Peekaboo: A Gray Hole Attack on Encrypted SCADA Communication using Traffic Analysis

Nunzio Marco Torrisi
Centro de Matemática, Computação e Cognição
Universidade Federal do ABC
Santo André, Brazil 09.210-170
Email: nunzio.torrisi@ufabc.edu.br

Ognjen Vuković and György Dán
School of Electrical Engineering
KTH, Royal Institute of Technology
Stockholm, Sweden
Email: {vukovic,gyuri}@ee.kth.se

Stefan Hagdahl
Security & Defense Solutions
Saab AB
Stockholm, Sweden
Email: stefan.hagdahl@saabgroup.com

***Abstract*—We consider a potential gray hole attack against SCADA substation to control center communications using DNP3. We propose a support vector machine-based traffic analysis algorithm that relies on message direction and timing information only, and we use trace-based simulations to show that even if SCADA traffic is sent through an encrypted tunnel, as often done in practice, the gray hole attack can be effectively performed based on the timing and direction of three consecutive messages. Our results show that the attacker does not need accurate system information to be successful, and could affect monitoring accuracy by up to 20%. We discuss possible mitigation schemes at different layers of the communication protocol stack, and show that a minor modification of message timing could help mitigate the attack.**

I. Introduction

Electric power systems have to be continuously monitored and controlled via Supervisory Control and Data Acquisition (SCADA) systems in order to be kept in a secure operating state. Meters at remote substations measure power flows and voltages, and the measurements are communicated to one or more SCADA control centers over a communication infrastructure using some SCADA communication protocol, such as Distributed Network Protocol 3 (DNP3) [1]. The dynamic visibility provided by SCADA systems has long been important in transmission systems and is becoming more important in power distribution systems, because the proliferation of intermittent distributed generation sources (e.g., solar) results in faster changes in power flows, which in turn requires that protection devices and integrated voltage and VAR control (iVVC) be adjusted in real time.

Motivated by the reliance of power systems on monitoring, estimation and control, a large body of recent work considered the impact of data integrity attacks on power system state estimation, from single systems [2], [3], [4], [5], [6], [7] to interconnected systems [8]. These works assume that the attacker is able to manipulate measurement data in lack of proper authentication.

Authentication is often indeed not possible in legacy remote terminal units (RTUs), and therefore in most SCADA systems the measured data are sent through an encrypted and authenticated tunnel between a substation gateway and the control center. Tunneling protects integrity and may provide confidentiality against an attacker that has access to one or more communication links or routers, and should be used to conform with NERC CIP. Since encryption hides the message contents from an attacker along the tunnel, one would expect that it would also make it impossible for an attacker to identify and to drop mission critical measurement and/or control messages without dropping all messages in a tunnel, and thus remain undetected or difficult to be detected.

In this paper we show through the example of DNP3, one of the two standardized SCADA substation to control center communication protocols, that targeted gray hole attacks may be feasible despite sending messages through an encrypted tunnel. We propose a support vector machine based traffic analysis attack that can distinguish between reports sent spontaneously by an RTU to the control center and messages sent by the RTU in response to messages by the control center. The attack is computationally simple, and is based on the inter-arrival times and directions of consecutive encrypted messages. We use measurement data sets from medium voltage substations to evaluate the effectiveness of the attack and its sensitivity, and to quantify the impact that the attack may have on monitoring accuracy. We finally discuss mitigation schemes to alleviate the attack. Our results give evidence to that the strict timing rules used in SCADA communication protocols facilitate traffic analysis attacks and appropriate countermeasures should be applied.

The rest of the paper is organized as follows. In Section II we review related work, and in Section III we give an overview of DNP3. We describe the system and the attack model in Section IV, followed by the attack in Section V. We evaluate the attack and propose a mitigation scheme in Section VI. Section VII concludes the paper.

II. Related Work

The vulnerability of SCADA systems to cyber attacks has received significant attention recently. In [9], the authors discuss challenges and difficulties of achieving all-encompassing component-level cyber security in power systems due to its cost and potential performance implications. False data injection attacks against common control system communication protocols were considered in [10], [11]; the authors proposed intrusion detection systems to detect the attacks based on neural networks [10] and based on the concepts of critical state analysis and state proximity [11]. Certain false data injection attacks can bypass the bad data detection algorithm used in SCADA state estimators [2], and can thus be used to deceive the system operators regarding the actual state of the system [2], [5], [3], [4], [6], [7]. Mechanism were proposed to protect against these attacks by securing a subset of measurements [5], [3], [4], and by securing a part of the SCADA infrastructure [5], [6], [7]. In [8], the authors showed that false data injection attacks against distributed state estimation in an interconnected power system can disable state estimation in the entire interconnected system, and proposed a detection and a mitigation scheme against such attacks. Our work differs from these recent works as we consider an attack that is limited to dropping messages, and we investigate the effectiveness of such an attack.

Related to ours are works that aim to identify application layer protocols sent through a tunnel using pattern recognition methods [12]. A support vector machine was used in [13] to identify protocols other than HTTP and SSH tunneled over HTTP or SSH by looking at the message size, the block cipher size (involved in the message encryption), and the MTU size. Application-layer protocols sent through an encrypted tunnel that carries traffic from many TCP connections simultaneously were classified in [14], [15] using a k-Nearest-Neighbor classifier based on Hidden Markov Models with the message size, the message direction, and message inter-arrival times as features. In [16], the authors compared Bayesian Networks, Decision Trees and Multilayer Perceptrons for the flow-based classification of six different types of Internet traffic, including peer-to-peer and content delivery traffic, and showed the importance of correctly classifying training instances. In [17], the authors proposed an unsupervised machine learning method for network traffic classification based on information entropy techniques. Furthermore, they combined the unsupervised method with a supervised learning method and showed that the combination can improve classification. Unlike these works that aim to identify different protocols in a tunnel, the attack we consider aims at classifying messages that belong to the same application layer protocol, DNP3, and we investigate the ability of such an attack to interfere with SCADA monitoring. To the best of our knowledge ours is the first work to consider a targeted gray hole attack against tunneled DNP3 traffic.

III. DNP3 Background

DNP3 is one of the two standardized communication protocols for SCADA substation to control center communication [1]. Its design follows the master/slave communication model; the *master* is the SCADA master station at the control center and the slaves, called *outstations*, are Remote Terminal Units (RTUs), Intelligent Electronic Devices (IEDs) and Programmable Logic Controllers (PLCs) at the substations.

A. Polling vs. Report by exception

DNP3 allows two types of data acquisition, *polling* and *report-by-exception*. In the case of polling, the master solicits data from an outstation and the outstation replies immediately with all data. In the case of report-by-exception, the outstation reports only the values that have changed since the last report by more than a predefined threshold, instead of reporting all data. The advantage of this choice is significant saving in bandwidth.

These two types of data acquisition can be combined, and result in four modes of operation for DNP3: (i) polled static, (ii) polled report-by-exception, (iii) quiescent, and (iv) unsolicited report-by-exception. In the case of (i) the master polls and the outstation reports all data. In case of (ii) the master polls but the outstation only reports changed values. In case of (iii) the master does not poll, an unsolicited response is generated by the outstation whenever a value changes by a predefined threshold. In case of (iv) the master polls periodically (typically at a low frequency) and an unsolicited response is generated by the outstation whenever a value changes by a predefined threshold. This last mode is the most commonly used in practice, as it allows for the detection of communication failures and keeps the bandwidth usage low.

B. DNP3 over IP

DNP3 includes a link layer specification (addressing, framing, etc), but it can also operate on top of a transport layer protocol, such as TCP and UDP, when used in IP networks [1]. In practice DNP3 is often used over UDP, because using UDP keeps the outstation implementation simple, using UDP does not require many connections to be kept alive in the master station, and if the operator has to pay for the amount of SCADA traffic then using UDP would also be less costly. Furthermore, DNP3 itself implements reliable transmission, hence reliability at the transport layer is not needed.

C. Sequence numbers and the Vulnerability

In order to achieve reliable transmission, every message is identified with a sequence number, and message reception is acknowledged so that lost messages can be retransmitted, if needed. For unsolicited responses DNP3 allows two retransmission strategies. One strategy allows the outstation to send a new unsolicited response without receiving the acknowledgement for the previous one, while the other strategy requires the outstation to wait for the acknowledgement before sending a new unsolicited response. An important feature of DNP3 is that the sequence numbers sent by an outstation in unsolicited responses are independent from the sequence numbers used in solicited responses (i.e., in response to polls).

This design choice makes a **gray hole attack** possible: in lack of signaling from the outstation to the master, as long as solicited responses are delivered, the master station can not tell if an attacker drops all unsolicited responses. This is the attack we consider, and we investigate whether the attack can be performed even if the DNP3 messages are sent through an encrypted tunnel, as is usually done in SCADA systems.

IV. System and Attack Model

We consider a master and an outstation that use DNP3 for communication over a wide area network; the outstation

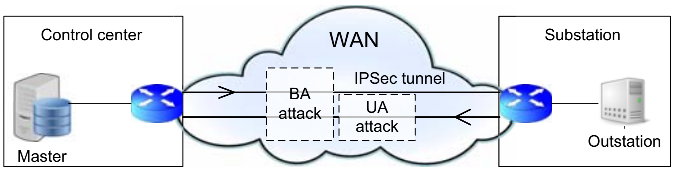

Fig. 1. Considered system: Master and outstation communicate using DNP3 through an IPSec tunnel over a WAN.

reports measurement data, such as power flow and voltage measurements. DNP3 is used in unsolicited report-by-exception mode, as commonly done in real systems: the outstation reports measurement data by replying to poll messages sent by the master or by sending an unsolicited response when the relative change of a measured value exceeds a configured reporting threshold Δ. We consider a modern WAN deployment, based on the TCP/IP protocol stack, and consider that UDP is used at the transport layer. For reliable delivery the master is configured to send a confirmation message for each unsolicited response it receives. If the outstation does not receive a confirmation for an unsolicited response, the outstation retries sending until the confirmation is received or until the number of retries exceeds a predefined threshold. Since the communication infrastructure is typically not trusted, end-to-end data integrity and confidentiality are achieved through establishing an IPSec tunnel for the DNP3 traffic between the substation gateway and the master station, in ESP mode [18]. In order to avoid non-mission critical data (such as video, voice or engineering data traffic) to interfere with DNP3, the tunnel typically carries DNP3 traffic only. There is thus one IPSec tunnel per DNP3 connection, as shown in Fig 1.

A. System Model

We denote the set of polling messages sent by the master by $\mathcal{M}^p = \{m_1^p, m_2^p, ...\}$, the set of solicited responses sent by the outstation by $\mathcal{M}^s = \{m_1^s, m_2^s, ...\}$, and the set of unsolicited responses (including retranmissions) sent by the outstation by $\mathcal{M}^u = \{m_1^u, m_2^u, ...\}$. We denote the set of all DNP3 messages sent by the outstation to the master by $\mathcal{M}^o = \mathcal{M}^s \cup \mathcal{M}^u$ and the set of messages exchanged by the master and the outstation by $\mathcal{M} = \mathcal{M}^p \cup \mathcal{M}^o$.

We denote by t_n^p the time instant when the master sends polling message $m_n^p \in \mathcal{M}^p$ ($n \in \mathbb{N}$), and by $t_n^s > t_n^p$ the time when the outstation replies with solicited response $m_n^s \in \mathcal{M}^p$. The time $t_n^s - t_n^p$ is determined by the one-way delay and the message processing time at the outstation, and is typically rather small compared to the polling period. Similarly, we denote by t_k^u the time when the outstation sends unsolicited response $m_k^u \in \mathcal{M}^u$ ($k \in \mathbb{N}$). If the response is not confirmed, the outstation sends a retransmission $m_{k+1}^u \in \mathcal{M}^u$ at time t_{k+1}^u. Note that the index of a message in a set is determined by the time the message is sent, e.g., $t_{n-1}^p < t_n^p < t_{n+1}^p$.

B. Attack Model

The goal of the attacker is to perform a gray hole attack on the data reported by the outstation to the master, while remaining undetected. The attacker has access to a component of the communication network between the substation and the control center, such as a switch, a router or a communication link. The attacker can observe the IPSec tunnels traversing the network component and can identify an IPSec tunnel that carries DNP3 traffic; it can observe the encapsulated DNP3 messages and it can *drop* individual messages. The attacker cannot observe the payload of the messages due to the use of IPsec in ESP mode, but for each message it intercepts it can observe the size of the message's payload, which it can use to differentiate between DNP3 messages and IPsec session management messages, similarly to [13], [14], [19].

Depending on the physical layer technology, the network topology and the routing, the messages sent by the master to the outstation ($\mathcal{M}^p$) and the messages sent by the outstation to the master ($\mathcal{M}^o$) may travel over separate physical links and paths. We therefore consider two models for the attack, the Unidirectional Access (UA) attack and the Bidirectional Access (BA) attack, shown in Fig 1. In the case of the *UA* attack, the attacker can only observe the messages sent from the outstation to the master, i.e., the messages in $\mathcal{M}^o$. In the case of the *BA* attack the attacker can observe the messages sent in both directions, i.e., the messages in $\mathcal{M}$.

Upon intercepting a message the attacker can record the actual time. We denote by t_n^a the time instant when the attacker observes message m_n; in case of the *BA* attack $m_n \in \mathcal{M}$, while in case of the *UA* attack $m_n \in \mathcal{M}^o$.

To perform the attack, the attacker should discard the unsolicited response messages; as long as no unsolicited responses are delivered to the master, the master cannot detect missing sequence numbers, since in DNP3 the sequence numbers are not related in the two directions. To remain undetected, the attacker should discard very few solicited responses as the master can notice the loss of solicited responses (in response to polls). Thus, in order to succeed the attacker has to identify whether an intercepted message is a DNP3 unsolicited response or a DNP3 solicited response. For a sequence of messages, we denote by $\mathcal{M}^{au}$ the set of messages the attacker classifies as unsolicited response.

V. Peekaboo: Binary Classifier Attacks

Clearly, there is a trade-off between correctly classifying the two kinds of messages. We formulate the goal of the attacker as maximizing the probability of correctly classifying an unsolicited response, while keeping the probability of incorrectly classifying a solicited response under a certain threshold c, or formally

$$\begin{aligned} \max \quad & P(m \in \mathcal{M}^{au} | m \in \mathcal{M}^u), \\ & \text{s.t. } P(m \in \mathcal{M}^{au} | m \in \mathcal{M}^s) < c. \end{aligned} \tag{1}$$

We describe two classes of attack algorithms to solve the problem based on past message inter-arrival times, and if available, based on past message directions.

The considered attacks identify the unsolicited responses by using a support vector machine (SVM) with an appropriately chosen feature space $X \subseteq \mathbb{R}^p$ [20]. Given l training feature vectors $x_n \in X$, $n = 1, \ldots, l$ and for each vector the corresponding class $y_n \in \{-1, 1\}$, an SVM is a supervised learning model that finds a hyperplane w that solves

$$\min_{w,\xi_n,b} \left(\frac{1}{2}|w|^2 \quad + \quad C \sum_{n=1}^{l} \xi_n \right) \qquad (2)$$

subject to

$$y_n(w * x_n - b) \quad \geq \quad 1 - \xi_n, \quad n = 1, \ldots, l, \qquad (3)$$

where $\xi_n \geq 0$ are slack variables, $C > 0$ is a constant that allows to trade-off between false negatives and false positives, $*$ is an operator that defines the type of the classifier, and b is a scalar. If the operator $*$ is the scalar product, then the classifier is linear and w defines a hyperplane in the feature space. If the operator $*$ is a non-linear kernel function, then the classifier is non-linear, and w defines a hyperplane in the transformed feature space. A widely used non-linear kernel function is the Gaussian radial basis function, for which the transformed feature space is a Hilbert space of infinite dimensions.

Given the trained SVM, i.e., w and b computed, the attacker constructs feature vector x_n for message m_n it intercepts, and decides whether to drop the message based on the sign of $w * x_n - b$. The *UA* and the *BA* attack models differ in the feature space, and are both parameterized by an integer $k > 0$.

UA(k) attack: Under the *UA(k)* attack, the attacker uses the k inter-arrival times between the last $k+1$ messages it observes. The feature vector that corresponds to message $m_n \in \mathcal{M}^o$ is $x_n = (t^a_n - t^a_{n-1}, \ldots, t^a_{n-k+1} - t^a_{n-k})^T$. The feature space of the SVM for the *UA*(k) attack is $\mathbb{R}^k$.

BA(k) attack: Under the *BA(k)* attack, the attacker uses the k inter-arrival times between the last $k+1$ messages together with the direction of the messages. The feature vector that corresponds to message $m_n \in \mathcal{M}$ is $x_n = (t^a_n - t^a_{n-1}, \ldots, t^a_{n-k+1} - t^a_{n-k}, d_n, \ldots, d_{n-k})^T$, where $d_n \in \{-d, d\}$ for some constant $d > 0$, depending on whether the message is sent by the outstation or by the master, respectively. Since the feature vector includes information about the message direction, the feature space for the *BA*(k) attack is $\mathbb{R}^{2k+1}$.

VI. Attack Impact and Mitigation

In the following we evaluate the efficiency of the attacks, we illustrate their potential impact and we consider potential mitigation schemes using traced-based simulations.

A. Measured traces

Our evaluation is based on three measurement data sets collected at medium voltage substations of a European power distribution system operator. The measurements were taken every 3 seconds over 7 consecutive days, and include the voltage and current phasors for the three phases. As RTUs typically report RMS voltage magnitude and active and reactive power flows, we computed these quantities from the traces.

Fig. 2 shows the complementary cumulative distribution function (CCDF) of unsolicited report inter-arrival times, i.e., the CCDF of $t^u_{k+1} - t^u_k$ ($k \in \{1, 2, ...\}$), assuming three different reporting threshold values $\Delta \in \{1\%, 5\%, 10\%\}$ based on one of the traces. We observe that the CCDF decays slower for higher values of Δ as unsolicited reports are sent less often due to the higher relative change required to trigger an unsolicited report. It is important to note that the range of inter-arrival times is very wide, between 2 and 4 orders of magnitude and correspondingly the standard deviations are high, 11.2s, 234s, and 932s for $\Delta = 1\%$, $\Delta = 5\%$ and $\Delta = 10\%$, respectively.

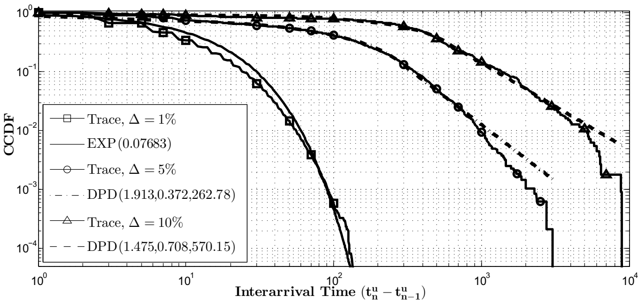

Fig. 2. CCDF of unsolicited response inter-arrival times with best fit Double Pareto distributions, DPD(α,β,ω), and Exponential distribution, Exp(λ).

The figure shows for each empirical CCDF the CCDF of the best fit double Pareto distribution [21] and the best fit exponential distribution, together with the parameters α, β, and ω, and λ, respectively. The figure shows that for higher threshold values Δ, the double Pareto distribution is a rather good fit and captures large part of the tail. The two regions with different power-law exponents are due to the different power demand dynamics during daytime (fast changing) and nightime (slow changing). For $\Delta = 1\%$ large inter-arrival times are rare because even small power flow and voltage fluctuations trigger unsolicited responses, and thus the exponential distribution seems to provide a very good fit.

B. Attack Success Rate

We evaluate the efficiency of the attacks for the scenario shown in Fig. 1, i.e., DNP3 traffic exchanged over UDP/IP between an outstation and a master station transmitted through an IPSec tunnel. The unsolicited reports are generated by the outstation based on the measurement data sets in response to voltage magnitude, and active and reactive power flow changes with a threshold of $\Delta = 1\%$. The master is configured to send polling messages every T_p seconds and the outstation sends a solicited report with the most recently measured values immediately after receiving a polling message. The round-trip time (RTT) between the master and the outstation, including the delay due to encryption, authentication and processing at the outstation, equals $1s$ in the baseline scenario.

Fig. 3 and Fig. 4 show the false negative and false positive rates $P(m \notin \mathcal{M}^{au} | m \in \mathcal{M}^u)$ and $P(m \in \mathcal{M}^{au} | m \notin \mathcal{M}^u)$, as a function of the polling period T_p and various k values for the UA attack and for the BA attack, respectively. The kernel function used is the Gaussian radial basis function. The false negative rates and the false positive rates of the *UA(k)* attacks are rather high, which would make the *UA(k)* attacks easy to detect. Interestingly, relying on more messages makes

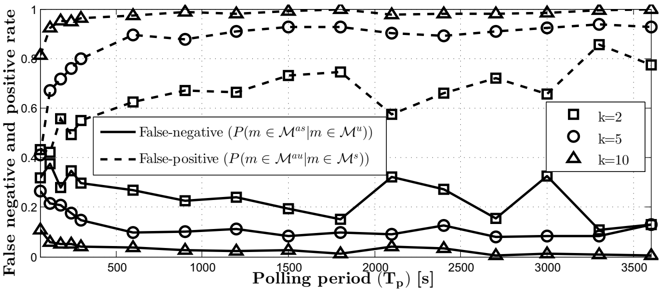

Fig. 3. False negative and positive rate vs. polling period for UA attack.

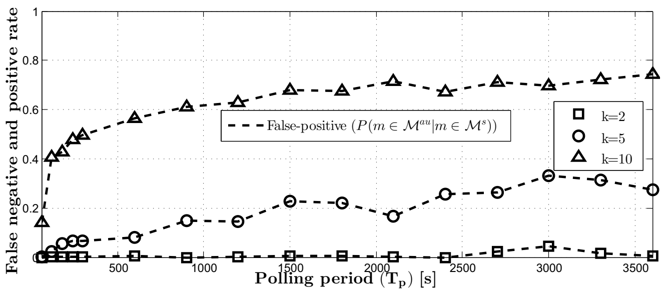

Fig. 4. False negative and positive rate vs. polling period for BA attack.

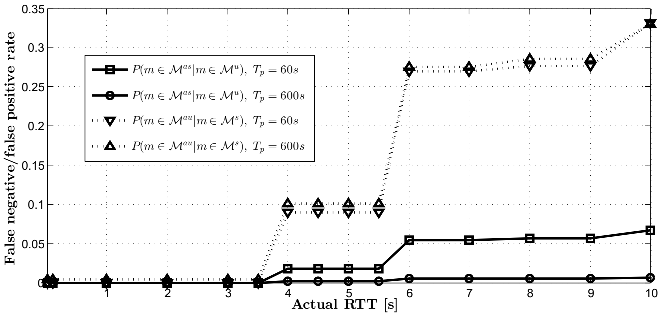

Fig. 5. False negative and positive rate vs. actual RTT. SVM is trained for RTT=1s.

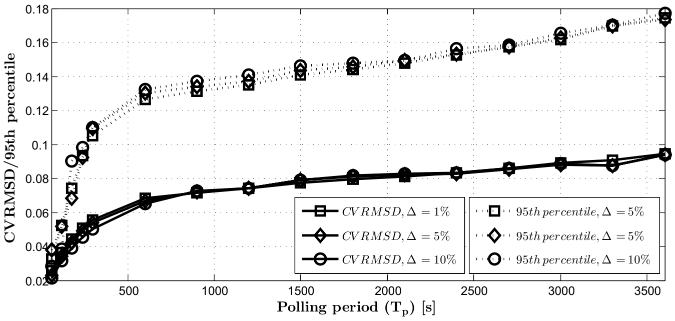

Fig. 6. $CV_{RMSE}(E_P^a(t))$ and $NRSE(E_P^a(t), t)$ for an active power flow and the *BA* attack vs. T_p. $\Delta \in \{1\%, 5\%, 10\%\}$.

the attack even weaker. The *BA(k)* attacks are, however, very effective. First, the false negative rate of the *BA(k)* attacks is zero (hence it is not shown). Second, the false positive rate is consistently lower for low k. The strongest attack is *BA(2)*, and hence we use it in the sequel. The *BA* attack's efficiency is due to the ability of the attacker to observe the messages in both directions; intuitively any report coming from the outstation shortly after a polling message is classified as a solicited request, and all others as unsolicited requests. Thus, for an attack to be successful, the attacker needs to be able to observe messages sent in both directions.

The results in Fig. 3 and Fig. 4 were obtained assuming that the attacker knows the (RTT) between the master and the outstation. Figure 5 shows the sensitivity of the false negative and of the false positive rate for the *BA(2)* attack using an SVM that was trained with RTT=1s as a function of the actual RTT. The figure shows that the *BA(2)* attack is effective as long as the actual RTT is below 4s, i.e., as long as the attacker's estimate of the RTT is within a factor of four, which is a rather wide margin of error. Above a factor of four the false negative and the false positive rates start to increase and the attack could be detected. The stepwise increase in the misclassification rates at RTT 4s and 6s is due to that measurements in the data sets were taken every 3s.

C. Attack Impact

The results so far show that the *BA* attack could effectively be used for selectively dropping unsolicited reports and this way blind an operator. We quantify the potential effect of the attack on the situational awareness of an operator through the error that the attacker would introduce in the power flow measurements available to an operator under the attack. We define the error at time t as the difference between the measured value $P(t_n^s)$ received in the most recent solicited response and the measured value $P(t_k^u)$ the operator should have received in the most recent unsolicited response (had it not been dropped by the attacker), i.e., for $t_n^s < t < t_{n+1}^s$

$$E_P^a(t) = \begin{cases} P(t_k^u) - P(t_n^s) & if\ \exists t_k^u \text{ s.t. } t_n^s < t_k^u \le t < t_{k+1}^u, \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

We define the mean squared error over the time interval $[t_1, t_2]$ as $\overline{E_P^a(t)^2} = \frac{1}{t_2 - t_1}\int_{t_1}^{t_2} E_P^a(t)^2$, and the coefficient of variation of the root mean squared error as

$$CV_{RMSE}(t_1, t_2) = \sqrt{\frac{\overline{E_P^a(t_1, t_2)^2}}{\overline{P(t_1, t_2)}^2}}, \tag{5}$$

where $\bar{\cdot}$ stands for the mean. In practice, reacting to sudden short changes of $P(t)$ is important for proper operation of the power system, we therefore also compute the normalized root squared error for every time instant as

$$NRSE(t) = \sqrt{\frac{\overline{E_P^a(t)^2}}{\overline{P(t))}^2}}. \tag{6}$$

Fig. 6. shows CV_{RMSE} and the 95th percentile of $NRSE(t)$ over the 7 days measurement period as a function of the polling interval T_p for one of the active power flow measurements for the *BA* attack (the attacker successfully drops all unsolicited responses). Both the mean and the 95 percentile increase monotonically with T_p, with a decreasing marginal gain. These results indicate that under an attack the operator's observation of the active power flow would be almost 20% off in 5% of the time and it would be on average up to 10% off. Interestingly, the results are not sensitive to the reporting threshold Δ.

D. Attack Mitigation

Motivated by the effectiveness of the *BA* attack and its potential impact, we finally discuss a number of mitigation

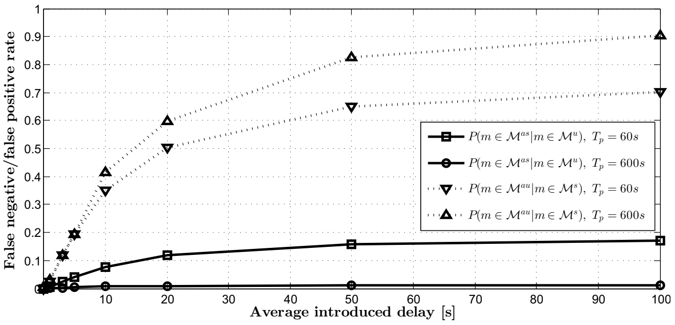

Fig. 7. False negative and positive rate vs. average introduced delay.

schemes. At the transport layer one could mitigate the attack by using TCP, as the attack would cause head of line blocking and would lead to a reset of the TCP connection. This mitigation may, however, not be feasible if the legacy equipment does not support TCP or server resources are insufficient.

At the application layer, the DNP3 solicited response could be extended by a field that contains the sequence number of the most recently sent unsolicited response. As an alternative, the outstation could introduce a random delay before sending a solicited response (in response to a poll). The random delay would make an attack using statistical pattern recognition more difficult. To assess this latter mitigation scheme, Figure 7 shows the false negative rate and the false positive rate as a function of the average delay introduced in the outstation for the case of an exponential distribution and the *BA(*2*)* attack. The choice of the exponential distribution is motivated by the observation that the inter-arrival times of unsolicited responses are well modeled by an exponential distribution for a small reporting threshold. The false negative and the false positive rates increase with a decreasing marginal gain with the introduced delay, and for a relatively small average delay of a few seconds they would be high enough for the *BA(k)* attack to be detected. An interesting open question is whether such delays would be compatible with legacy SCADA masters.

VII. Conclusion

We addressed the vulnerability of SCADA communication to a gray hole attack, in which an attacker drops unsolicited reports sent by an outstation to a SCADA master, while letting through solicited reports in order to avoid detection. We showed that such a gray hole attack is possible even if messages are sent through an encrypted tunnel, because due to the strict timing rules used in SCADA protocols traffic analysis can effectively be used to classify protocol messages. We proposed a support vector machine based traffic analysis algorithm, used trace-based simulations to evaluate the attack, and showed that an attacker would not need exact knowledge of system parameters for a successful attack. We quantified the impact of the attack in terms on monitoring accuracy, and showed that the operator's observation can be up to 10% off on average, and up to 20% off in 5% of the time. Finally, we discussed potential mitigation schemes, and showed that the attack can be mitigated by introducing a random delay before answering to poll messages.

Acknowledgment

N. Torrisi was supported by the Swedish-Brazilian Research and Innovation Centre (CISB) and by the SAAB group (Grant CNPq-SAAB No. 236756/2012-5). Dán and Vuković were supported by EIT ICTLabs activity SES 14306.

References

[1] DNP3 IEEE WG, "IEEE Standard for Electric Power Systems Communications-Distributed Network Protocol (DNP3)," *IEEE Std 1815-2012 (Revision of IEEE Std 1815-2010)*, pp. 1–821, 2012.

[2] Y. Liu, P. Ning, and M. Reiter, "False data injection attacks against state estimation in electric power grids," in *Proc. of the 16th ACM conference on Computer and Communications Security (CCS)*, 2009, pp. 21–32.

[3] R. B. Bobba, K. M. Rogers, Q. Wang, H. Khurana, K. Nahrstedt, and T. J. Overbye, "Detecting false data injection attacks on dc state estimation," in *Preprints of the First Workshop on Secure Control Systems, CPSWEEK, Stockholm, Sweden*, April 2010.

[4] T. T. Kim and H. V. Poor, "Strategic protection against data injection attacks on power grids," *IEEE Trans. on Smart Grid*, vol. 2, no. 2, pp. 326–333, Jun 2011.

[5] G. Dán and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Proc. of IEEE SmartGridComm*, October 2010.

[6] A. Giani, E. Bitar, M. Garcia, M. McQueen, P. Khargonekar, and K. Poolla, "Smart grid data integrity attacks: Characterizations and countermeasures," in *Proc. of IEEE SmartGridComm*, October 2011.

[7] O. Vuković, K. C. Sou, G. Dán, and H. Sandberg, "Network-aware mitigation of data integrity attacks on power system state estimation," *IEEE JSAC: Smart Grid Communications Series*, vol. 30, no. 6, pp. 176–183, July 2012.

[8] O. Vuković and G. Dán, "Detection and localization of targeted attacks on fully distributed power system state estimation," in *Proc. of IEEE SmartGridComm*, October 2013, pp. 390–395.

[9] G. Dán, H. Sandberg, G. Björkman, and M. Ekstedt, "Challenges in power system information security," *IEEE Security and Privacy*, 2011.

[10] W. Gao, T. Morris, B. Reaves, and D. Richey, "On SCADA control system command and response injection and intrusion detection," in *eCrime Researchers Summit (eCrime), 2010*. IEEE, 2010, pp. 1–9.

[11] A. Carcano, A. Coletta, M. Guglielmi, M. Masera, I. N. Fovino, and A. Trombetta, "A multidimensional critical state analysis for detecting intrusions in SCADA systems," *IEEE Trans. on Industrial Informatics*, vol. 7, no. 2, pp. 179–186, 2011.

[12] K. Fukunaga, *Introduction to statistical pattern recognition*. Academic Press, 1990.

[13] M. Dusi, M. Crotti, F. Gringoli, and L. Salgarelli, "Tunnel hunter: Detecting application-layer tunnels with statistical fingerprinting," *Computer Networks*, vol. 53, no. 1, pp. 81–97, 2009.

[14] C. V. Wright, F. Monrose, and G. M. Masson, "On inferring application protocol behaviors in encrypted network traffic," *J. Mach. Learn. Res.*, vol. 7, pp. 2745–2769, Dec. 2006. [Online]. Available: http://dl.acm.org/citation.cfm?id=1248547.1248647

[15] G. Maiolini, A. Baiocchi, A. Iacovazzi, and A. Rizzi, "Real time identification of ssh encrypted application flows by using cluster analysis techniques," in *Proc. of IFIP/TC6 Networking*, 2009, pp. 182–194.

[16] M. Soysal and E. G. Schmidt, "Machine learning algorithms for accurate flow-based network traffic classification: Evaluation and comparison," *Performance Evaluation*, vol. 67, no. 6, pp. 451–467, 2010.

[17] J. Yuan, Z. Li, and R. Yuan, "Information entropy based clustering method for unsupervised internet traffic classification," in *Proc. of IEEE ICC*, 2008, pp. 1588–1592.

[18] S. Kent, "IP Encapsulating Security Payload (ESP)," RFC 4303 (Standard Track), Internet Engineering Task Force, 2005. [Online]. Available: http://www.ietf.org/rfc/rfc4303.txt

[19] X. Tan, X. Su, and Q. Qian, "The classification of ssh tunneled traffic using maximum likelihood classifier," in *Proc. of IEEE Conf. on Electronics, Comm. and Control (ICECC)*, 2011, pp. 2347–2350.

[20] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural computation*, vol. 13, no. 7, pp. 1443–1471, 2001.

[21] W. J. Reed and M. Jorgensen, "The double Pareto-lognormal distribution - a new parametric model for size distribution," *Communications in Statistics - Theory and Methods*, vol. 33, no. 8, pp. 1733–1753, 2004.