Defending Synchrophasor Data Networks Against Traffic Analysis Attacks

Biplab Sikdar *Senior Member, IEEE,* and Joe Chow *Fellow, IEEE*

***Abstract*—The use of synchrophasor data for observation and control is expected to enhance the operation and efficiency of the next generation of power systems. However, the specific characteristics of the data generated by synchrophasors makes them particularly vulnerable to cyber attacks. This paper presents a set of strategies to protect the anonymity of synchrophasor data against passive traffic analysis attacks. Considering the periodic nature of synchrophasor data, we propose defense mechanisms based on packet concatenation and random packet drops as a countermeasure against attacks that may use the timing as well as data volume information to compromise the network. In contrast to existing defenses against traffic analysis attacks, our scheme can be easily deployed using the current networking infrastructure as it is based on end-to-end principles and does not require any specialized routers. The proposed defense mechanisms are evaluated using both analysis and simulations.**

***Index Terms*—Network security, smart grid, synchrophasor network**

I. Introduction

THE addition of synchrophasor measurements is expected to provide a number of important features in smart grids [1]. The existing and ongoing deployment of phasor measurement units (PMUs) for providing the synchrophasor data are envisioned to be an integral part of smart grids and provide valuable information about the state of the system. PMU data serves to facilitate a number of applications while enhancing others, such as real-time monitoring of the system, state estimation, disturbance monitoring, instability prediction, wide area protection and control, etc. [1], [2]. Given their importance in the maintenance and control of the power generation and distribution system, monitoring and manipulation of PMU data are particularly attractive avenues for malicious attackers that intend to disrupt and damage the power infrastructure [3]. Additionally, the synchrophasor measurement data is usually transferred over public domain networks such as the Internet, thereby making it susceptible to a number of attacks. This paper investigates the susceptibility of the PMU data measurement and collection network against a class of passive attacks and develops defense strategies against the attacks.

Synchrophasor data collected at geographically diverse locations are usually routed to data concentrators in central locations. These central locations are either owned by or serve independent system operators (ISOs) and transmission owners (TOs) that use the data in their wide area monitoring systems. The most common communication protocol used by PMUs to transmit their data is defined in the IEEE C37.118 standards. While the networking infrastructure, the communication protocols, as well as the measurement and collection devices may all be vulnerable to a variety of malicious attacks, this paper focuses on a set of passive attacks that may be launched in the network.

The focus of this paper is on passive attacks that are mainly concerned with privacy and anonymity issues. In passive attacks, the attacker does not alter the data or the system resources, but attempts to learn or make use of the information from the system. In particular, *traffic analysis* based passive attacks aim to determine the identity and location of the communicating hosts by observing the timing and length of messages across links in the network. In the context of synchrophasor measurement data, such attacks can be used to determine the network locations of the PMUs and central stations, as well as the path taken by their data. Subsequently, this information may be used to launch sophisticated attacks on the PMUs, monitoring stations and the network routers, leading to loss of data, increased response times and restricted availability of resources. As a first step in understanding the vulnerabilities of synchrophasor data transmissions, this paper limits its focus on traffic analysis attacks.

The inherent characteristics of synchrophasor measurement data make the problem of defending them against traffic analysis attacks quite challenging. The data measured and reported by the PMUs includes frequencies, phasors, analog values and digital values [4]. PMUs are usually equipped with a global positioning system (GPS) and produce data with accuracy better than 0.1% that are timestamped with precision better than 1 μs. The data is generated periodically at fixed intervals and the data packets thus generated have the same size (since the same set of values is measured each time). The constant packet size makes it easier for attackers to correlate the traffic generated by a PMU as it propagates across links. Additionally, the requirements of applications such as real-time system monitoring, wide area control and protection etc. imply that the data must be sent to the central locations without delay. Consequently, traditional obfuscation approaches such as batching and introduction of random delays [5] comes with a penalty. Finally, we note that while encryption may conceal the contents of a data packet, they cannot hide the transmission timing information exploited by traffic analysis attacks. To address the challenges listed above, we use a strategy that mixes data concatenation and selective packet dropping as a countermeasure against traffic analysis attacks.

B. Sikdar and J. Chow are with the Department of Electrical, Computer and Systems Engineering, Rensselaer Polytechnic Institute, Troy, NY, 12180 USA.

This work was supported in part by grants from NYPA and NYSERDA.
Manuscript received October 16, 2010; revised April 15 2011.

Existing work on securing networks carrying PMU data is limited, and to the best of our knowledge, there is none that addresses traffic analysis attacks. The focus of previous work on the analysis and prevention of traffic analysis based attacks has focused mainly on *mix* networks [5], [6], [7], [8], [9]. These solutions are based on the use of specialized mix routers that have not been deployed in the Internet. Thus the focus of this paper is on developing techniques that are readily deployable in the existing networking infrastructure. To achieve this goal, we propose a strategy that is purely based on operations carried out at the end hosts, i.e., the synchrophasors generating the measurement data.

This paper makes two contributions. First, it presents and evaluates a set of schemes, specifically tailored for synchrophasor measurement data, for defeating traffic analysis attacks. Second, it uses an information theoretic measure to quantify the degree of anonymity provided by the proposed obfuscation strategies. The proposed schemes are evaluated using simulations in realistic networking scenarios.

The rest of this paper is organized as follows. Section II presents the related work and Section III presents an overview of traffic analysis attacks and describes the threat model assumed in this paper. The proposed defense mechanism against traffic analysis attacks on synchrophasor data is proposed in Section IV. Section V uses an information theoretic measure to quantify the degree of anonymity associated with the proposed defense mechanism. Section VI presents simulation results to evaluate the performance of the proposed scheme. Finally, Section VII concludes the paper.

II. Related Work

A survey of various security issues related to smart grids is presented in [3] while [11] presents an overview of the potential sources and cost of security failures in smart grids deployments. The VIKING project focusing on developing a resilient and secure power transmission and distribution system (including the possibility of cyber attacks) is described in [12].

Existing literature on actual security mechanisms and protocols specifically designed for PMUs is limited. Design principles and engineering practices for developing authentication protocols for smart grids is discussed in [10]. However, security issues related to the advanced metering infrastructure (AMI) has received considerable attention [13], [14]. A third-party escrow mechanism for authenticated anonymous meter readings is presented in [15]. Privacy preserving metering aggregation and comparison using Paillier encryption has been proposed in [16] while [17] describes an architecture for secure metering that relies on trusted components outside of the meter. The goal of these techniques is primarily to obfuscate the value of the meter readings and they do not address the problem of traffic analysis attacks considered in this paper.

Existing literature on the analysis and prevention of traffic analysis based attacks has focused mainly on *mix* networks [5], [6], [7], [8], [9]. These techniques require the use of specialized mix routers that perform security operations. Since mix node routers are not readily available in the current Internet, these strategies are not suitable for practical deployment. In contrast, our focus is on developing strategies that can be deployed using the current networking infrastructure without any changes to the network components. To achieve this goal, we propose a strategy that is purely based on operations carried out at the PMUs[1].

III. Traffic Analysis Attacks

Traffic analysis attacks are primarily based on finding correlations between various traffic parameters, such as timing and volume information, on two links in a network. The underlying principle behind traffic analysis attacks is that parameters such as inter-packet interval etc. vary widely between different packet flows and thus can be used as identifiers to distinguish and isolate flows. In the most commonly considered traffic analysis attacks, the adversary divides time into fixed-size bins and correlates the parameter of interest (such as packet count) across two links.

The most common traffic analysis attack in literature is one based on timing analysis. The basic aim of a timing analysis attack is to find a correlation between the timings of the packets of a flow seen on one link with the timings of another flow on another link. The presence of a strong correlation is taken as an indication that the two flows are the same. The most commonly used random variable for evaluating the timing correlations is the interval between two packets [5].

In order to minimize the effect of packet drops on the inter-packet times, the number of packets in a given interval may also be used in traffic analysis attacks. Other random variables such as the sizes of the packets observed on two flows in two links (or the total number of bytes transferred by two flows over a given interval) may also be used in traffic analysis attacks.

A. Threat Model

The threat model assumed in this paper is that the adversary has a partially global view of the network, i.e., the adversary has the ability of monitor some but not all of the links in the network. We assume that the adversary monitors and collects both the interarrival times, as well as the sizes (and thus the data volumes) of packets that pass through compromised links. The attacks launched by the adversary are passive and neither the content nor the flow of packets in the networks are altered in any way. We assume that the characteristics of the traffic flow from the synchrophasors is known to the adversary. This assumption is based on the fact that the nature of measurements and the data originating from synchrophasors is, in general, well known. However, the adversary cannot correlate a packet on an input link of a router to a packet on one of its output links (or on any other link), using either timing, size or content information. The use of packet concatenation in our scheme prevents correlations based on size, encryption prevents correlations based on packet contents, and both packet concatenation and dropping prevent correlations based on timing. All data originating from a given

[1]Additional differences with individual schemes proposed in literature are presented in Section IV-B.

synchrophasor is considered to be part of a single flow. The network topology is assumed to be known to the adversary.

Under the threat model described above, the paper considers the following traffic analysis attack: Given that a synchrophasor data flow with known characteristics is present at a specific input link of a router, and a number of flows with indistinguishable packets at the output links of the router, the adversary wishes to determine the output link that contains the synchrophasor data flow. In addition, given that a synchrophasor data flow with known characteristics is present at a specific input link of a router, the attacker may wish to determine which downstream router carries the same flow at one of its input links. Such threats are one of the most important and usually considered in security literature [5], [18] since it allows the adversary to establish the path taken by the flow (and thus the routers that handle the flow) as well as the origin and destination of the flow. Consequently, the adversary may attack these nodes to disrupt the flow or compromise the nodes to read or alter the contents of the data.

IV. Countermeasures Against Traffic Analysis Attacks

In this paper we assume an adversary model where the attacker is able to observe the times at which packets are sent on various links as well as the lengths of the packets. To counter the traffic analysis attacks, we propose a strategy that uses a mix of data concatenation at the source and random packet dropping at the routers along the path taken by the synchrophasor data from the PMUs to the central location.

A. Data Concatenation Strategy

A possible traffic analysis attack is to correlate the sizes of the packets and intervals between them to identify a traffic flow. Synchrophasor data flows are particularly vulnerable to this type of attack since the data they generate is of fixed length and at constant intervals. Given that the data generated is of fixed length, simply randomizing the inter-packet times may not be sufficient in defeating traffic analysis attacks since the attacker can correlate the total volume of traffic passing through the links. An additional constraint for the defense systems here is the strict delay requirements of synchrophasor data, which may limit the delays that may be introduced to randomize the inter-packet times.

As the first step of defense against traffic analysis attacks, we propose a scheme where PMUs randomly concatenate the data from a number of contiguous measurements to form a single packet. The concatenation of a random number of data samples produces a packet of random size and thus avoids sending a stream of fixed length packets that is relatively easy to identify. The maximum number of measurement data that may be concatenated depends on the latency requirements and the time between two successive data samples (i.e. data packets) generated by a PMU[2].

As a way of defeating the concatenation based strategy, an attacker may look at the packet size and be able to exploit the fact that concatenated packets are an integral multiple of a single packet. Thus we allow for finer granularity in the length of a concatenated packet by allowing the PMU to append an arbitrary number of bytes from the last measurement that is included in a given concatenated packet. Thus if three measurements are considered for inclusion in a concatenated packet, an arbitrary number of bytes may be included from the third measurement. The remainder of the data from the third sample is sent in the next concatenated packet. Control fields may be introduced and marked to indicate that a packet contains partial information from a measurement at its beginning and/or end.

The random packet concatenation strategy introduces a delay in the synchrophasor measurement data and if this is not controlled, the additional delay may be detrimental to the performance and effectiveness of various applications that use the data. Thus an upper bound needs to be placed on the number of packets that may be concatenated. Typical applications that use synchrophasor measurement data usually require that the data be delivered within a few hundred milliseconds, depending on the application, for it to be useful [19], [20]. Given the sampling rate of the synchrophasor measurement units and compensating for the network latency, the maximum number of packets that may be concatenated can be easily obtained.

The algorithm describing the random packet concatenation strategy is shown in Algorithm 1. Before the transmission starts, the algorithm initializes itself by obtaining the number of hops, h, and the average propagation time, T_p, to the destination. These variables can be obtained through the use of tools like traceroute [24]. Depending on the latency requirements, δ, of the application, the maximum latency introducable by the concatenation mechanism, Δ, is determined: $\Delta = \delta - T_p$. The maximum number of packets that may be concatenated is then $N_m = \Delta/\varphi$ and the largest size of a concatenated packet is $V_m = \lceil N_m D \rceil$, where φ is the interval between two data packets generated by a synchrophasor and D is the length of each packet in bytes. When a new packet is generated by the synchrophasor, it is added to the tail of the queue of packets to be transmitted. The number of packets to be concatenated is determined (using the function concatenate in the algorithm) whenever a new packet arrives at an empty queue or whenever a concatenated packet is transmitted and leaves behind a non-empty queue. The number of packets n to be concatenated is drawn uniformly between $n = U(0, N_m]$ where U denotes a uniformly distributed random variable. Note that the proposed strategy allows for a packet to be split and a concatenated packet to be comprised of less than D bytes (for $n < 1$). This allows for greater variability in the size of the transmitted packets. The function concatenate also determines the time that must elapse before the next transmission and sets a timer (denoted by $timer_c$) with that value. The new concatenated packet, comprising of n data packets, is sent when the timer expires. Now, n concatenated packets corresponds to $\lceil nD \rceil$ bytes. If the current queue length at time t is Q, the new concatenated packet can be sent immediately if $\lceil nD \rceil \leq Q$ and the timer is thus set to 0. Otherwise, we need to wait for another $\lceil nD \rceil - Q$ bytes of data (corresponding to $\lceil (\lceil nD \rceil - Q)/D \rceil$ new data packets) to arrive

[2]In this paper we use the terms PMU and synchrophasor interchangeably.

Algorithm 1 Random packet concatenation scheme.

Given: h, T_p, δ, φ, D;
Initialize: $\Delta = \delta - T_p$;
Initialize: $N_m = \Delta/\varphi$;
Initialize: $V_m = \lceil N_m D \rceil$;
Initialize: $Q = 0$;
while (1) **do**
  if (new packet arrival) **then**
    enqueue packet;
    $Q = Q + D$;
    if $(timer_c == 0)$ **then**
      $(n, timer_c) =$ concatenate();
    end if
  end if
  if $((timer_c == 0)$ && $(Q > 0))$ **then**
    transmit concatenated packet;
    $Q = Q - \lceil nD \rceil$;
    if $Q > 0$ **then**
      $(n, timer_c) =$ concatenate();
    end if
  end if
end while
function concatenate()
{
  $n = U(0, N_m]$;
  if $(\lceil nD \rceil \leq Q)$
    $timer_c = 0$;
  else
    $timer_c = [\lceil t/\varphi \rceil + \lceil (\lceil nD \rceil - Q)/D \rceil - 1]\,\varphi - t$;
  end if
  return(n,$timer_c$);
}

before we can send the concatenated packet. If the current time is t seconds, we need to wait $\lceil t/\varphi \rceil \varphi - t$ seconds for the first new data packet to arrive and another $(\lceil (\lceil nD \rceil - Q)/D \rceil - 1)/\varphi$ seconds for the remaining new data packets to arrive. The timer in this case is thus set to

$$\begin{aligned} timer_c &= \left\lceil \frac{t}{\varphi} \right\rceil \varphi - t + \left(\left\lceil \frac{\lceil nD \rceil - Q}{D} \right\rceil - 1 \right) \varphi \\ &= \left(\left\lceil \frac{t}{\varphi} \right\rceil + \left\lceil \frac{\lceil nD \rceil - Q}{D} \right\rceil - 1 \right) \varphi - t. \end{aligned} \quad (1)$$

Finally, a new concatenated packet is sent whenever the timer decrements to zero and the queue length is positive. Note that a concatenated packet is sent as soon as the last data packet to complete the desired size of the concatenated packet is generated by the synchrophasor.

An artifact of the packet concatenation strategy is that it introduces a randomization in the inter-packet times. However, the upper bound on the acceptable delays imposed by the applications using the data limits the maximum inter-packet times. Consequently, traffic analysis attacks that observe and use the number of packets in a given window and the data volume as the random variables for correlation may be able to defeat a strategy based only on packet concatenation. To address this and other issues, we next introduce a preventive strategy based on random packet drops.

Algorithm 2 Random drop strategy.

Given: h, λ, p_{drop}, D_{max};
Initialize: $timer_d = 0$;
while (1) **do**
  if $(timer_d == 0)$ **then**
    $D_d = U[1, D_{max}]$;
    $rand = U[0, 1]$;
    if $(rand < p_{drop})$ **then**
      $TTL = U[1, h]$;
    else
      $TTL = TTL_{max}$;
    end if
    transmit dummy packet
    $timer_d = EXP(\lambda)$
  end if
end while

B. Random Drop Strategy

While concatenating the data to generate packets of random lengths also serves to eliminate packet transmissions at regular intervals, an attacker can correlate the gaps between two packets and the lengths of the two packets to detect a flow (for example, two packets spaced by three times the usual interval with the second packet being three times as large suggests packet concatenation). Thus to break the regular spacing between packets from a PMU while still having packets of random length, in addition to the concatenation strategy, we suggest the insertion of dummy packets that are randomly dropped at the routers. With the concatenation strategy providing random packet lengths, insertion of dummy packets at random intervals helps to break the timing information inherent in the PMU data generation process. The random drop mechanism is introduced to reduce the correlations between the traffic at an incoming link of a router and that on the outgoing link of that flow, and also between the traffic of the same flow on different hops in its path from the source to the destination.

The random drop strategy proposed in this paper is shown in Algorithm 2. The random drop strategy works independently of the packet concatenation scheme presented earlier. The random drop strategy is based on the insertion of packets of random length at random intervals. Based on a configurable rate λ of dummy packets per second, the drop scheme inserts dummy packets at exponentially distributed inter-arrival times. Among all continuous time distributions with base $[0, \infty)$ and a given mean, the exponential distribution has the highest entropy. Thus in our scheme, the packet interarrival time is generated using the exponential distribution. The length of each dummy packet, D_d, is also random and is uniformly distributed according to $U[1, D_{max}]$, where D_{max} is the maximum packet length allowed in the network. The maximum packet size is usually determined when a connection is set up by protocols such as the Transmission Control Protocol

(TCP) and can thus be easily obtained. Each dummy packet may be dropped at any of the intermediate nodes in its path from the source to the destination. To implement the random drop strategy, we use the time to live (TTL) feature built into the Internet Protocol (IP), the default layer 3 protocol in the Internet [24]. The TTL feature limits the amount of time a packet spends in the network. The TTL value of each packet is initialized at the source node and decremented by one at each router that it passes. If the TTL value reaches zero at a router before the packet reaches its destination, it is dropped. Given the number of hops h (as described in the previous subsection) and a desired drop rate, p_{drop}, for each packet, we generate a TTL value at random for each packet using the distribution $U[1, h]$. Thus packets are dropped randomly inside the network, depending on where a packet's TTL reaches zero. Dummy packets that are not to be dropped are given a TTL value greater than h, i.e. TTL_{max}, the maximum TTL value allowed. Note that not all dummy packets are dropped in the network in order to provide cover to the synchrophasor traffic on the last hop.

Anonymizing schemes based on the insertion of cover traffic and random packet drops have been proposed earlier in literature, in the context of mix networks [6], [7], [5], [8], [9]. These techniques require the use of specialized mix nodes on each hop of the path. Each mix node is assumed to be capable of operations such as encryption, delaying, reordering and dropping packets, and insert dummy packets. However, mix networks have not been deployed in the Internet and are thus a valid option for practical and immediate needs. Our solution can be deployed using the current hardware and software deployed in the Internet and is based on end-to-end principles, and all modifications are made at the end hosts (the synchrophasors) and not the routers. In addition, existing dummy packet based anonymizing techniques are based on the insertion of a constant rate cover traffic [7], [9]. In contrast, we introduce random dummy traffic, which is capable of anonymizing synchrophasor traffic using a much lower overhead. Also, all our dummy traffic is inserted at the synchrophasor end and does not require insertions at the routers. Finally, unlike other drop based anonymizing schemes that have been proposed in literature, our mechanism can be implemented without requiring any changes to the intermediate routers or requiring the use of specialized routers such as mix nodes [5].

V. Information Theoretic Measure of Anonymity

In this section we use an information theoretic measure to describe the anonymity provided by the proposed obfuscation strategies. We use the notion of equivocation [21] to define the anonymity, as have been done in [22] in the case of mix networks. Consider a network with a number of flows, each following its own path. We use the term "session" to describe the set of all paths in the network and we denote a session by $\mathbf{S}$. The adversary observes the network at compromised points and wishes to determine the flows constituting a session.

In our case, a session will include one or more flows of synchrophasor data. In addition, a session may include flows of other traffic using the network (for example, email, file downloads etc.) with arbitrary origin and destination nodes. We assume that the flows constituting a session, i.e., a session $\mathbf{S}$, can be modeled as an independent identical random variable with density function $p(\mathbf{S})$. We assume that the adversary is aware of the distribution $p(\mathbf{S})$ which aids it in determining the flows in the session. To consider the worst case passive adversary situation, the adversary is assumed to have a global view of the network and can monitor the transmission time and the length (in bytes) of all packets in the network. On an arbitrary link (A, B), the transmission time and length of the i-th packet on that link are denoted by $T_{A,B}(i)$ and $V_{A,B}(i)$, respectively. The sequence of transmission times and packet lengths on link (A, B) are denoted by

$$\tau_{A,B} = \{T_{A,B}(1), T_{A,B}(2), \cdots\} \quad (2)$$

$$\nu_{A,B} = \{V_{A,B}(1), V_{A,B}(2), \cdots\} \quad (3)$$

and the global adversary has access to the transmission times and packet lengths over all links in the network. Denoting the set of all links by $\mathbf{E}$, the sets of all observations by the adversary, τ and ν, are given by

$$\tau = \{\tau_{A,B} \mid (A, B) \in \mathbf{E}\} \quad (4)$$

$$\nu = \{\nu_{A,B} \mid (A, B) \in \mathbf{E}\} \quad (5)$$

The presence or absence of traffic on the links that the adversary observes provides some information to the adversary about the flows in the network, and we denote the set of all links observed by the adversary by $\mathbf{L}$. In addition, the adversary uses the timing and traffic volume information to correlate the packet transmissions across links and determine the flows. In the obfuscation strategies proposed earlier, we have essentially sought to control the conditional distribution $q(\tau, \nu|\mathbf{S})$ governing the packet transmission times and packet lengths, in order the decrease the likelihood of an attacker determining the flows in a session based on its observations τ and ν. Using the notion of equivocation [21], we define the anonymity in terms of the conditional uncertainty of the information we wish to hide (i.e. $\mathbf{S}$) given the observations of the adversary (i.e. $(\tau, \nu,$ and $\mathbf{L})$ [22]. Then, a distribution $q(\tau, \nu|\mathbf{S})$ is said to have anonymity α if

$$\frac{H(\mathbf{S}|\tau, \nu, \mathbf{L})}{H(\mathbf{S}|\mathbf{L})} \geq \alpha \quad (6)$$

where we have used the notation $H(X|Y)$ to denote the entropy of random variable X given random variable Y, and

$$H(X|Y) = -E[\log p(X|Y)] \quad (7)$$

The anonymity metric α lies in the range $[0, 1]$ with $\alpha = 0$ denoting no anonymity and $\alpha = 1$ denoting perfect anonymity, i.e., a scenario where observing the transmission times and packet lengths provides no additional information to the adversary about the flows in the session than by just observing the links. Also, using Fano's inequality [23], it can be shown that the probability p_e that the adversary makes an error in

identifying the session is bounded by

$$p_e \geq \frac{H(\mathbf{S}|\tau,\nu,\mathbf{L})}{\log|\mathcal{S}|} \geq \frac{\alpha H(\mathbf{S}|\mathbf{L})}{\log|\mathcal{S}|} \tag{8}$$

where $\mathcal{S}$ is the set all sessions. This provides a physical interpretation for the anonymity metric.

In [22] it was shown that there is a tradeoff between the anonymity provided by the network and the latency it introduces in the packets. We now show that the concatenation scheme introduced in this paper introduces a bounded delay on the packets in the network. Also, it is easy to see that the expected volume of dummy packets introduced in the network per second by a synchrophasor data flow using our scheme is bounded by $\lambda D_{max}/2$. Using this in conjunction with the bounds on the delays introduced by the concatenation scheme, achievable bounds may thus be placed on the anonymity achieved by the strategies proposed in this paper.

Result 1: The expected delay, $E[T]$, incurred by an arbitrary packet due to the concatenation strategy proposed in Section IV is bounded by

$$E[T] \leq \frac{D}{2V_m}\frac{\lceil\frac{V_m}{D}\rceil\left(\lceil\frac{V_m}{D}\rceil+1\right)}{2}\varphi. \tag{9}$$

To prove the result above, we first obtain the expected concatenation delay, given that the size of the concatenated packet is V, i.e., $E[T|V]$. When the last concatenated packet is transmitted, the buffer may still contain a fragment of the last data packet, since we do not restrict the size of concatenated packets to be a multiple of integral data packets. Now, the number of new data packets required to form a concatenated packet of size V is at most $\lceil V/D\rceil$. Note that the first data packet that contributes to the concatenated packet may already be in the buffer since part of it may have been used in the previous concatenated packet. Thus this data packet has to wait for another $\lceil V/D\rceil\varphi$ seconds for the new data packets to arrive. Similarly, the first of the $\lceil V/D\rceil$ new data packets has to wait for $(\lceil V/D\rceil-1)\varphi$ seconds for the remaining $\lceil V/D\rceil-1$ new data packets to arrive and so on. Thus the expected per packet delay is

$$\begin{aligned} E[T|V] &\leq \frac{1}{\lceil\frac{V}{D}\rceil+1}\left[\left\lceil\frac{V}{D}\right\rceil+\left(\left\lceil\frac{V}{D}\right\rceil-1\right)+\cdots+0\right]\varphi \\ &= \frac{1}{\lceil\frac{V}{D}\rceil+1}\left[\frac{\lceil\frac{V}{D}\rceil\left(\lceil\frac{V}{D}\rceil+1\right)}{2}\right]\varphi \\ &= \frac{1}{2}\left\lceil\frac{V}{D}\right\rceil\varphi \end{aligned} \tag{10}$$

The size of each concatenated packet in bytes is uniformly distributed in the range $[1, V_m]$. Then, unconditioning on V, the expected delay due to concatenation is bounded by

$$\begin{aligned} E[T] &= E[E[T|V]] \\ &\leq \sum_{V=1}^{V_m}\frac{1}{V_m}\frac{1}{2}\left\lceil\frac{V}{D}\right\rceil\varphi \\ &\leq \frac{1}{2V_m}D\sum_{i=1}^{\lceil\frac{V_m}{D}\rceil} i\varphi \\ &= \frac{D}{2V_m}\frac{\lceil\frac{V_m}{D}\rceil\left(\lceil\frac{V_m}{D}\rceil+1\right)}{2}\varphi \end{aligned} \tag{11}$$

which proves the expression in Result 1.

Finally, we use the result above to characterize the anonymity provided by the proposed scheme, using the results from [22]. Theorem 3 of [22] provides an explicit characterization of the achievable anonymity α for a given latency introduced by an anonymizing scheme. Thus the use of the delay bound from Result 1 above into Theorem 3 of [22] provides a direct characterization of the anonymity that is provided by our scheme. The details of Theorem 3 of [22] are omitted here for brevity and because the application of Result 1 in Theorem 3 is quite straightforward.

VI. RESULTS

In this section we present simulation results to verify the performance of the proposed strategies against traffic analysis attacks. For the results reported, we simulated a multi-hop, linear network. For the simulated network, we randomly picked one synchrophasor flow and evaluated the effectiveness of the proposed strategies by obtaining the false positive and false negative rates associated with traffic analysis attacks. The false positive and false negative rates were obtained by comparing the outcome of the attacks (described below) on the chosen flow on different links in its path, as well as other flows in the network. The reported results are the average of results for 100 flows. The end to end network delay was assumed to be $T_p = 100$ msec and the maximum allowable delay for the synchrophasor data was assumed to be $\delta = 500$ msec, allowing $\Delta = 400$ msec as the maximum allowable delay due to concatenation. It was assumed that the synchrophasors generate 20 samples or data packets per second (i.e. $\varphi = 50$ msec). The length of each simulation run was 100 minutes.

A cross correlation based test was used as the timing and volume based traffic attack [5], [8]. We assume that the adversary is able to observe an arbitrary number of links in the network and its objective is to determine if a flow on a given link is the same as a flow on another link. The adversary is assumed to use both the timing and well as the traffic volume for the attacks. The statistical correlation test used by the adversary works by collating the observations into adjacent windows of fixed length W [5], [8]. For the k-th window, the adversary counts the number of packets, X_k, and the total number of bytes transmitted, Y_k, on a given link, and compares them with the corresponding values X'_{k+d} and Y'_{k+d} on another link, where the lag d is selected by the adversary. To

Attack	Dummy traffic rate		
	$\lambda = 4$	$\lambda = 10$	$\lambda = 20$
count	0.00	0.01	0.02
volume	0.05	0.08	0.04

TABLE I
CROSSOVER ERROR RATES

complete the attack, the adversary computes the correlations

$$r_x(d) = \frac{\sum_i ((X_i - \mu_x)(X'_{i+d} - \mu'_x))}{\sqrt{\sum_i (X_i - \mu_x)^2} \sqrt{\sum_i (X'_{i+d} - \mu'_x)^2}} \quad (12)$$

$$r_y(d) = \frac{\sum_i ((Y_i - \mu_y)(Y'_{i+d} - \mu'_y))}{\sqrt{\sum_i (Y_i - \mu_y)^2} \sqrt{\sum_i (Y'_{i+d} - \mu'_y)^2}} \quad (13)$$

where μ_x and μ'_x are the average number of packets on the two links while μ_y and μ'_y are the average traffic volumes. The evaluated correlations are compared against pre-set thresholds ρ_x and ρ_y and if at least one of the correlations exceeds its threshold, the adversary infers that the flows on the two links is the same. Different window (W) and lag (d) values were used in our simulations. A lag of $d = 0$ led to the most effective attacks (as also reported in [5], [8]). As in [8] we use $W = 60$ seconds, unless otherwise noted.

The effectiveness of the correlation based tests is dependent on the choice of the thresholds. It is intuitive that while high thresholds will reduce the rate of false positives, it will in turn increase the rate of false negatives, and vice versa. Existing literature has suggested the use of *crossover error rate* [5], [8] to evaluate the effectiveness of traffic analysis attacks. The crossover error rate is defined as the value of the false positive and false negative rates when the thresholds are chosen so as to make the two rates equal. The crossover error rate varies in the range $[0, 0.5]$ and a low crossover error rate signifies an effective attack.

The crossover error rates for the proposed obfuscation strategies is shown in Table I. The table shows the crossover rates for three different rates of dummy traffic (note that $\lambda = 20$ implies that the dummy traffic rate is the same as the rate of data packets generated by the synchrophasor). The table shows the results for traffic analysis attacks where only the number of packets (labeled "count") and only the data volume (labeled "volume") is considered. As expected, the performance of the obfuscation strategies improves as more dummy traffic is added, particularly for the count based attack.

Next we consider the impact of various parameters on the effectiveness of the proposed defense strategies against traffic analysis attacks. To observe the effect of various parameters, we use plots with the false positive rate as the x-axis and the false negative rate subtracted from one as the y-axis to plot the receiver operator characteristic (ROC) curves [5]. The effectiveness of the attack (equivalently, the weakness of the defense strategy) is indicated by the area under the curve, or equivalently, how close the curve is to the upper left-hand corner.

We first consider the impact of the number of hops in the path on the effectiveness of the defense strategies in Figure 1.

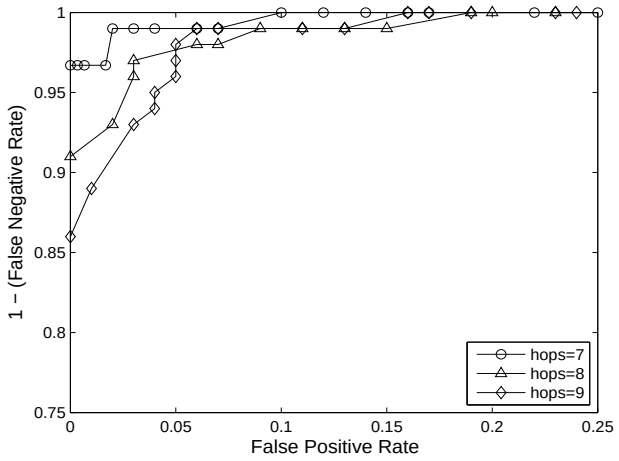

Fig. 1. False positive and false negative rates for simulation results for synchrophasor measurement data traffic with different path lengths.

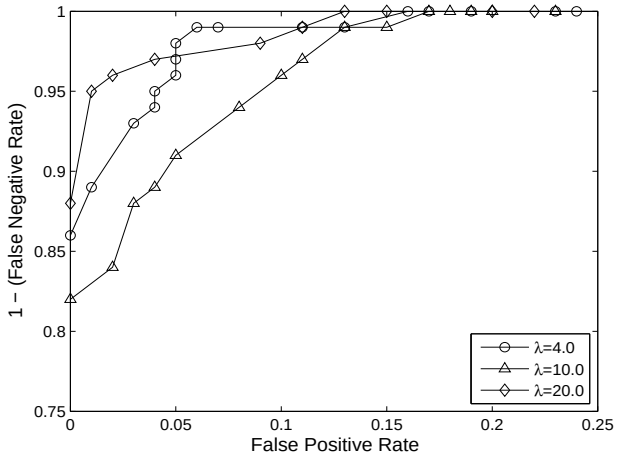

Fig. 2. False positive and false negative rates for simulation results for synchrophasor measurement data traffic with different dummy packet rates.

The results shown here are for the traffic volume based attack. The figure shows the false positive and false negative rates when the the path traversed by the synchrophasor data was 7, 8 and 9 hops. We see that the proposed obfuscation strategy performs better as the number of hops in the paths increases. This is because a longer path allows for greater variability through the random dropping strategy, thereby reducing the effectiveness of the traffic analysis attacks.

Finally, Figure 2 shows the ROC curves for the proposed scheme for different dummy traffic rates, for a path of 9 hops. As before, the results shown here are for the traffic volume based attack. We observe that there is no discernible relationship between the dummy traffic rate and the accuracy of the volume based traffic analysis attack.

VII. CONCLUSION

This paper presents a set of strategies to defeat traffic analysis attacks on networks carrying synchrophasor data. The proposed strategies defend against both data volume based and

timing based attacks through a mix of random data concatenation and random packet dropping strategies. In contrast to traditional methods of defeating traffic analysis attacks, the proposed strategies only require modifications at the end hosts and do not require specialized routers. The performance of the proposed schemes are evaluated theoretically as well as through simulations.

REFERENCES

[1] S. Horowitz, A. Phadke and B. Renz, "The Future of Power Transmission," *IEEE Power and Energy Magazine*, vol.8, no.2, pp.34-40, March-April 2010.

[2] R. Burnett, M. Butts, and P. Sterlina, "Power system applications for phasor measurement units," *IEEE Computer Applications in Power*, vol. 7, no. 1, pp. 8-13, January 1994.

[3] H. Khurana, M. Hadley, N. Lu and D. Frincke, "Smart-Grid Security Issues," *IEEE Security and Privacy*, vol. 8, no. 1, pp. 81-85, January-February 2010.

[4] A. Armenia and J. Chow, "A Flexible Phasor Data Concentrator Design Leveraging Existing Software Technologies," *IEEE Transactions on Smart Grid*, vol. 1, no. 1, pp. 73-81, June 2010.

[5] B. Levine, M. Reiter, C. Wang and M. Wright, "Timing Attacks in Low-Latency Mix Systems," *Proceedings of Financial Cryptography*, pp. 251-265, Key West, FL, February 2004.

[6] D. Chaum, "Untraceable electronic mail, return addresses, and digital pseudonyms," *Communications of the ACM*, vol. 24 no. 2, pp. 84-88, February, 1981.

[7] A. Jerichow, J. Muller, A. Pfitzmann, B. Pfitzmann and M. Waidner, "Real-Time Mixes: A Bandwidth-Efficient Anonymity Protocol," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 4, pp. 495-509, May 1998.

[8] V. Shmatikov and M.-H. Wang, "Timing analysis in low-latency mix networks: Attacks and defenses," *Proceedings of European Symposium on Research in Computer Security*, pp. 18-33, Hamburg, Germany, September 2006.

[9] O. Berthold and H. Langos, "Dummy traffic against long-term intersection attacks," *Proceedings of International Workshop on Privacy Enhancing Technologies*, pp. 110-128, 2002.

[10] H. Khurana, R. Bobba, T. Yardley, P. Agarwal and E. Heine. "Design Principles for Power Grid Cyber-Infrastructure Authentication Protocols," *Proceedings of HICSS*, Hawaii, January, 2010.

[11] P. McDaniel and S. McLaughlin, "Security and Privacy Challenges in the Smart Grid," *IEEE Security and Privacy*, Vol. 7, No. 3, pp. 75-77, May/June 2009.

[12] A. Giani, S. Sastry, K. Johansson and H. Sandberg, "The VIKING project: An initiative on resilient control of power networks," *Proceedings of the Symposium on Resilient Control Systems*, pp.31-35, August 2009.

[13] F. Cleveland, "Cyber security issues for Advanced Metering Infrastructure (AMI)," *Proceedings of IEEE Power and Energy Society General Meeting*, July 2008.

[14] M. LeMay, G. Gross, C. Gunter and S. Garg, "Unified architecture for large-scale attested metering," *Proceedings of Hawaii International Conference on System Sciences*, 2007.

[15] C. Efthymiou and G. Kalogridis, "Smart Grid Privacy via Anonymization of Smart Metering Data," *Proceedings of IEEE SmartGridComm*, pp. 238-243, October 2010.

[16] F. Garcia and B. Jacobs, "Privacy-friendly energy-metering via homomorphic encryption," *Proceedings of Workshop on Security and Trust Management*, Athens, Greece, September 2010.

[17] A. Molina-Markham, P. Shenoy, K. Fu, E. Cecchet and D. Irwin, "Private memoirs of a smart meter," *Proceedings of ACM BuildSys*, Zurich, Switzerland, November 2010.

[18] Y. Zhu, X. Fu, B. Graham, R. Bettati and W. Zhao, "Correlation-based traffic analysis attacks on anonymity networks," *IEEE Transactions on Parallel and Distributed Systems*, vol. 21, no. 7, pp. 954-967, July 2010.

[19] I. Kamwa, R. Grondin and Y. Hebert, "Wide area measurement based stabilizing control of large power systems: A decentralized/hierarchical approach," *IEEE Transactions on Power Systems*, vol. 16, no. 1, pp.136-153, February 2001.

[20] J. Stahlhut, T. Browne, G. Heydt and V. Vittal, "Latency viewed as a stochastic process and its impact on wide area power system control signals," *IEEE Transactions on Power Systems*, vol. 23, no. 1, pp. 8491, February 2008.

[21] C. Shannon, "Communication theory of secrecy systems," *Bell Systems Technical Journal*, vol. 28, no. 4, pp. 656-715, October 1949.

[22] P. Venkitasubramaniam and L. Tong, "Anonymous networking with minimum latency in multihop networks," *Proceedings of IEEE Symposium on Security and Privacy*, pp. 18-32, Oakland, CA, May 2008.

[23] R. Fano, *Transmission of Information; A Statistical Theory of Communications*, M.I.T. Press, Cambridge, MA, 1961.

[24] W. Stallings, *Data and Computer Communications*, Prentice Hall, Upper Saddle River, NJ, 2007.

[25] A. Serjantov, R. Dingledine and P. Syverson, "From a Trickle to a Flood: Active Attacks on Several Mix Types," *Proceedings of Information Hiding Workshop*, pp. 36-52, October 2002.

PLACE PHOTO HERE

Biplab Sikdar Biplab Sikdar received the B. Tech degree in electronics and communication engineering from North Eastern Hill University, Shillong, India, M. Tech degree in electrical engineering from Indian Institute of Technology, Kanpur and Ph.D in electrical engineering from Rensselaer Polytechnic Institute, Troy, NY, USA in 1996, 1998 and 2001, respectively. He is currently an Associate Professor in the Department of Electrical, Computer and Systems Engineering of Rensselaer Polytechnic Institute, Troy, NY, USA. His research interests include wireless MAC protocols, transport protocols, network security and queuing theory. Dr. Sikdar is a member of Eta Kappa Nu and Tau Beta Pi and is an Associate Editor of the IEEE Transactions on Communications.

PLACE PHOTO HERE

Joe H. Chow Joe H. Chow is a Professor of Electrical, Computer, and Systems Engineering at Rensselaer Polytechnic Institute, Troy, New York. He received his BS degrees in Electrical Engineering and Mathematics from the University of Minnesota, and the MS and PhD degrees from the University of Illinois, Urbana-Champaign. He worked at General Electric Company before joining RPI in 1987. His current research includes Voltage-Sourced Converter based Flexible AC Transmission Systems, and the analysis, modeling, and control of large power systems using synchronized phasor measurements.