the control center. For instance, the RTUs could be targets of denial-of-service (DoS) or deceptions attacks injecting false data [6].

Power networks, being systems where control loops are closed over communication networks, represent an important class of networked control systems (NCS). Unlike other IT systems where cyber security mainly involves encryption and protection of data, here cyber attacks may influence the physical processes through the digital controllers. Therefore focusing on encryption of data alone may not be enough to guarantee the security of the overall system, especially its physical component. In order to increase the resilience of these systems, one needs appropriate tools to first understand and then to protect NCS against cyber attacks. Some of the literature has already tackled these problems such as false data injection in power system state estimation [6], security constrained control [7], and replay attacks [8].

Our work analyzes the cyber security of the PSSE in the SCADA system. In current implementations of PSSE algorithms there are bad data detection (BDD) schemes [9], [10] designed to detect random outliers in the measurement data. Such schemes are based on high measurement redundancy and are performed at the end of the state estimation process. Although such methods can detect basic attacks, they may fail in the presence of more intelligent attackers that wish to stay undetected, in which case the false data could be introduced in a coordinated manner so that it looks consistent to the detection mechanism, thus bypassing it. We explore the latter by considering scenarios where deception attacks are performed by sending false information to the

This work was supported in part by the European Commission through the VIKING project, the Swedish Research Council, the Swedish Foundation for Strategic Research, and the Knut and Alice Wallenberg Foundation.

H. Sandberg, A. Teixeira, and K. H. Johansson are with the Automatic Control Lab, Royal Institute of Technology, Stockholm, Sweden. {andretei,hsan,kallej}@ee.kth.se

S. Amin and S. S. Sastry are with the TRUST Center, University of California, Berkeley. {saurabh,sastry}@eecs.berkeley.edu

control center. A related study was performed in [6] for linear state estimators, assuming the attacker has perfect model knowledge. Here we instead assume the attacker only possesses a perturbed model. Such a model may correspond to a partial model of the true system, or an out-dated model. We characterize the attacker by defining a set of objectives, and propose policies to synthesize stealthy deceptions attacks, both for linear and nonlinear estimators. We show that the more accurate model the attacker has access to, the larger deception attack he can perform undetected. Specifically, we quantify trade-offs between model accuracy and possible attack impact for different BDD schemes.

The outline of this paper is as follows. We present the main concepts behind state estimation in power systems, the attacker model, and problem formulation in Section II. The properties of the estimation algorithm which are deployed in practice are discussed in Section III. In Section IV, two common BDD methods are reviewed. The analysis of stealthy deception attacks with partial knowledge is performed in Section V. An example that illustrates the results is presented in Section VI, followed by the conclusions in Section VII.

II. Stealthy Deception Attacks on PSSE

We focus on additive deception attacks aimed toward manipulating the measurements to be processed by the PSSE in such a manner that the resulting systematic errors introduced by the adversary are either undetected or only partially detected by a BDD method. We call such attacks *stealthy deception attacks* on the PSSE. We are also interested in finding the class of stealthy deception attacks that do not pose significant convergence issues for the estimator. Attacks affecting the convergence of the PSSE are related to *data availability*, as they can be seen as DoS attacks. However the focus of this work is on deception attacks, which are related to *data integrity*. Note that the non-convergence of the PSSE without any attack can have several reasons, such as low measurement redundancy and topology and parameter errors. Since this is not related to the security of the PSSE, we assume the estimator converges if no attack is performed.

A. PSSE

The basic PSSE problem is to find the best n-dimensional state x for the measurement model

$$z = h(x) + \epsilon, \tag{1}$$

in a weighted least square (WLS) sense. Here z is the m-dimensional vector of measurements, h is a nonlinear function modeling the power network, and $\epsilon \sim \mathcal{N}(0, R)$ is a vector of independent zero-mean Gaussian variables with covariance matrix $R = \text{diag}(\sigma_1^2, \ldots, \sigma_m^2)$. For an electric power network with N buses, the state vector $x = (\theta^\top, V^\top)^\top$, where $V = (V_1, \ldots, V_N)^\top$ is the vector of bus voltage magnitudes and $\theta = (\theta_2, \ldots, \theta_N)^\top$ the vector of phase angles. Without loss of generality, bus 1 is considered as the reference bus with $\theta_1 = 0$, so the state dimension is $n = 2N - 1$. Detailed formulae relating measurements z and state x may be found in [11].

Defining the residual vector $r(x) = z - h(x)$, we can write the WLS problem as

$$\min_{x \in \mathbb{R}^n} J(x) = \frac{1}{2} r(x)^\top R^{-1} r(x).$$

The PSSE yields a *state estimate* $\hat{x}$ as a minimizer to this minimization problem. The *measurement estimates* are defined as $\hat{z} := h(\hat{x})$. The WLS estimate $\hat{x}$ satisfies the following first order necessary condition for optimality

$$F(\hat{x}) := \nabla J(\hat{x}) = -H^\top(\hat{x}) R^{-1} r(\hat{x}) = 0, \tag{2}$$

where $H = dh/dx$ is the $m \times n$ dimensional measurement Jacobian matrix. The solution $\hat{x}$ of the nonlinear equation $F(\hat{x}) = 0$ may be obtained by the *Newton method* in which a linear equation is solved at each iteration to compute the correction $\Delta x^k := x^{k+1} - x^k$:

$$[F'(x^k)](\Delta x^k) = -F(x^k), \quad k = 0, 1, \ldots, \tag{3}$$

where the Hessian matrix $[F'(x^k)] = \nabla^2 J(x^k)$ is given by

$$[F'(x^k)] = H^\top(x^k) R^{-1} H(x^k) + \sum_{i=1}^{m} \frac{r_i(x^k)}{\sigma_i^2} \nabla^2 r_i(x^k).$$

The iterates (3) guarantee the convergence to a local minimum as long as the generated sequence $\{x^k\}$ converges and the matrices $[F'(x^k)]$ remain non-singular during the iteration process. A nearly singular Hessian matrix $[F'(x^k)]$ can result in a convergence failure. A precise statement of local convergence is presented in the Appendix.

The second order information in $[F'(x^k)]$ is computationally expensive, and its effect often negligible when applied to PSSE. Thus, the symmetric approximation is used in practice

$$[F'(x^k)] \approx H^\top(x^k) R^{-1} H(x^k) =: K^k$$

where K^k is called the *gain* (or information) matrix. This approximation leads to the *Gauss-Newton* steps obtained by solving the so called *normal equations*:

$$\left(H^\top(x^k) R^{-1} H(x^k)\right)(\Delta x^k) = H^\top(x^k) R^{-1} r(x^k), \tag{4}$$

for $k = 0, 1, \ldots$ For an observable power network, the measurement Jacobian matrix $H(x^k)$ is full column rank. Consequently, the gain matrix $K^k = \sum_{i=1}^{m} \frac{H_i^\top(x^k) H_i(x^k)}{\sigma_i^2}$ in (4) is positive definite and the Gauss-Newton step generates a descent direction, *i.e.*, for the direction $\Delta x^k = x^{k+1} - x^k$ the condition $\nabla J(x^k)^\top \Delta x^k < 0$ is satisfied. We now present the attacker model.

B. Attacker Model

The goal of a stealthy deception attacker is to compromise the telemetered measurements available to the PSSE such that: 1) The PSSE algorithm converges; 2) For the targeted set of measurements, the estimated values at convergence are close to the compromised ones introduced by the attacker; and 3) The attack remains fully undetected by the BDD scheme.

As a consequence of the attacker's stealthy action, the incorrect state estimates generated by the PSSE can have

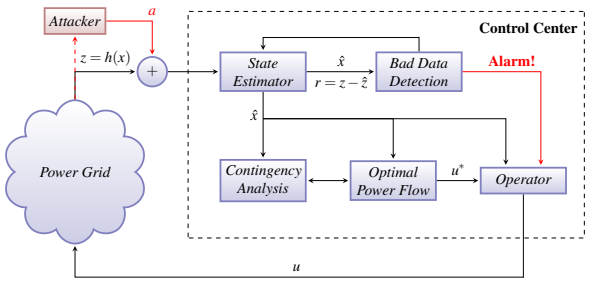

Fig. 1. The state estimator under a cyber attack

different effects on other power management functions. In fact, as depicted in Figure 1, the state estimate is used as an input to other software applications, in particular the contingency analysis and optimal power flow.

Let the corrupted measurement be denoted z^a. We assume the following additive attack model

$$z^a = z + a, \tag{5}$$

where $a \in \mathbb{R}^m$ is the attack vector introduced by the attacker. The vector a has zero entries for uncompromised measurements. Under attack, the normal equations (4), give the estimates

$$\tilde{x}^{k+1} = \tilde{x}^k + \left(H^\top(\tilde{x}^k)R^{-1}H(\tilde{x}^k)\right)^{-1} H^\top(\tilde{x}^k)R^{-1}r^a(\tilde{x}^k),$$

for $k = 0, 1, \ldots$, where $\tilde{x}^k$ is the *biased* estimate at iterate i, and $r^a(\tilde{x}^k) := z^a - h(\tilde{x}^k)$. If the local convergence conditions hold, then these iterations converge to $\hat{x}^a$, which is the biased state estimate resulting from the use of z^a. Thus, the convergence behavior can be expressed as the following statement:

1) The sequence $\{\tilde{x}^0, \tilde{x}^1, \ldots\}$ generated by the mapping $G(x) = x + (H^\top(x)R^{-1}H(x))^{-1}H^\top(x)R^{-1}r^a(x)$ converges to a fixed point $\hat{x}^a$ of G in a region $\mathcal{S}^a_\vartheta$,

where $\mathcal{S}^a_\vartheta$ is a closed ball in $\mathbb{R}^n$ of radius ϑ governed by the conditions required for the local convergence to hold. We will occasionally use the notation $\hat{x}^a(z^a)$ to emphasize the dependence on z^a.

The BDD schemes for PSEE are based on checking if the weighted p-norm of the measurement residual is below some threshold τ, which is selected based on permissible false-alarm rate. Thus, the attackers action will be undetected by the BDD scheme provided that the following condition holds:

2) The measurement residual under attack $r^a := r(\hat{x}^a) = z^a - h(\hat{x}^a)$, satisfies the condition $\|Wr(\hat{x}^a)\|_p < \tau$.

Finally, let the target set be represented by $\mathcal{I}_{tgrt}$ containing indices of the measurements which are targeted by the attacker. For each $i \in \mathcal{I}_{tgrt}$, the attacker would like the estimated measurement $\hat{z}^a_i := h_i(\hat{x}^a(z^a))$ to be equal to the actual corrupted measurement z^a_i. However, such a condition may not be satisfied since corrupted measurements may not be consistent with the model, and can result in violation of conditions 1), and 2) mentioned above. Therefore, we arrive at the following condition which will additionally govern the synthesis of attack vector a:

3) The attack vector a is chosen such that $|z^a_i - \hat{z}^a_i| < \eta$ for $i \in \mathcal{I}_{tgrt}$, where η is a small positive constant.

The aim of a stealthy deception attacker is then to find and apply an attack a that satisfies conditions 1), 2), and 3). In Section V, we take a similar approach as in [6] to synthesize stealthy attack policies of the form of $a = \tilde{H}c$, where $\tilde{H}$ is the imperfect model known by the attacker. Unlike in [6], we do not assume the attacker has the exact model of the system and we consider both linear and nonlinear estimators.

III. PSSE Iterates as Linear WLS Problems

As seen in the previous section, solving the normal equation is the corner stone of the estimation algorithm. In this section we take a closer look on the normal equation and show that it can be seen as the solution for a linear least squares problem. This is quite useful as it provides a unified interpretation of the residual for both the linear and nonlinear estimation algorithms.

The normal equation can be interpreted as the solution of a linear least squares problem. In particular, writing $H(x^k)$ as H, and Δx^k as Δx, and $r(x^k) = z - h(x^k)$ as Δz for notational convenience, and defining $\Delta\bar{z} = R^{-1/2}\Delta z$ and $\bar{H} = R^{-1/2}H$, the $k-$th iteration as given by equation (4) is the solution of the linear least squares problem

$$\min_{\Delta x}(\Delta\bar{z} - \bar{H}\Delta x)^\top(\Delta\bar{z} - \bar{H}\Delta x).$$

It can be obtained as a solution of the overdetermined system of equations

$$\bar{H}\Delta x \cong \Delta\bar{z}. \tag{6}$$

Given that $\bar{H}$ has full column rank and using the notation of the pseudo-inverse $\bar{H}^\dagger := (\bar{H}^\top\bar{H})^{-1}\bar{H}^\top$,

$$\Delta x = \bar{H}^\dagger\Delta\bar{z} = (\bar{H}^\top\bar{H})^{-1}\bar{H}^\top\Delta\bar{z}.$$

For the approximate (linear) model

$$\Delta\bar{z} = \bar{H}\Delta\bar{x} + \bar{\epsilon}$$

where $\bar{\epsilon} = R^{-1/2}\epsilon$, the measurement residual can be expressed as

$$\bar{r} = \bar{S}\bar{\epsilon}, \tag{7}$$

where $\bar{S} = (I - \bar{H}(\bar{H}^\top\bar{H})^{-1}\bar{H}^\top)$ is called the weighted sensitivity matrix. Since the matrix $\bar{T} = \bar{H}(\bar{H}^\top\bar{H})^{-1}\bar{H}^\top$ is symmetric and orthogonal with range space $\mathrm{Im}(\bar{H}(\bar{H}^\top\bar{H})^{-1}\bar{H}^\top))$ same as $\mathrm{Im}(\bar{H})$, we call it the *orthogonal projector* onto $\mathrm{Im}(\bar{H})$ and denote it by $\mathcal{P}_{\mathrm{Im}(\bar{H})}$. Such matrix is known as the *hat matrix* in the power system literature [11], [12]. Consequentially, we see that $\bar{S}$ in (7) is the orthogonal projector onto the null-space (kernel) of $\bar{H}^\top$, *i.e.* $\bar{S} = (I - \mathcal{P}_{\mathrm{Im}(\bar{H})}) = \mathcal{P}_{\mathrm{Ker}(\bar{H}^\top)}$.

IV. Bad Data Detection

The measurements used in PSSE may be corrupted by random errors and so a necessary security capability of the PSSE is BDD [11], [12], [10]. Traditionally, the bad data is understood as a result of parameter errors which corrupt the values of modeled circuit elements, incorrect network topology descriptions, and gross measurement errors due to device failures and incorrect meter scans. However, in view of new security threats, bad data can be deliberately introduced by an active adversary which manipulates the communication between remote RTUs and the SCADA system.

Through BDD the PSSE detects measurements corrupted by errors whose statistical properties exceed the presumed standard deviation or mean. This is achieved by hypothesis tests using the statistical properties of the weighted measurement residual (7). We now introduce two of the BDD hypothesis tests widely used in practice, the *performance index test* and the *largest normalized residual test*. These indices are used to model the BDD objective in Section II-B.

1) Performance index test: For the measurement error $\bar{\epsilon} \sim \mathcal{N}(0, I)$, the random variable $y := \sum_{i=1}^{m} \bar{\epsilon_i}^2$ has a chi-square distribution with m degrees of freedom (χ^2_m) with $\mathbb{E}\{y\} = m$. Consider the quadratic cost function evaluated at the optimal estimate $\hat{x}$

$$J(\hat{x}) = \bar{r}^\top \bar{r} = \bar{\epsilon}^\top \bar{S} \bar{\epsilon}. \tag{8}$$

Recalling that $\mathrm{rank}(\bar{H}) = n$, $\mathrm{Im}(\bar{H}) \oplus \mathrm{Ker}(\bar{H}^\top) = \mathbb{R}^m$, and using the definition of orthogonal projector, we note that $\bar{S} = \mathcal{P}_{\mathrm{Ker}(\bar{H}^\top)}$, and we have $\mathrm{rank}(\bar{S}) = m - n$. Therefore, in the absence of bad data, the quadratic form $\bar{\epsilon}^\top \bar{S} \bar{\epsilon}$ has a chi-squares distribution with $m - n$ degrees of freedom, *i.e.* $J(\hat{x}) \sim \chi^2_{m-n}$ with $\mathbb{E}\{J(\hat{x})\} = m - n$. The main idea behind the performance index test is to use $J(\hat{x})$ as an approximation of y and check if $J(\hat{x})$ follows the distribution χ^2_{m-n}. This can be posed as a hypothesis test with a null hypothesis H_0, which if accepted means there is no bad data, and an alternative bad data hypothesis H_1 where

$$H_0 : \mathbb{E}\{J(\hat{x})\} = m - n, \quad H_1 : \mathbb{E}\{J(\hat{x})\} > m - n$$

Defining $\alpha \in [0,\ 1]$ as the significance level of the test corresponding to the false alarm rate, and $\tau_\chi(\alpha)$ such that

$$\int_0^{\tau_\chi(\alpha)} g^\chi(u) du = 1 - \alpha, \tag{9}$$

where $g^\chi(u)$ is the probability distribution function (pdf) of χ^2_{m-n}, and noting that $J(\hat{x}) = \|R^{-1/2} r(\hat{x})\|_2$ the result of the test is

$$\text{reject } H_0 \text{ if } \|R^{-1/2} r\|_2 > \sqrt{\tau_\chi(\alpha)},$$
$$\text{accept } H_0 \text{ if } \|R^{-1/2} r\|_2 \leq \sqrt{\tau_\chi(\alpha)}.$$

2) Largest normalized residual test: From (7), we note that $\bar{r} \sim \mathcal{N}(0, \bar{S})$ and equivalently $r \sim \mathcal{N}(0, \Omega)$ with $\Omega = R^{1/2} \bar{S} R^{1/2}$. Now consider the normalized residual vector

$$r^N = D^{-1/2} r, \tag{10}$$

with $D \in \mathbb{R}^{m \times m}$ being a diagonal matrix defined as $D = \mathrm{diag}(\Omega)$. In the absence of bad data each element $r_i^N,\ i = 1, \ldots, m$ of the normalized residual vector then follows a normal distribution with zero mean and unit variance, *i.e.* $r_i^N \sim \mathcal{N}(0, 1),\ \forall i = 1, \ldots, m$. Thus, bad data could be detected by checking if r_i^N follows $\mathcal{N}(0, 1)$. Posing this as hypothesis test for each element r_i^N

$$H_0 : \mathbb{E}\left\{r_i^N\right\} = 0, \quad H_1 : \mathbb{E}\left\{|r_i^N|)\right\} > 0$$

Again defining $\alpha \in [0,\ 1]$ as the significance level of the test and $\tau_\mathcal{N}$ such that

$$\int_{-\tau_\mathcal{N}(\alpha)}^{\tau_\mathcal{N}(\alpha)} g^\mathcal{N}(u) du = 1 - \alpha, \tag{11}$$

where $g^\mathcal{N}(u)$ is the pdf of $\mathcal{N}(0, 1)$, and noting (10), the result of the test is

$$\text{reject } H_0 \text{ if } \|D^{-1/2} r\|_\infty > \tau_\mathcal{N}(\alpha)$$
$$\text{accept } H_0 \text{ if } \|D^{-1/2} r\|_\infty \leq \tau_\mathcal{N}(\alpha)$$

We observe that for the case of single measurement with bad data, the largest normalized residual element $|r_i^N|$ corresponds to the corrupted measurement [11]. It is clear that both tests may be written as $\|W r(\hat{x})\|_p < \tau$, for suitable W, p, and τ.

V. Deception Attacks on Linear State Estimator

Several scenarios of stealthy deception attacks on PSSE for the DC case have been analyzed in [6]. The authors of [6] considered linear models, which were fully known by the attacker, and focused on additive attack policies that would guarantee the measurement residual to remain unchanged for the linear least squares algorithm. The feasibility of such attack policies was then analyzed for several IEEE benchmarks under different resource constraints of the attacker (for e.g., number of sensors the attacker could corrupt) and attacker objectives (for e.g., random attack, targeted attack). The main result related to attack policies was that if the attack vector a was in the range space of H, then the measurement residual $r^a = (z + a) - H\hat{x}$ would be the same as the residual r when there was no attack. Thus, such attack vectors would not increase the residual. Such undetectable errors have been analyzed previously within the power system's community, see [9], [13].

In this section we analyze how the attacker may fulfill the objective Section II-B, and thereby remain undetected.

A. Attack Synthesis

In general a stealthy attack requires the corruption of more measurements than the targeted ones, see [6], [14]. This relates to the fact that a stealthy attack must have the attack vector a fitting the measurement model, which for the weighted linear case is equivalent to have $a \in \mathrm{Im}(\bar{H})$.

We now present a general methodology for synthesizing stealthy attacks for the linear case with specific target constraints. Suppose the attacker wishes to compute an attack vector a such that $\bar{z}^a = \bar{z} + a$ satisfies a set of goals, encoded by $a \in \mathcal{G}$, and the attack is stealthy, *i.e.* $a \in \mathrm{Im}(\bar{H})$. Assuming the attacker knows the weighted measurement model $\bar{H}$, such attack could be computed by solving the optimization problem

$$\begin{aligned} &\min_{a} \|a\|_p \\ &\text{s.t. } a \in \mathcal{G},\ a \in \mathrm{Im}(\bar{H}) \ , \end{aligned} \tag{12}$$

corresponding to the "least-effort" attack in the p-norm sense. An interesting case is that of $p = 0$, which means the attacker is computing the attack with minimum cardinality, e.g., minimizing the number of sensors to corrupt. Another particular formulation is the 2-norm case with a single attack target, $z_a^i = z_i + 1$ or $a_i = 1$. By recalling that $a \in \mathrm{Im}(\bar{H})$ means that $a = \bar{H}c$ for a given c, the optimization problem may be recast as

$$\begin{aligned} &\min_{c} \|\bar{H}c\|_2^2 \\ &\text{s.t. } e_i^\top \bar{H} c = 1 \end{aligned} \ , \tag{13}$$

where e_i is a unitary vector with 1 in the i-th component. Recall $\bar{T} = \mathcal{P}_{\mathrm{Im}(\bar{H})} = \bar{H}\bar{H}^\dagger$.

Proposition 1: The optimal solution a^* to the optimization problem (13) is given by $a^* = \frac{\bar{T}}{\bar{T}_{ii}} e_i$

Proof: The Lagrangian of this optimization problem is $L(c, \nu) = c\bar{H}^\top \bar{H} c + \nu(e_i^\top \bar{H} c - 1)$ and the KKT conditions [15] for an optimal solution (c^*, ν^*) are

$$\begin{cases} \bar{H}^\top \bar{H} c^* + \nu^* \bar{H}^\top e_i = 0 \\ e_i^\top \bar{H} c^* - 1 = 0 \end{cases} . \tag{14}$$

Since it is assumed the power network is observable, the solution for the first equation is $c^* = \nu^* \bar{H}^\dagger e_i$. Including this in the second equation results in $\nu^* e_i^\top \bar{T} e_i = 1$ which is equivalent to $\nu^* = \frac{1}{\bar{T}_{ii}}$ with $\bar{T}_{ii}$ being the i-th diagonal element of $\bar{T}$. We then have that $a^* = \bar{H}c^* = \frac{\bar{T}}{\bar{T}_{ii}} e_i$. ∎

In the power system's literature, the hat matrix $\bar{T}$ is known to have information regarding measurement redundancy and correlation. This result highlights a new meaning: each column of $\bar{T}$ actually corresponds to an optimal attack vector yielding a zero residual.

B. Relaxing the Assumptions on Adversarial Knowledge

Here we consider the scenario where the attacker is performing an attack according to (12), but having only a partial or corrupted knowledge of the measurement model. Such knowledge may be obtained, for instance, by recording and analyzing data sent from the RTUs to the control center using suitable statistical methods. The corrupted measurement model may also correspond to an out-dated model or an estimated model using the power network topology, usual parameter values and uncertain operating point. We further assume that the covariance matrix R is known.

In the following analysis we provide bounds on the measurement residual under this kind of attack scenario. These bounds give some insights on what attacks may go undetected, given the model uncertainty. For the moment we assume there are no random errors in the measurements and so we consider the weighted measurements $\bar{z} = \bar{H}x$.

Let the perturbed measurement model known by the attacker be denoted by $\tilde{H}$, such that

$$\tilde{H} = \bar{H} + \Delta\bar{H}, \tag{15}$$

and consider the linear policy to compute attacks on the measurements to be $a = \tilde{H}c$, resulting in the corrupted set of measurements $\bar{z}^a = \bar{z} + a$. Recall the objectives of the attacker as defined in Section II-B.

The third objective, being undetected, depends both on the desired bias on the flow measurements a and on the model uncertainty $\Delta\bar{H}$. The measurement residual under attack, $r^a := \bar{r}(\bar{z}^a)$, can be written as

$$\bar{r}(\bar{z}^a) = \bar{S}(\bar{z} + \tilde{H}c) = \bar{S}\bar{z} + \bar{r}_a. \tag{16}$$

Using (15) and the fact that $\bar{S} = \mathcal{P}_{\mathrm{Ker}(\bar{H}^\top)}$, we can rewrite it as

$$\bar{r}(\bar{z}^a) = \bar{S}(\bar{z} + \bar{H}c) + \bar{S}\Delta\bar{H}c = \bar{S}\Delta\bar{H}c. \tag{17}$$

We denote $\bar{r}_a = \bar{S}\Delta\bar{H}c$ as the residual due to the attack, since it only depends on c and $\Delta\bar{H}$. Furthermore, we see that $\|\bar{r}_a\| \leq \|\bar{S}\|\|\Delta\bar{H}\|\|c\| = \|\Delta\bar{H}\|\|c\|$, since $\bar{S}$ is an orthogonal projector, showing that the residual norm is linear in terms of the model uncertainty. However, this bound does not capture an important property of the sensitivity matrix $\bar{S}$, *i.e.*, $\bar{S}$ is the orthogonal projector onto $\mathrm{Ker}(\bar{H}^\top)$. To show this, assume $\tilde{H} = \delta\bar{H}$ for some nonzero δ, yielding $\Delta\bar{H} = (1-\delta)\bar{H}$. From the previous result we have $\|\bar{r}_a\| \leq \|(1-\delta)\bar{H}\|\|c\|$. However, since $\bar{S}$ is the orthogonal projector onto $\mathrm{Ker}(\bar{H}^\top)$ and this subspace is the orthogonal complement of $\mathrm{Im}(\bar{H})$ we know that $\bar{r}_a = \bar{S}\Delta\bar{H}c = 0$. Therefore, although there is model uncertainty, the residual is still zero. This reasoning indicates that there is a geometrical meaning in the residual, since all the model perturbations $\Delta\bar{H}$ spanning $\mathrm{Im}(\bar{H})$ will yield a zero residual. To further explore this property, we will make use of the so-called principal angles and projection theory described in [16]. The main results and definitions used in this work are now given.

Definition 1 ([16]): Let M_1 and M_2 be subspaces of $\mathbb{C}^m$. The smallest principal angle $\gamma_1 \in [0, \pi/2]$ between M_1 and M_2 is defined by

$$\begin{aligned} \cos(\gamma_1) = &\max_{u \in M_1} \max_{v \in M_2} |u^H v| \\ &\text{subject to } \|u\| = \|v\| = 1 \end{aligned} \tag{18}$$

Lemma 1 ([16]): Let $\mathcal{P}_1, \mathcal{P}_2 \in \mathbb{R}^{m \times m}$ be orthogonal projectors of M_1 and M_2, respectively. Then the following holds

$$\|\mathcal{P}_1 \mathcal{P}_2\|_2 = \cos(\gamma_1) \tag{19}$$

Proposition 2: Let γ_1 be the smallest principal angle between $\mathrm{Ker}(\bar{H}^\top)$ and $\mathrm{Im}(\tilde{H})$. The residual increment due to a deception attack following the policy $a = \tilde{H}c$ satisfies

$$\|\bar{r}_a\|_2 \leq \cos\gamma_1 \|a\|_2. \tag{20}$$

Proof: Recall the so-called hat matrix defined by $\bar{T} = \bar{H}\bar{H}^\dagger$, which is the orthogonal projector onto $\mathrm{Im}(\bar{H})$ and

define $\tilde{T} = \mathcal{P}_{\mathrm{Im}(\tilde{H})} = \tilde{H}\tilde{H}^{\dagger}$. The residual under attack in Eq. (16) may be rewritten as

$$\bar{r}_a = \bar{S}\tilde{T}\tilde{H}c, \tag{21}$$

since $\tilde{T}\tilde{H} = \tilde{H}$. The residual norm can be upper bounded as

$$\|\bar{r}_a\|_2 \leq \|\bar{S}\tilde{T}\|_2\|\tilde{H}c\|_2 = \cos\gamma_1\|a\|_2, \tag{22}$$

where γ_1 is the smallest principal angle between $\mathrm{Ker}(\bar{H}^{\top})$ and $\mathrm{Im}(\tilde{H})$. ∎

Analyzing the example where $\tilde{H} = \delta\bar{H}$, we see that $\mathrm{Im}(\tilde{H}) = \mathrm{Im}(\bar{H})$ is orthogonal to $\mathrm{Ker}(\bar{H}^{\top})$. Hence the smallest principal angle between these subspaces is $\gamma_1 = \frac{\pi}{2}$, yielding $\|\bar{r}_a\|_2 \leq \cos(\gamma_1)\|a\|_2 = 0$.

Thus we achieved a tighter bound that explores the geometrical properties of the residual subspace. In brief, γ_1 measures how close the subspaces $\mathrm{Ker}(\bar{H}^{\top})$ and $\mathrm{Im}(\tilde{H})$ are from each other. In order for the model uncertainty not to affect the residual, it is desired that $\mathrm{Ker}(\bar{H}^{\top})$ and $\mathrm{Im}(\tilde{H})$ are as close to orthogonal as possible. For some insights on the physical interpretation of this geometrical property, see Section VI.

C. Stealthy Attacks

Consider the measurement residual under attack in (16). Taking into account the random error vector $\bar{\epsilon}$ we can rewrite the residual as

$$\bar{r}(\bar{z}^a) = \bar{S}\bar{\epsilon} + \bar{S}a. \tag{23}$$

The residual then has the following distribution $\bar{r}(\bar{z}^a) \sim \mathcal{N}(\bar{r}_a, \bar{S})$. Note that due to the model uncertainties the residual has a non-zero mean, which increases the chances of triggering an alarm in the BDD. Recall that one of the attacker's objective is to keep such probability as low as possible, *i.e.* $\|Wr(\hat{x}^a)\|_p < \tau$. We now provide insights on how such objective may be fulfilled for the two BDD schemes presented in Section IV.

1) Performance index test: Recall that without any attack on the measurements we have $J(\hat{x}) \sim \chi^2_{m-n}$. Under attack the cost function $J_a(\hat{x}) = \bar{r}(\bar{z}^a)^{\top}\bar{r}(\bar{z}^a)$ will have the so-called *non-central chi-squares* distribution [17], due to the non-zero mean which affects all the statistical moments of the χ^2_{m-n} distribution. We denote $J_a(\hat{x}) \sim \chi^2_{m-n}(\lambda)$ where $\lambda = \|\bar{S}a\|_2^2$. Recalling the relationship between the false alarm probability α and the detection threshold $\tau_\chi(\alpha)$ in (9), in the presence of attacks we have

$$\int_{\tau_\chi(\alpha)}^{\infty} g_\lambda(u)du = \alpha + \delta_\lambda(\lambda), \tag{24}$$

with $g_\lambda(u)$ being the pdf of $\chi^2_{m-n}(\lambda)$. We call $\delta_\lambda(\lambda)$ the increase in the alarm probability that the attacker must minimize to remain undetected. It is not possible to attack the PSSE and guarantee that no alarm is triggered, due to the presence of random measurement errors. Therefore we assume the attacker has an upper limit on $\delta_\lambda(\lambda)$ which is considered acceptable, $\bar{\delta}_\lambda$. Given reasonable values of α, the attacker is able to compute feasible values of λ by solving

$$\int_{\tau_\chi(\alpha)}^{\infty} g_\lambda(u)du \leq \alpha + \bar{\delta}_\lambda. \tag{25}$$

Under the reasonable assumption that $\delta_\lambda(\lambda)$ increases with λ, since the mean of $\chi^2_{m-n}(\lambda)$ is shifted along the positive direction and its variance increases as λ increases, we provide the following result.

Proposition 3: Suppose that $\delta_\lambda(\lambda)$ increases with λ. Given α and $\bar{\delta}_\lambda$ an attack is stealthy regarding the performance index test if the following holds

$$\cos\gamma_1\|a\|_2 \leq \sqrt{\bar{\lambda}(\alpha, \bar{\delta}_\lambda)} \tag{26}$$

where $\bar{\lambda}(\alpha, \bar{\delta}_\lambda)$ is the maximum value of λ for which (25) is satisfied.

Proof: First note that from our assumption $\delta_\lambda(\lambda)$ increases with λ. Therefore stealthy attack vectors satisfy $\|\bar{r}_a\|_2 \leq \sqrt{\bar{\lambda}}$, as this implies by definition that $\lambda \leq \bar{\lambda}$ and $\delta_\lambda(\lambda) \leq \bar{\delta}_\lambda$. The rest of the proof follows from Prop. 2. ∎

2) Largest normalized residual test: Recall that the residuals without attack follow a normal distribution $\bar{r} \sim \mathcal{N}(0, \bar{S})$, whereas under attack we have $\bar{r}_a \sim \mathcal{N}(d, \bar{S})$ with $d = \bar{S}a$. Each element of the normalized residual vector then has distribution $r^N_{a_i} \sim \mathcal{N}(d^N_i, 1)$ with $d^N_i = D^{-1/2}_{ii}d_i$ being the bias introduced by the attack vector. Similarly as before, defining $\bar{\delta}_d$ as the maximum admissible increase in the alarm probability and given α, the biases d^N_i providing the required level of stealthiness satisfy the inequality

$$\int_{-\tau_{\mathcal{N}}(\alpha)}^{\tau_{\mathcal{N}}(\alpha)} g^{\mathcal{N}}_{d^N_i}(u)du \geq 1 - \alpha - \bar{\delta}_d, \tag{27}$$

with $g^{\mathcal{N}}_{d^N_i}(u)$ being the pdf of $r^N_{a_i}$.

Proposition 4: Given α and $\bar{\delta}_d$ an attack is stealthy regarding the largest normalized residual test if the following holds

$$\|D^{-1/2}\|_2\cos\gamma_1\|a\|_2 \leq \bar{d}^N(\alpha, \bar{\delta}_d)\ , \tag{28}$$

where $\bar{d}^N(\alpha, \bar{\delta}_d)$ is the maximum value of $\|d^N\|_\infty$ for which (27) is satisfied with $d^N_i = \|d^N\|_\infty$.

Proof: Clearly it is sufficient to require (27) to hold for $|d^N_i| = \|d^N\|_\infty$, as this corresponds to the worst-case bias. Note that the increase in alarm probability δ_d increases with $|d^N_i|$ due to the symmetrical nature of $g^{\mathcal{N}}_{d^N_i}(u)$. Thus (27) reaches equality for $\|d^N\|_\infty = \bar{d}^N$ and a sufficient condition for (27) to hold is to have $\|d^N\|_\infty \leq \bar{d}^N$. Recalling $d^N = D^{-1/2}\bar{S}a$ and $\|\cdot\|_\infty \leq \|\cdot\|_2$, we conclude the attack is stealthy if $\|D^{-1/2}\bar{S}a\|_2 \leq \bar{d}^N$, which is satisfied by $\|D^{-1/2}\|_2\|\bar{S}a\|_2 \leq \bar{d}^N$. The rest follows from Proposition 2. ∎

The main result of this section is as follows:

Theorem 1: Given the perturbed model $\tilde{H}$, the false-alarm probability α and the maximum admissible increase in alarm probability $\bar{\delta}$, an attack following the policy $a = \tilde{H}c$ is stealthy if

$$\|a\|_2 \leq \beta(\alpha, \bar{\delta})\ , \tag{29}$$

where $\beta(\alpha,\bar{\delta})$ is given by:

- $\beta(\alpha,\bar{\delta}) = \frac{\sqrt{\bar{\lambda}(\alpha,\bar{\delta}_\lambda)}}{\cos\gamma_1}$, for the performance index test;
- $\beta(\alpha,\bar{\delta}) = \frac{\bar{d}^N(\alpha,\bar{\delta}_d)}{\|D^{-1/2}\|_2\cos\gamma_1}$, for the largest normalized residual test.

Proof: Assuming the BDD method is the performance index and taking $\beta(\alpha,\bar{\delta}) = \frac{\sqrt{\bar{\lambda}(\alpha,\bar{\delta}_\lambda)}}{\cos\gamma_1}$, the proof directly follows from Proposition 3. For the largest normalized residual, defining $\beta(\alpha,\bar{\delta}) = \frac{\bar{d}^N(\alpha,\bar{\delta}_d)}{\|D^{-1/2}\|_2\cos\gamma_1}$ the proof follows from Proposition 4. ■

Note that in the scenario analyzed here, the designer of the BDD scheme chooses both the detection method as well as the false-alarm probability α. These elements are fixed and usually unknown to the attacker, who defines the maximum risk $\bar{\delta}$ he is willing to take and has some knowledge of the power network $\tilde{H}$, that is used to compute the attack vector a. However α can be estimated by reasonable values and the same happens for the degrees of freedom of the chi-squares distribution. Although the exact value of γ_1 is not accessible to an attacker tampering only with RTUs, additional knowledge such as the topology of the network may be used to compute worst-case estimates of γ_1, as it is shown in the next section.

VI. Case Study

An interesting analysis is to understand what is the worst-case uncertainty for the attacker, $\Delta\bar{H}$, maximizing the orthogonality between $\mathrm{Im}(\tilde{H})$ and $\mathrm{Im}(\bar{H})$. This corresponds to maximizing the effect of the attack vector a on the measurement residual. From the attacker's view, this could lead to a set of robust attack policies. As for the control center this could be useful to implement security measures based on decoys, for instance. It is known that the network model used in the PSSE can be kept in the databases of the SCADA system with little protection. Thus a possible defensive strategy would be, for instance, to disseminate a perturbed model with fake but "genuine" looking parameter values in the database which, if retrieved and used by an attacker, would produce large residuals and increase the detection of intelligent attacks.

The first observation at this point is that it is of little interest to consider cases when only the maximum magnitude of the model perturbation is considered,*i.e.* $\|\Delta\bar{H}\| \leq \omega$. Note that this formulation only tells us that the uncertainty is within a ball of radius ω from the nominal model $\bar{H}$. Thus one can always choose a worst-case perturbation satisfying $\|\Delta\bar{H}\| = \omega$ which is orthogonal to $\bar{H}$, yielding $\|\bar{S}\bar{T}_\Delta\| = 1$. Hence scenarios where the uncertainty is more structured are of greater interest.

We now apply the previous results to the scenario where the attacker knows the exact topology of the network but has an error on the transmission line's parameters of $\pm 20\%$. The detectability of attacks in this scenario is intimately related to the detectability of parameter or topology errors [13], [18]. Consider the power network in Fig. 2 with the data in Table I. The network shown in Fig. 2 corresponds to the bus-branch model of a, possibly larger, power network

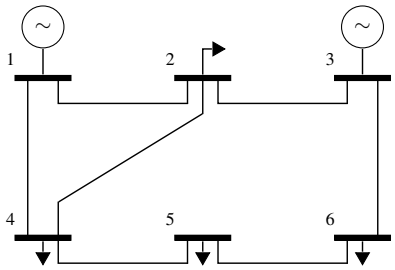

Fig. 2. Power network with 6 buses

TABLE I
Data of the network in Fig. 2

Branch	From bus	To bus	Reactance (pu)	Parameter Error
$b1$	1	4	0.370	-20%
$b2$	1	2	0.518	+20%
$b3$	6	5	1.05	-20%
$b4$	6	3	0.640	-20%
$b5$	5	4	0.133	-20%
$b6$	4	2	0.407	-20%
$b7$	3	2	0.300	+20%

computed by the EMS after analyzing which buses and branches are energized, based on measurements from RTUs such as breaker status. This model is then used by the PSSE, together with the list of available measurements, to compute the measurement model. In this example we consider the linear case where $z = Hx$. The parameter errors in Table. I were computed so that $\cos(\gamma_1) = \|\bar{S}\tilde{T}\|_2$ is maximized for errors up to $\pm 20\%$, corresponding to the worst-case uncertainty. This actually corresponds to the constrained maximization of a convex function, which was solved using the numerical solvers available in MATLAB.

In Fig. 3 we show how the maximum 2-norm of a stealthy

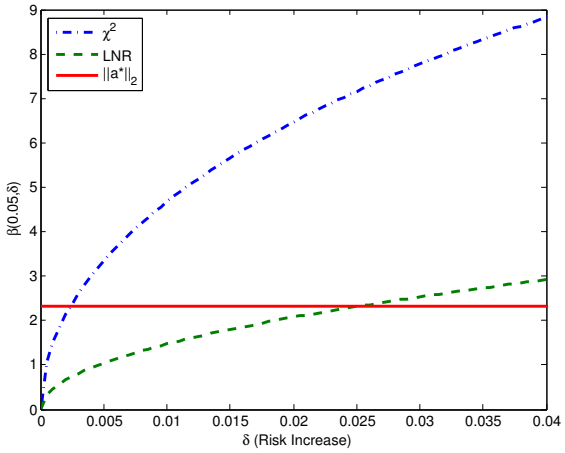

Fig. 3. Attack stealthiness as a function of the detection risk. The solid line represents the 2-norm of the optimal attack vector a^* constrained by $a_{b_1} = 1$, where a_{b_1} is the power flow in branch b_1. The curves denoted as χ^2 and *LNR* represent the value of $\beta(0.05,\delta)$ for the performance index test and largest normalized residual test, respectively. From these results, we conclude that the *LNR* test is more sensitive to this kind of attacks.

attack vector $\beta(\alpha, \delta)$ in terms of Theorem 1 varies with respect to the increased detection risk δ, for $\alpha = 0.05$. As it is seen, the performance index test allows for larger attacks than the largest normalized residual test. Since attacks following $a = \bar{H}c$ have a similar meaning to multiple interacting bad data, this validates the known fact that largest normalized residual test is more robust to such bad data than the performance index test [11]. Note that the norm of the optimal attack vector in the sense of (13) when targeting the power flow between buses 1 and 4 is also shown. We see that such attack would have a small risk, even for the largest normalized residual.

VII. Conclusions

In this work we provided methods to analyze cyber-security of PSSE in scenarios where the attacker has a limited knowledge of the network and unlimited resources. In particular we proposed a framework to model such attackers, which is capable of taking into account resource constraints. We also explored and considered two BBD methods widely used and showed that such tools do not guarantee security against cyber-attacks.

References

[1] A. Giani, S. Sastry, K. H. Johansson, and H. Sandberg, "The VIKING project: an initiative on resilient control of power networks," in *Proc. 2nd Int. Symp. on Resilient Control Systems*, Idaho Falls, ID, USA, Aug. 2009, pp. 31–35.

[2] A. Cárdenas, S. Amin, and S. Sastry, "Research challenges for the security of control systems." in *Proc. 3rd USENIX Workshop on Hot topics in security*. USENIX, July 2008, p. Article 6.

[3] "Electricity grid in U.S. penetrated by spies," *The Wall Street Journal*, p. A1, April 8th 2009.

[4] "Cyber war: Sabotaging the system," *CBSNews*, November 8th 2009.

[5] "Final report on the August 14th blackout in the United States and Canada," U.S.-Canada Power System Outage Task Force, Tech. Rep., April 2004.

[6] Y. Liu, M. K. Reiter, and P. Ning, "False data injection attacks against state estimation in electric power grids," in *Proc. 16th ACM Conf. on Computer and Communications Security*, New York, NY, USA, 2009, pp. 21–32.

[7] S. Amin, A. Cárdenas, and S. Sastry, "Safe and secure networked control systems under denial-of-service attacks." in *HSCC*, ser. Lecture Notes in Computer Science, R. Majumdar and P. Tabuada, Eds., vol. 5469. Springer, 2009, pp. 31–45.

[8] Y. Mo and B. Sinopoli, "Secure control against replay attack," in *Proc. 47th Annual Allerton Conf.*, Monticello, IL, USA, Sep. 2009, pp. 911–918.

[9] K. A. Clements, G. R. Krumpholz, and P. W. Davis, "Power system state estimation residual analysis: An algorithm using network topology," in *IEEE Trans. Power App. Syst.*, Apr. 1981.

[10] L. Mili, T. V. Cutsem, and M. Ribbens-Pavella, "Bad data identification methods in power system state estimation - a comparative study," in *IEEE Trans. Power App. Syst.*, Nov. 1985.

[11] A. Abur and A. Exposito, *Power System State Estimation: Theory and Implementation*. Marcel-Dekker, 2004.

[12] A. Monticelli, "Electric power system state estimation," in *Proc. IEEE*, vol. 88, no. 2, Feb. 2000.

[13] F. F. Wu and W.-H. E. Liu, "Detection of topology errors by state estimation," *IEEE Trans. Power Syst.*, no. 1, Feb. 1989.

[14] H. Sandberg, A. Teixeira, and K. H. Johansson, "On security indices for state estimators in power networks," in *Preprints of the 1st Workshop on Secure Control Systems, CPS Week 2010*, Stockholm, Sweden.

[15] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004.

[16] A. Galántai, "Subspaces, angles and pairs of orthogonal projections," *Linear and Multilinear Algebra*, vol. 56, no. 3, pp. 227–260, Jun. 2006.

[17] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, 1982.

[18] W.-H. E. Liu, F. F. Wu, and S.-M. Lun, "Estimation of parameter errors from measurement residuals in state estimation," *IEEE Trans. Power Syst.*, no. 1, Feb. 1992.

Appendix

Convergence of Newton's method

For Newton method applied to WLS estimation problem, we have $F(x) = -H(x)^\top R^{-1}(z - h(x))$. Assuming that $[F'(x)]$ is nonsingular, following (3) we define

$$G(x) = x - [F'(x)]^{-1}F(x). \tag{30}$$

$G : \mathbb{R}^n \to \mathbb{R}^n$. A solution $x^* = G(x^*)$ is called the *fixed point* of G. Since G arises as an iteration function for the equation $F(x) = 0$, x^* is a fixed point of G if and only if $F(x^*) = 0$. The local convergence theorem for Newton iterates is as follows:

Theorem 2: Let F be continuously differentiable function, and $[F'(x)]$ be nonsingular with elements continuous in the ball $\mathcal{S} := \{x \in \mathbb{R}^n | \parallel x - x^0 \parallel < \epsilon\}$. Let us define

$$c := \max_{\xi \in \mathcal{S}} \parallel G'(\xi) \parallel_\infty .$$

Suppose the following conditions are satisfied

(A1) $c < 1$

(A2) $\parallel G(x^0) - x^0 \parallel < (1 - c)\epsilon$

then

- There exists a unique solution of $F(x) = 0$ in $\mathcal{S}$,
- the sequence $\{x^0, x^1, x^2, \ldots\}$ generated by G will converge to the fixed point x^* of G in $\mathcal{S}$,
- $\parallel x^i - x^* \parallel < \frac{c}{1-c} \parallel x^i - x^{i-1} \parallel$.