Blockchain Based Decentralized Replay Attack Detection for Large Scale Power Systems

Paritosh Ramanan*†,Dan Li† and Nagi Gebraeel†

Abstract—Large scale power systems are comprised of regional utilities with assets that stream sensor readings in real time. In order to detect cyberattacks, the globally acquired, real time sensor data needs to be analyzed in a centralized fashion. However, owing to operational constraints, such a centralized sharing mechanism turns out to be a major obstacle. In this paper, we propose a blockchain based decentralized framework for detecting coordinated replay attacks with full privacy of sensor data. We develop a Bayesian inference mechanism employing locally reported attack probabilities that is tailor made for a blockchain framework. We compare our framework to a traditional decentralized algorithm based on the broadcast gossip framework both theoretically as well as empirically. With the help of experiments on a private Ethereum blockchain, we show that our approach achieves good detection quality and significantly outperforms gossip driven approaches in terms of accuracy, timeliness and scalability.

Index Terms—blockchain, decentralized analytics, data privacy, power systems, global replay attacks

I. INTRODUCTION

Affordable sensor and communication technologies have given rise to a growing wave of industrial digitization. The power industry has been at the forefront of this trend that has culminated into a digital transformation of the power grid. Such levels of digitization have led to automation and digital control components that are collectively referred to as Industrial Control Systems (ICS). Until recently, ICS used specialized communication and control protocols that made them relatively immune to cyber-attacks. However, with increase in Industrial Internet-Of-Things (IIoT) enabled assets, traditional ICS have gradually become heavily integrated with standard IT components. Meanwhile, on the physical level, the grid has evolved into a complex network with a high degree of interdependency among utility providers [1], [2]. Such an interdependent, digitized grid has an increased vulnerability to various kinds of cyberattacks [3]. Tackling these vulnerabilities requires competing utilities to share sensitive information with a trusted, centralized entity that can quickly assess cybersecurity related threats. The process of sharing data with a centralized entity can be challenging for many utilities due to the presence of a single point of failure, privacy concerns and competing market dynamics. Therefore, this paper proposes a blockchain based, decentralized methodology for detecting the probability of a global network cyberattack while preserving data privacy. Our approach is particularly useful in the context of IIoT enabled assets that yield real time sensor data to help monitor the power network.

A significant portion of ICS-focused cyberattacks involve data manipulation. Often the intent of these attacks is to impact asset reliability either by accelerating physical or efficiency degradation or causing sudden breakdowns. The *Stuxnet* worm has often been referenced as a classic example of an ICS-focused cyberattack with data manipulation [4], [5]. Another popular example was the Aurora Generator Test in which a 2.25 MW substation generator was destroyed through a planned cyberattack that caused an out-of-sync closing of protective relays [6], [7]

ICS cyberattacks involving data manipulation have been classified into three major types, false data injections [8], replay attacks [9], [10] and covert attacks [11]. This paper considers ICS replay attacks where a malicious agent replays sensor measurements representing normal operating conditions in order to mask underlying malicious control actions. We focus on investigating coordinated, large scale scenarios where an ICS replay attack is mounted on more than one regional utility provider in the power network. We refer to such attacks as global network cyberattacks, or global attacks for short. ICS replay attacks are difficult to detect. Most of the existing detection algorithms do not distinguish between replay cyberattacks and naturally occurring equipment or controller faults.

Current state-of-the-art ICS attack detection methods are intended for individual plant sites and assets only. Utility stakeholders typically rely solely on such localized detection mechanisms to raise alarms. However, without the knowledge of a rapidly evolving state of global cyber health, the local mechanism will only raise an alarm when local assets are either i) under an attack themselves, or ii) are facing the system repercussions of a network-wide attack. By the time a local alarm has been raised, utilities would have lost precious time which could otherwise be used to localize, diagnose and isolate faults. Moreover, local detection schemes are inherently based on some underlying hypothesis test, which means that there is always some level of unavoidable false alarms. In large scale settings like the one considered in this paper, detection schemes could significantly overstate the cyber threat level since false alarm rates increase with the number of hypotheses tests being conducted [12]. Without active collaboration with other stakeholders, utilities might waste resources trying to diagnose false alarms unnecessarily. Therefore, a major stumbling block for global attack detection arises due to the inability of different utility stakeholders to collaborate and share their local ICS attack detection insights and alarms. Such roadblocks are inevitable despite the fact that stakeholders might be equipped with state of the art ICS attack detection mechanisms.

Consequently, utility stakeholders are left with a tough choice

*School of Computational Science and Engineering, Georgia Institute of technology, Atlanta, GA, USA 30332
†H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of technology, Atlanta, GA, USA 30332.
paritoshpr@gatech.edu,danli@gatech.edu,nagi@isye.gatech.edu

between one of two options. Either, they pursue a purely localized approach with state of the art ICS attack detection, not withstanding its various blinding limitations. Alternatively, they could adopt a real time sensor and cyber incident data sharing strategy that relies on a centralized coordinator. The centralized entity would be in charge of real time processing and analytics of sensor data being streamed from a large number of assets, while providing alerts pertaining to the global cyber health of the network as a whole. Many such centralized data sharing programs have been pursued in the past. For instance, the U.S. Department of Energy operates the Cyber Risk Information Sharing Program (CRISP) [13] which provides utility members with a platform to share cyber incident/sensor data.

However, such centralized data sharing programs come with their own set of challenges. First, sharing data through a centralized repository presents several privacy risks along with efficiency and agility issues [14] with respect to sensor data. Second, in the absence of legal obligation to participate, many critical utility providers do not pursue membership of data sharing cooperatives like CRISP due to a variety of business as well as logistical reasons. Third, programs like CRISP also require hardware upgrades to existing IT infrastructure and associated costs as well [13]. Lastly, stakeholders might harbor inherent distrust in the centralized coordinator itself which is usually the federal government agencies and/or regulatory bodies. Such distrust could be due to apprehensions pertaining to unintended use of sensor data by regulators to detect violations of critical infrastructure rules and impose penalties.

The unique characteristics of the blockchain could be leveraged to achieve the best of both worlds: the privacy and agility of a purely localized approach coupled with the accuracy and robustness of the centralized method. The blockchain could allow stakeholders to use state of the art ICS attack detection mechanisms locally while being able to share their alarm statistics with all other stakeholders in real time. Further, the underlying consensus mechanism of the blockchain would be useful in estimating the global cyber health status by pooling local alarms in a fully trustworthy and transparent fashion. While the blockchain provides a robust computational platform for global aggregation of alarms, the local ICS attack detection algorithms would eliminate the need to move low level sensor data ensuring full data privacy. As a result, stakeholders would be more amenable to participating in such a decentralized mechanism that provides prompt insights while potentially reducing false alarm rates drastically. The significant reduction in setup and maintenance costs due to elimination of a centralized aggregator [15] is an understated advantage of a blockchain driven approach.

In this paper, we propose the use of a permissioned blockchain for detecting globally coordinated replay attacks in a decentralized fashion. Blockchains typically rely on a consensus among multiple mistrusting parties to achieve a consistent global state. As a result, blockchain based platforms are decentralized in nature and do not involve a centralized chain of command. Our framework therefore ensures data privacy by allowing individual utilities to run their own detection algorithms locally. This leads to full ownership of cyber incident sensor data with complete privacy. It also eliminates the associated cost of setting up a trusted, centralized third party. Moreover, we utilize blockchain driven Smart Contracts (SC) [16] to estimate the likelihood of a global replay attack based on alarms and insights aggregated from the various utilities.

In order to demonstrate the blockchain's ability to rapidly generate and propagate global insights, we perform an in depth comparison with well established information diffusion mechanisms like Broadcast Gossip (BG). Gossip protocols are an important type of diffusion technique aimed at estimating the global state of a system in a peer-to-peer fashion [17], [18]. As a result, in this paper, we develop a BG based global replay attack detection detection framework for benchmarking purposes. We compare and contrast our blockchain driven approach against the BG based framework in a theoretical as well as in an empirical manner. Our results indicate that the blockchain provides a sound computational platform allowing for global aggregation of local outputs in a timely, accurate and reliable fashion. The major contributions of this work are summarized in detail below:

- We develop a decentralized mechanism that relies on Bayesian inference in order to detect a globally coordinated replay attack with full regional data privacy.
- We introduce Theorem III.1, specifically geared towards maintaining computational efficiency of Bayesian inference on a blockchain platform.
- With the help of Theorem III.1, we design a blockchain based framework for computing the Global Attack Probability (GAP) with only one global multiplication and addition steps.
- We propose the BG framework as a benchmark and theoretically compare its performance with the blockchain based approach with the help of Theorem V.1.
- We implement our framework on an Ethereum based private blockchain network to demonstrate its scalability and applicability for varying degrees of cyber threat parameters.

Our results conclusively show that the blockchain driven framework is vastly superior to conventional, state-of-the-art information diffusion paradigms, like BG, both in terms of computational performance as well as accuracy of results.

II. RELATED WORK

ICSs are a special class of information systems, involving interaction between information technology (IT) and operational technology (OT) [19]. Like all information systems, cyberattacks also threaten many ICSs involved in applications such as water supply, manufacturing, power systems, and energy [20]–[22]. Attacks like DoS, DDoS, phishing, spoofing and eavesdropping that target generic IT systems can also target ICSs but they can often be effectively detected and isolated by monitoring network traffic [23]–[25]. On the other hand, data attacks are a more significant threat to ICSs due to their interaction between information, communication and the underlying physical processes [26] that could potentially lead to critical infrastructure damage.

A. Data Manipulation Attack Detection

In data manipulation attacks, attackers manipulate controller/sensor data or even data at rest in order to damage critical assets through malicious control actions and incorrect state estimations leading to degraded asset performance. It has been shown that data attacks can be designed to bypass basic verification methods relying on Cyclic Redundancy Check (CRC), User Datagram Protocol (UDP) and Transmission Control Protocol (TCP). [27], [28].

Numerous model-based detection frameworks have been proposed as an added layer of protection for ICS attacks involving

data manipulation [8], [10], [29]–[39]. Most detection algorithms rely on differences between actual measurements and those estimated by a model of the physical system. Differences between the estimated (or predicted) and observed measurements or states (i.e., residuals) can be used to detect possible cyberattacks. In most cases, a sequential goodness-of-fit testing procedure serves as the basis for detecting the attack. However, these approaches have been shown to be inefficient in detecting replay attacks, primarily because the observed measurements (replayed data) often match measurements estimated by the system's model.

Replay attacks can be detected based on several types of strategies. The first type is watermarking the control actions [10], [34], [40]–[43]. This type of strategy is based on the assumption that the operator can add watermarks (a small bias or noise) to the control actions, which is unknown to the attacker. The second type is observer-based anomaly detection [44], [45]. This type of strategy is based on the assumption that some of the sensors can be protected or immune from manipulation. The local attack detection algorithm we adopt in this paper [45] belongs to the second type, however, it does not assume the operators can proactively protect the sensors, but that some certain types of sensors are immune to manipulation. In fact, any replay attack detection technique can be adapted to our global framework, as long as the detection algorithm can be used to obtain the probability of attack. The advantage of adopting the method in [45] is that it detects the attack based on statistical inference, which naturally outputs the probability of attack.

B. Blockchain Driven Methods for Security and Privacy

The use of blockchain has been proposed as a gateway to ensuring data privacy and security in a wide variety of application areas. A lightweight, private blockchain paradigm for enhancing the security and privacy of IIoT driven manufacturing platforms has been proposed in [46]. Further, [47] provides a secure, permissioned blockchain based approach for energy trading between electric vehicles and the grid. Use of blockchain has been considered in power systems in numerous works in the recent past. Blockchains have been proposed as a means to establish a decentralized, secure technique for transactive energy [48] as well as to handle and trace back energy losses in microgrids that incorporate PV nodes [49]. Blockchain driven approaches are also being considered as the perfect computational platforms for collaboratively detecting attacks and anomalies on a global scale with full data privacy [50]. However, despite its immense potential, efforts exploring the use of the blockchain towards detecting globally coordinated replay attacks on power network ICSs largely remains understated.

C. Operational Dynamics of Blockchains

The operational dynamics of a blockchain implementation are dictated by their individual application or use case. In cases where no assumptions can be made about the identity or intent of the participating entities, a permissionless blockchain implementation is the norm [51]. On the other hand, in applications where the parties are well-established real-world entities which elicit a fair degree of trust, a permissioned blockchain implementation is preferred [52]. The major distinguishing factor between a permissioned and a permissionless blockchain is the underlying consensus protocol which ensures a consistent global state among all parties. Permissionless blockchain implementations involve computationally intensive consensus protocols that are designed to operate without any assumption of trust among the mistrusting parties. Popular consensus mechanisms for permissionless blockchains include Proof-of-Work (PoW), Proof-of-Stake (PoS), delegated Proof-of-Stake (dPoS), Proof-of-ElapsedTime (PoET) [52]. However, permissioned blockchains place an emphasis on high performance involving consensus protocols that leverage the identity of participating entities. In permissioned blockchains, consensus protocols are primarily a fault-tolerant means of achieving global consistency by inducing a tamper proof record [52].

D. Consensus Mechanisms for Permissioned Blockchains

Consensus mechanisms in permissioned settings are usually dominated by Proof-of-Authority (PoA) based algorithms. Rather than relying on solutions to complex mathematical-based challenges like PoW, PoA relies on the authority of real-world entities part of a permissioned ledger. A majority of the authorities have to achieve consensus before a block is permanently added to the chain. As a result, PoA improves security without relying on computational challenges, since an attacker must hack a majority of authorities in order to rescind all transactions [53]. PoA chains are also known to have low latency, deterministic block creation process, as well as faster block creation times which are more important in a permissioned setting [52], [53]. PoA is more suited to a consortium or permissioned settings due to its security, easier maintenance as well as accountability of the authorities themselves [53]. Therefore, in this paper, we adopt a permissioned blockchain driven approach since our problem setting involves a *consortium of real-world entities* such as utility stakeholders with their own ICSs desirous of detecting a global replay attack.

Most PoA algorithms are based on the concept of Byzantine fault tolerance (BFT) [51], [52]. However, non-BFT algorithms like Raft [52] have also been proposed and form an integral part of many permissioned blockchain implementations such as Corda and GoQuorum. The class of BFT driven PoA mechanisms can in turn be divided into two categories: classical and modern. The classical PoA category consists of protocols such as Istanbul BFT [54] (offered in Hyperledger Besu, and GoQuorum) with the consistency of the ledger as the priority. The modern category consists of a new family of BFT protocols such as Clique (offered in Hyperledger Besu, Ethereum, OpenEthereum and GoQuorum) and Aura [51] (offered in OpenEthereum). The modern PoA protocols emphasize better performance (i.e. low latency and block creation times) over consistency.

We summarize the characteristic features of widely used consensus protocols i.e. Clique, IBFT, Aura and Raft with the help of Table I. Table I provides a qualitative comparative analysis of each protocol. For an in-depth discussion of the characteristic features presented in Table I, we refer the reader to Appendix VII. Due to its low latency, high decentralization potential, increased degree of fault tolerance as well as the capability to achieve eventual consistency, we choose the Clique consensus protocol for our experiments.

III. PROBLEM FORMULATION

We primarily use the blockchain as a computational platform for aggregating outputs from local replay attack detection algorithms. In developing our blockchain based global attack detection framework, we consider a power network that is divided topologically into a set of distinct regions denoted by $\mathcal{R} = \{1, 2, 3 \ldots n\}$. Each region can be thought of as a utility provider with multiple power plants. A blockchain architecture is used to assess the

TABLE I: Comparative Analysis of State-of-the-art Consensus Protocols for Permissioned Blockchains

protocol	PoA type	validators	decentralization	fault tolerance	consistency	block latency
Clique	modern BFT	at least 1	high	$< 1/2$ validators	eventual consistency	low; only 1 message round required
IBFT	classical BFT	at least 4	medium	$< 1/3$ validators	consistent but stalling likely	high; 3 message rounds required
Aura	modern BFT	at least 1	high	$< 1/2$ validators	no consistency guarantees	medium; 2 message rounds required
Raft	non BFT	exactly 1 leader, multiple verifiers, many learners	low; restricted to leader	$< 1/2$ crash faults	consistent (assumes no forking)	low; only leader can mint blocks

global state of the entire network based on insights that are aggregated in a decentralized, privacy preserving manner. The aggregation determines the overall probability of a global network attack.

We first consider a single region i comprised of p generators monitored and controlled by a single ICS where the state of each generator can be represented by m variables. We assume that the dynamics of this system of generators can be characterized by a linear time-invariant (LTI) model at time t described by Equations (1), (2).

$$x_{t+1} = Ax_t + Bu_t + v_t, \tag{1}$$
$$y_t = Cx_t + w_t, \tag{2}$$

In Equations (1), (2), $x_t \in \mathbb{R}^{mp}$ represents the unobserved state of the system, $u_t \in \mathbb{R}^{mp}$ is the control action while $y_t \in \mathbb{R}^{mp}$ denotes the sensor measurements, which are assumed to be noisy realizations of the state. $A \in \mathbb{R}^{mp \times mp}$ is the state transition matrix and represents the system dynamics. $B \in \mathbb{R}^{mp \times mp}$ is the input matrix and represents how the control action impacts the system state. C is the measurement matrix and represents the relationship between the measurements y_t and the state x_t. $v_t \in \mathbb{R}^{mp}$ and $w_t \in \mathbb{R}^{mp}$ represent process and measurement noise at time t, respectively. Such a type of modeling framework has also been used extensively in prior art [10], [11], [30].

In this setting, the Kalman filter is known to be the optimal state estimator [55]. The residuals r_t are defined as the differences between the actual measurements y_t and the predicted measurements $C\hat{x}_{t|t-1}$ as estimated by the Kalman filter. The control action u_t is calculated using a linear-quadratic Gaussian (LQG) controller, which is the optimal controller under the LTI setting [56]. The Kalman filter and the LQG controller are estimated based on Equations (3)-(6).

$$\hat{x}_{t|t-1} = A\hat{x}_{t-1|t-1} + Bu_{t-1}, \tag{3}$$
$$\hat{x}_{t|t} = \hat{x}_{t|t-1} + Kr_t, \tag{4}$$
$$r_t = y_t - C\hat{x}_{t|t-1}, \tag{5}$$
$$u_t = L\hat{x}_{t|t}. \tag{6}$$

In Equation (6), $L = -(B^TSB + U)^{-1}B^TSA$ where S is the solution to the Ricatti equation [56] denoted by Equation (7).

$$S = A^TSA + W - A^TSB(B^TSB + U)^{-1}B^TSA \tag{7}$$

In Equation (7), $W, U \in \mathbb{R}^{mp \times mp}$ are positive semidefinite

matrices used by the LQG controller to minimize its objective function $J = E[x_t^T W x_t + u_t^T U u_t]$ based on state variables and control actions, respectively. Therefore, W, U primarily denote the degree of aggression while regulating the state variables.

A. Threat Model

As mentioned in Section I, in a replay attack, the attacker gains access to both sensors and controller. This can be achieved by intruding into the ICS through its supervisory computers or other connected hardware. After the access is obtained, the attack is implemented in two stages. The first stage involves eavesdropping and recording of sensor measurements $\{y_t^r\}$ without data manipulation, which can be denoted by

$$y_t^r = y_t. \tag{8}$$

The second stage is the manipulation of the control actions u_k by injecting bias or noise, while replaying the sensor measurements y_t^r in replacement of y_t. The second stage is represented by Equations (9) and (10)

$$u'_t = u_t + a_t, \tag{9}$$
$$y'_t = y_{t-\delta}^r, \tag{10}$$

where u'_t and y'_t are the control action and measurements under attack, respectively, a_t is the bias or noise injected at time t, and δ is the time difference between the onset of the two stages of the replay attack. In this way, the system state is altered due to Equation (9), but the attack is hard to detect by monitoring the sensor data due to Equation (10).

In our framework, we assume that the adversary is outside the blockchain and is solely focused on injecting data into a utility stakeholder's ICS in order to carry out a replay attack. In our framework, the stakeholders themselves trust each other and share the insights from their local ICS detection algorithms with each other through the permissioned blockchain. Furthermore, we also assume that an attacker targets at most less than 50% of the regional utility stakeholders participating in the consortium. Therefore, based on the PoA protocol [51], we can state that an attacker controlling hacked authorities cannot revert transactions and attack the network. We assume that all regions are equally likely to experience a replay attack which are mutually independent. Independence is assumed here for mathematical convenience. More importantly, a global network attack is assumed to occur when at least two regions report local attacks.

Given our threat model, the problem formulation can be viewed as having two components, a local and a global component. The local component represents regional plants that belong to a single utility provider. We assume that each region executes the local algorithms aimed at detecting replay attacks on their plants' ICS. We refer to the local component as the *regional detection model*. The global component, i.e. the *network detection model*, concerns the detection of a coordinated global attack at the network level.

B. Regional Detection Model

As shown in [45], replay attacks can be detected by monitoring the covariance matrix of the residuals r_t. Using this approach, let σ^i be a random variable representing the probability of an alarm triggered by region i. That is, $\sigma^i = 1$ if a replay attack is detected and $\sigma^i = 0$ otherwise. We define $\hat{\sigma}^i$ as the ground truth that represents whether a replay attack is indeed underway. There are two classic errors that can occur in this setting. Type-I error, α, represents the probability of a false alarm, i.e., the algorithm triggers an alarm when there is no attack. The Type-II error, β, represents the probability of a false negative, i.e., where the detection algorithm fails to detect a true replay attack. These errors can be defined more formally in terms of a region i as follows

$$\alpha_i = Pr(\sigma^i = 1|\hat{\sigma}^i = 0), \tag{11}$$

$$\beta_i = Pr(\sigma^i = 0|\hat{\sigma}^i = 1). \tag{12}$$

In Equations (11), (12) $\hat{\sigma}^i = 1$ in the event the system is truly under a replay attack whereas $\hat{\sigma}^i = 0$ otherwise.

C. Network Detection Model

Consider a power network comprised of n regions each reporting an alarm based on their local belief of an attack. Recall that a global network attack is triggered if there are two or more distinct regional alarms, i.e. at least two regions detect an attack. Let the set $S = \{s_0, s_1 \ldots s_n\}$ denote all the scenarios that represent no global network attack, where $s_0 \in \{0\}^n$ indicates a scenario where no regional alarms have been triggered and $s_i \in \{0,1\}^n$ denotes scenarios where only one region i triggers an alarm. Consequently, for a scenario s_k, $k > 0$, only its kth element $s_k^k = 1$. Next, let $\boldsymbol{\sigma} \in \{0,1\}^n = [\sigma^1, \sigma^2 \ldots \sigma^n]$ be a vector of random variables representing regional alarm events. Given $\boldsymbol{\sigma}$, the probability of no global network attack can be expressed as $Pr(S|\boldsymbol{\sigma})$. Similarly, $Pr(\boldsymbol{\sigma}|s_k)$ defines the probability of observing $\boldsymbol{\sigma}$ given a scenario s_k Since regional attacks are assumed to be independent, the following expressions hold true: (13)-(16).

$$Pr(s_k) = \prod_{i=1}^{n} Pr(s_k^i), \tag{13}$$

$$Pr(\boldsymbol{\sigma}|s_k) = \prod_{i=1}^{n} Pr(\sigma^i|s_k^i), \tag{14}$$

$$Pr(S|\boldsymbol{\sigma}) = \sum_{s_k \in S} \frac{Pr(\boldsymbol{\sigma}|s_k)Pr(s_k)}{Pr(\boldsymbol{\sigma})}, \tag{15}$$

$$Pr(\boldsymbol{\sigma}) = Pr(\sigma^1 \cap \sigma^2 \ldots \cap \sigma^n) = \prod_{i=1}^{n} Pr(\sigma^i). \tag{16}$$

The occurrence of the event denoted by s_k^i is only relevant in the context of the ground truth $\hat{\sigma}^i$. Therefore, we can state that $Pr(s_k^i) = Pr(\hat{\sigma}^i = s_k^i)$, where $Pr(\hat{\sigma}^i = s_k^i)$ reflects the prior probability of the existing ground truth $\hat{\sigma}^i$ being equal to s_k^i. Consequentially, substituting Equations (13) and (14) in (15), we obtain the relation represented by (17).

$$\sum_{s_k \in S} Pr(\boldsymbol{\sigma}|s_k)Pr(s_k) = \sum_{s_k \in S} \prod_{i=1}^{n} Pr(\sigma_i|\hat{\sigma}^i = s_k^i)Pr(\hat{\sigma}^i = s_k^i). \tag{17}$$

We also know that

$$Pr(\sigma^i) = Pr(\sigma^i|\hat{\sigma}^i = 0)Pr(\hat{\sigma}^i = 0) + Pr(\sigma^i|\hat{\sigma}^i = 1)Pr(\hat{\sigma}^i = 1). \tag{18}$$

We note that $Pr(\sigma^i)$ and $Pr(\sigma_i|s_k^i)Pr(\hat{\sigma}^i = s_k^i)$ in Equations (17), (18), respectively, can be computed locally allowing us to propose the the following theorem.

Theorem III.1. *The probability of no global attack denoted by* $Pr(S|\sigma)$ *is governed by the equality,*

$$Pr(S|\boldsymbol{\sigma}) = \Big(\prod_{i=1}^{n} \frac{a_i}{Pr(\sigma^i)}\Big)\Big(1 + \sum_{i=1}^{n} \frac{b_i}{a_i}\Big)$$

where,

$$a_i = Pr(\sigma^i|\hat{\sigma}^i = 0)Pr(\hat{\sigma}^i = 0), \ \ b_i = Pr(\sigma^i|\hat{\sigma}^i = 1)Pr(\hat{\sigma}^i = 1)$$

Proof. Based on our assumptions, we derive the relations represented by Equations (19), (20).

$$Pr(\boldsymbol{\sigma}|s_0)Pr(s_0) = \prod_{i=1}^{n} Pr(\sigma^i|\hat{\sigma}^i = 0)Pr(\hat{\sigma}^i = 0), \tag{19}$$

$$Pr(\boldsymbol{\sigma}|s_k)Pr(s_k) = \prod_{i=1}^{n} Pr(\sigma^i|\hat{\sigma}^i = s_k^i)Pr(\hat{\sigma}^i = s_k^i). \tag{20}$$

Dividing Equations (20) with (19), we get

$$\psi_k = \frac{Pr(\boldsymbol{\sigma}|s_k)Pr(s_k)}{Pr(\boldsymbol{\sigma}|s_0)Pr(s_0)} = \prod_{i=1}^{n} \frac{Pr(\sigma^i|\hat{\sigma}^i = s_k^i)Pr(\hat{\sigma}^i = s_k^i)}{Pr(\sigma^i|\hat{\sigma}^i = 0)Pr(\hat{\sigma}^i = 0)}. \tag{21}$$

Since, $s_k^i = 1$ only when $i = k$, we can rewrite Equation (21) as

$$\psi_k = \Big[\prod_{\substack{i=1 \\ i \neq k}}^{n} \frac{Pr(\sigma^i|\hat{\sigma}^i = 0)Pr(\hat{\sigma}^i = 0)}{Pr(\sigma^i|\hat{\sigma}^i = 0)Pr(\hat{\sigma}^i = 0)}\Big] \frac{Pr(\sigma^k|\hat{\sigma}^k = 1)Pr(\hat{\sigma}^k = 1)}{Pr(\sigma^k|\hat{\sigma}^k = 0)Pr(\hat{\sigma}^k = 0)}. \tag{22}$$

Equations (21), (22) imply that,

$$Pr(\boldsymbol{\sigma}|s_k)Pr(s_k) = Pr(\boldsymbol{\sigma}|s_0)Pr(s_0)\frac{b_k}{a_k}, \quad \forall k > 0. \tag{23}$$

Recall from Equation 17, that

$$Pr(\boldsymbol{\sigma}|s_0)Pr(s_0) = \prod_{i=1}^{n} Pr(\sigma^i|\hat{\sigma}^i = 0)Pr(\hat{\sigma}^i = 0) = \prod_{i=1}^{n} a_i. \tag{24}$$

Since a_k, b_k can be purely computed by region k we obtain Equation (25) by combining Equations (23) and (24)

$$\sum_{s_k \in S} Pr(\boldsymbol{\sigma}|s_k)Pr(s_k) = \Big(\prod_{i=1}^{n} a_i\Big)\Big(1 + \sum_{i=1}^{n} \frac{b_i}{a_i}\Big). \tag{25}$$

Based on Equations (25) and (16), we obtain

$$Pr(S|\boldsymbol{\sigma}) = \frac{\sum\limits_{s_k \in S} Pr(\boldsymbol{\sigma}|s_k)Pr(s_k)}{\prod\limits_{i=1}^{n} Pr(\sigma^i)} = \Big(\prod_{i=1}^{n} \frac{a_i}{Pr(\sigma^i)}\Big)\Big(1 + \sum_{i=1}^{n} \frac{b_i}{a_i}\Big)$$

.   □

The entities a_i, b_i can be computed in a simplified manner as follows

$$a_i = Pr(\sigma^i, \hat{\sigma}^i = 0) = Pr(\sigma^i | \hat{\sigma}^i = 0) Pr(\hat{\sigma}^i = 0), \quad (26)$$

$$b_i = Pr(\sigma^i, \hat{\sigma}^i = 1) = Pr(\sigma^i | \hat{\sigma}^i = 1) Pr(\hat{\sigma}^i = 1). \quad (27)$$

Moreover, the prior distribution can be updated in a purely local fashion using:

$$Pr(\hat{\sigma}^i = 0) = \frac{a_i}{a_i + b_i}, \quad (28)$$

$$Pr(\hat{\sigma}^i = 1) = \frac{b_i}{a_i + b_i}. \quad (29)$$

Theorem III.1 indicates that the global attack probability can be computed through one global multiplication and addition of the two terms, $\frac{a_i}{Pr(\sigma^i)}, \frac{b_i}{a_i}$. We therefore incur significant computational benefits especially on a blockchain based framework where computation is expensive. Figure 1 represents the system architecture for our blockchain based global attack detection mechanism. At the lowest level, we have plants and their associated sensors feeding data into controllers that are directly managed by the Industrial Control System (ICS). The ICS stores the data in a local data store on which the regional detection algorithm is applied. Insights from the regional detection model are shared using Theorem III.1 on the blockchain. Every regional utility stakeholder utilizes the latest available information on the ledger to be able to obtain knowledge of a global replay attack.

IV. BLOCKCHAIN BASED FRAMEWORK FOR GLOBAL REPLAY ATTACK DETECTION

The introduction of Smart Contracts (SCs) [16], [57] have been an important addition to the blockchain paradigm. An SC is typically a snippet of code that resides on the blockchain. It can contain complex program logic on the blockchain and can be invoked by any party having access to the blockchain. Once invoked, an SC is self-triggering and proceeds to alter the state of the ledger with the help of the underlying consensus protocol. As a result, an SC can be used for executing business logic through consensus among mistrusting parties paving the way for a decentralized application. Therefore, due to their versatility, blockchain driven Smart Contracts (SC) provide an ideal environment for the Network Detection model in a fully decentralized fashion.

Solidity [1] is a popular SC oriented programming language that can be leveraged for developing highly versatile decentralized applications. A key constraint of languages like Solidity is the lack of floating point arithmetic operations [2]. This is done to primarily reduce the computational burden on the underlying consensus protocols. As a result, one has to define a precision conversion factor for converting floating point values to integers before feeding them as inputs to the SC on the front end side. Obviously, the choice of the precision factor could have a wide impact on the detection accuracy of a global replay attack. In this section, we propose a Solidity based Smart Contract(SC) design that embodies our global replay attack detection paradigm discussed in Section III.

[1] https://github.com/ethereum/solidity
[2] https://solidity.readthedocs.io/en/v0.5.3/types.html#fixed-point-numbers

A. Smart Contract Design

Table II, III depict the attributes and functions that form an integral part of the SC design for our framework. In Table II, the precision factor is denoted by D and arrays $aOPs, bOa$ store the integer values $\frac{Da_i}{Pr(\sigma^i)}, \frac{Db_i}{a_i}$ of all regions respectively. In Table III, the functions $updateData, aggregateValues$ can be asynchronously invoked by any region for updating the SC with their local values and obtaining the corresponding global aggregates respectively.

TABLE II: Solidity based Smart Contract Attributes

Attribute	Type	Description
D	uint256	Precision for floating point conversion.
n	uint	The total number of regions i.e. $|\mathcal{R}|$
$aOPs[]$	uint256	Array for storing $\frac{Da_i}{Pr(\sigma^i)}, \forall i \in \mathcal{R}$
$bOa[]$	uint256	Array for storing $\frac{Db_i}{a_i}, \forall i \in \mathcal{R}$

TABLE III: Solidity based Smart Contract Function

Function	Invoker	Description
$updateData$	Region i	sets $aOPs[i], bOa[i]$
$aggregateValues$	Region i	returns $\prod_{j=1}^{n} aOPs[i], \sum_{j=1}^{n} bOa[i]$

B. Blockchain Based Global Attack Detection Algorithm

Algorithm 1 presents the details of blockchain based global replay attack detection which is executed in a decentralized fashion. We assume an off chain interaction informs all the regions about the SC address, total number of regions as well as the precision factor. Every region initially determines its corresponding local alarm value σ^i. Let $x_i = \frac{a_i}{Pr(\sigma^i)}, y_i = \frac{b_i}{a_i}$ represent the local statistical values which are converted to integers using the precision factor D and pushed to the SC by invoking $SC.updateData$. Any region can asynchronously invoke $SC.aggregateValues$ in order to obtain $x^b = \prod_{i=1}^{n} Dx_i$ and $y^b = \sum_{i=1}^{n} Dy_i$. A global estimate of no attack can then be computed locally using Theorem III.1. A complement of the result can be used to infer the presence of a global attack. At each epoch, the prior distribution pertaining to local alarm values gets updated locally.

Algorithm 1 Decentralized blockchain based algorithm

owner region deploys and initializes SC on blockchain
for i=1,2,3…n **in parallel do**
    initialize D, n and obtain SC address.
    while true **do**
        determine the message σ^i
        compute $Pr(\sigma^i)$ using Equation (18)
        compute x_i, y_i using Equations (26),(27)
        invoke $SC.updateData\left(Dx_i, Dy_i, i\right)$
        $x^b, y^b \leftarrow SC.aggregateValues()$
        using Theorem III.1 compute $1 - Pr(S|\boldsymbol{\sigma})$ such that,
            $Pr(S|\boldsymbol{\sigma}) = x^b(D + y^b)/D^{(n+1)}$
        update prior distribution using Equations (28),(29)
    end while
end for
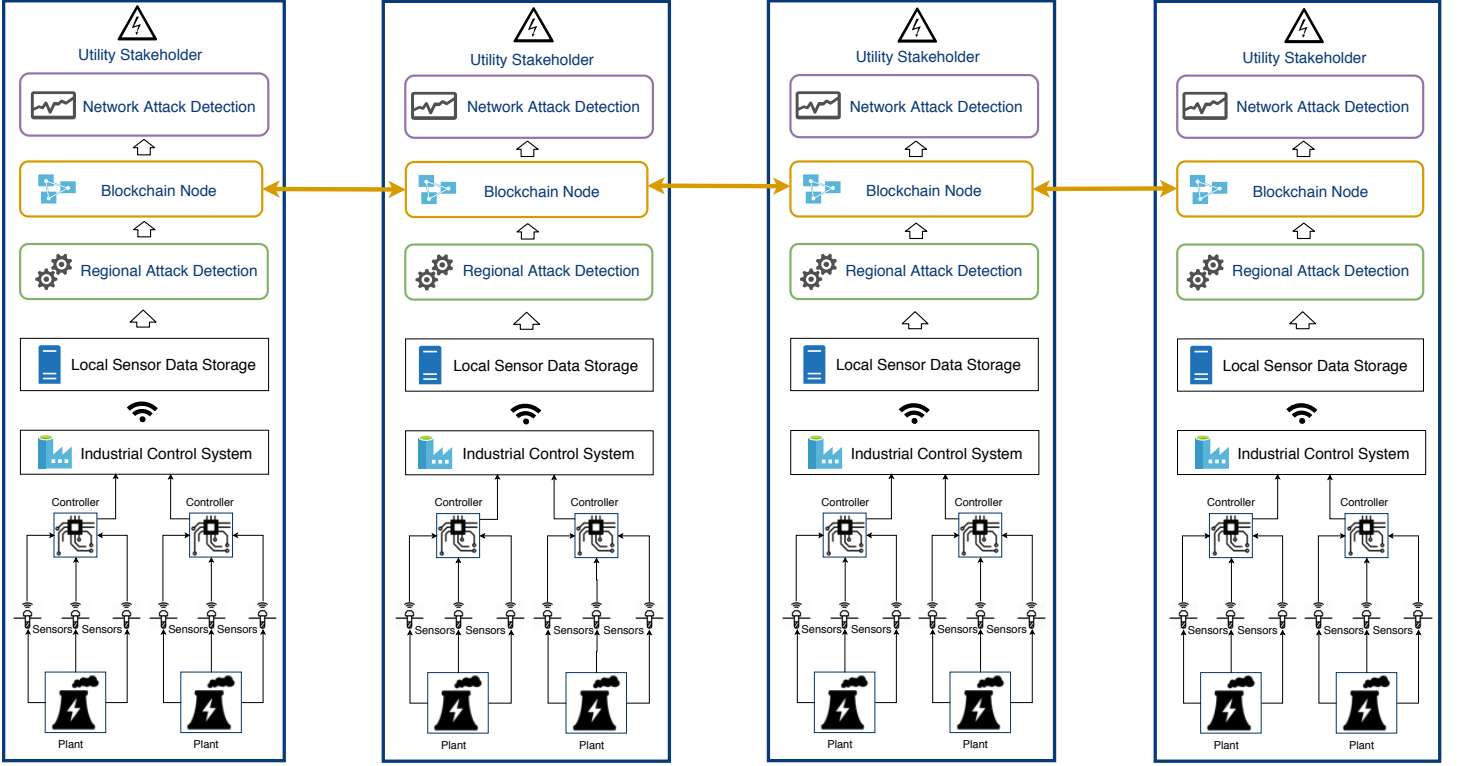
Blockchain based Global Replay Attack Detection

Fig. 1: System Architecture for blockchain based global replay attack detection

V. PERFORMANCE COMPARISON OF BLOCKCHAIN AND BG DRIVEN APPROACHES

For developing a decentralized global replay attack detection, we assume the existence of an underlying connectivity graph that represents the system level interconnection among utilities [1]. In this graph, each vertex represents a utility or region. An edge exists if there is a shared transmission line between the corresponding utilities. Without a central aggregator, a diffusion mechanism must take place over the existing connectivity graph in order to detect globally coordinated replay attacks. Therefore, we develop a novel diffusion algorithm based on state-of-the-art BG as a benchmark strategy for our blockchain based approach [18]. BG can be used to compute the global average of values held by the vertices of the connectivity graph. BG avoids the computational bottlenecks found in other gossip protocols while converging to the global average in expectation [18]. Being inherently decentralized, consensus driven and peer-to-peer in nature, BG forms the ideal benchmark for comparing the performance of a blockchain driven framework.

A. BG based Reformulation of Global Attack Detection

Since the total number of agents in the system is known, BG can be used to estimate the global sum and product terms present in Theorem III.1 as well. Recall $x_i, y_i, \forall i \in \mathcal{R}$ from Algorithm 1. For computing the sum y, the reformulation is trivial and consists of estimating the global sum from the global average $\sum_{k=1}^{n} y_i/n$.

In order to use BG to calculate the product $x = \prod_{i=1}^{n} x_i$, we

reformulate $x = e^u$, where $u = \sum_{i=1}^{n} u_i$, and $u_i = \log x_i$. We leverage the BG protocol to compute $\sum_{i=1}^{n} u_i/n$ which can be used to estimate x in a purely peer-to-peer fashion.

B. Performance Analysis

We wish to compare and contrast the precision factor based error introduced in the blockchain framework against the expected asymptotic error in a BG driven framework. As a result, we propose Theorem V.1 which helps characterize the conditions favorable for the BG to remain competitive with our blockchain framework. For Theorem V.1, we consider a set of agents connected by graph G with Laplacian L. For our analysis of Theorem V.1, we assume a constant value of β across all regions. Further, we let $\lambda_{n-2}(L)$ and $\lambda_1(L)$ denote the second smallest and the largest eigenvalues of L where $\gamma \in (0,1)$ represents the mixing parameter [18]. The mixing parameter dictates the contribution of values recieved from each neighbor during the BG protocol at each vertex [18]. As illustrated in [18], the expected BG error bound for computing the average $\bar{z} = \sum_{i=1}^{n} z_i/n$ where z_i is the local value held by agent $i \in \{1, 2, \dots n\}$ is given by

$$\lim_{t\to\infty} \Delta(z^t)^2 \le (\Delta z^0)^2(1-r), \text{ where } r = \frac{\gamma^{\frac{\lambda_{n-2}}{\lambda_1}}}{1 - \frac{1-\gamma}{2n}\lambda_1}. \quad (30)$$

Theorem V.1. *Let $\Delta p_b, \Delta p_g^\infty$ denote the precision error induced in a blockchain framework and the limiting asymptotic mean*

square error from BG respectively. $\Delta p_g^\infty \leq \Delta p_b$ *if*

$$D \leq \frac{1}{\beta \sqrt{n(1-r)\left[\left(\frac{\Delta x^0}{x}\right)^2 + \left(\frac{\Delta y^0}{1+y}\right)^2\right]}}$$

where x^0, y^0 *represent the initial error of* x, y *respectively in case of BG.*

Proof. We consider the error on $p = x \cdot (1+y)$ to obtain

$$\Delta p^2 = p^2 \left[\left(\frac{\Delta x}{x}\right)^2 + \left(\frac{\Delta y}{1+y}\right)^2\right]. \qquad (31)$$

Analyzing error on x, y, we have

$$\Delta x^2 = |x|^2 \left[\sum_{i=1}^n \left(\frac{\Delta x_i}{x_i}\right)^2\right], \ \Delta y^2 = \sum_{i=1}^n (\Delta y_i)^2. \qquad (32)$$

With a precision factor of D, we know that $\Delta x_i, \Delta y_i \leq \frac{1}{D}$ leading to the relations

$$\left(\frac{\Delta x}{x}\right)^2 \leq \frac{1}{D^2} \sum_{i=1}^n \left(\frac{1}{x_i}\right)^2, \ \left(\frac{\Delta y}{1+y}\right)^2 \leq \left[\frac{n}{(1+y)D}\right]^2. \qquad (33)$$

We know that,

$$a_i = Pr(\sigma^i | s^i = 0) Pr(s^i = 0) \implies x_i = Pr(s^i = 0 | \sigma^i), \qquad (34)$$

where,

$$x_i = \begin{cases} \beta, & \text{when } \sigma_i = 1 \\ 1 - \alpha, & \text{when } \sigma_i = 0 \end{cases}$$

Let p_b denote the global probability of no attack obtained with a blockchain based framework with precision factor D. Since $\beta, \alpha \approx 0$ and $\frac{1}{x_i} \leq \frac{1}{\beta}$, we obtain

$$\left(\frac{\Delta p_b}{p_b}\right)^2 \leq \frac{1}{D^2}\left[\frac{n}{\beta^2} + \left(\frac{n}{1+y}\right)^2\right]. \qquad (35)$$

Since $\frac{1}{(1+y)^2} \leq 1$, we can safely assume that $\frac{1}{(1+y)^2} \in O(\frac{1}{n\beta^2})$ thereby leading to the following error estimate for the blockchain based algorithm

$$\left(\frac{\Delta p_b}{p_b}\right)^2 \leq \left(\frac{n}{D}\right)^2 \cdot \frac{1}{n\beta^2}. \qquad (36)$$

On the other hand for the BG algorithm recall that $x = e^u$, where $u = \sum_{i=1}^n u_i$, $u_i = \log x_i$. Based on error analysis we obtain

$$(\Delta x)^2 = \left(\frac{\partial x}{\partial u}\Delta u\right)^2 \implies \left(\frac{\Delta x}{x}\right)^2 = \Delta u^2. \qquad (37)$$

Let p_g denote the global probability of no attack obtained from a BG based framework. We can therefore say that,

$$\left(\frac{\Delta p_g}{p_g}\right)^2 \leq \Delta u^2 + \left(\frac{\Delta y}{1+y}\right)^2 \qquad (38)$$

Substituting the expected upper bound on error terms $\Delta u^2, \Delta y^2$ from Equation (30), we obtain

$$\left(\frac{\Delta p_g}{p_g}\right)^2 \leq n^2(1-r)\left[(\Delta u^0)^2 + \left(\frac{\Delta y^0}{1+y}\right)^2\right] \qquad (39)$$

Based on Equations (36) and (39), for the BG algorithm to completely outperform the blockchain based method, the following
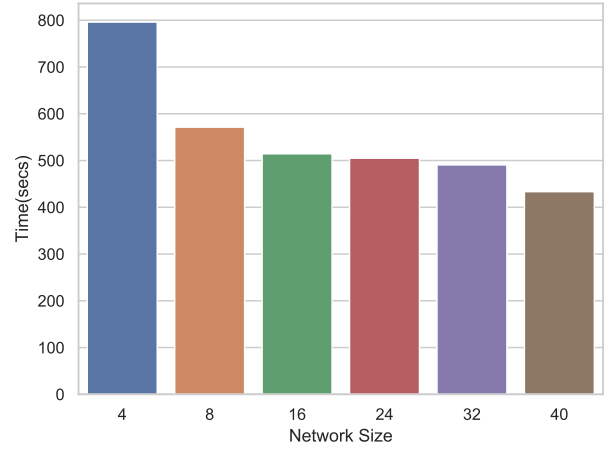

Scaling Study with Varying Number of Regions

Fig. 2: Computational Scaling Study comparing Simulation Time against varying number of total regions for IEEE 3012 bus case

condition must be satisfied

$$n^2(1-r)\left[(\Delta u^0)^2 + \left(\frac{\Delta y^0}{1+y}\right)^2\right] \leq \left(\frac{\Delta p_b}{p_b}\right)^2 \leq \left(\frac{n}{D}\right)^2 \cdot \frac{1}{n\beta^2}. \qquad (40)$$

Since we know that $\Delta u^0 = \frac{x^0}{x}$ we can state that

$$D \leq \frac{1}{\beta \sqrt{n(1-r)\left[\left(\frac{\Delta x^0}{x}\right)^2 + \left(\frac{\Delta y^0}{1+y}\right)^2\right]}}. \qquad (41)$$

$\square$

Theorem V.1 shows us that the precision factor D has an inverse relation to the initial BG error, $\Delta x^0/x + \Delta y_0^2/y$. It can also be noted that $r(\gamma)$ is a monotonously increasing function with $0 \leq r(\gamma) \leq \frac{\lambda_{n-2}}{\lambda_1}$ [18]. Therefore, Theorem V.1 indicates a direct relationship between the precision factor D and the mixing parameter γ as well as the eigenvalue ratio denoted by $\frac{\lambda_{n-2}}{\lambda_1}$.

Based on observations from Theorem V.1, we can postulate several constraints on BG in order for it to match the detection quality of a blockchain driven approach with a high precision factor D. First, BG must preferably start with low initial error with respect to the global values of x, y. Second, since $\gamma \to 1$ is more favorable for BG, only scant perturbation of the local estimate can be allowed when new neighbor messages are received. Therefore, a BG framework cannot afford a drastic change in the overall network mean, making a low initial error state on all nodes imperative to its success. Lastly, for good detection quality, the BG prefers a highly connected underlying graph as indicated by the requirement for $\frac{\lambda_{n-2}}{\lambda_1} \to 1$.

The prerequisite constraints for the BG to outperform the blockchain present significant implementation challenges especially in a large scale power network with rapidly evolving global cyber health status. As a result, a blockchain driven framework is a highly favorable option for delivering an accurate, reliable and timely estimate of the global attack probability.

VI. EXPERIMENTAL RESULTS

Our blockchain results utilize a Geth[3] based, Ethereum private blockchain network orchestrated on a 2.7 GHz Intel Xeon
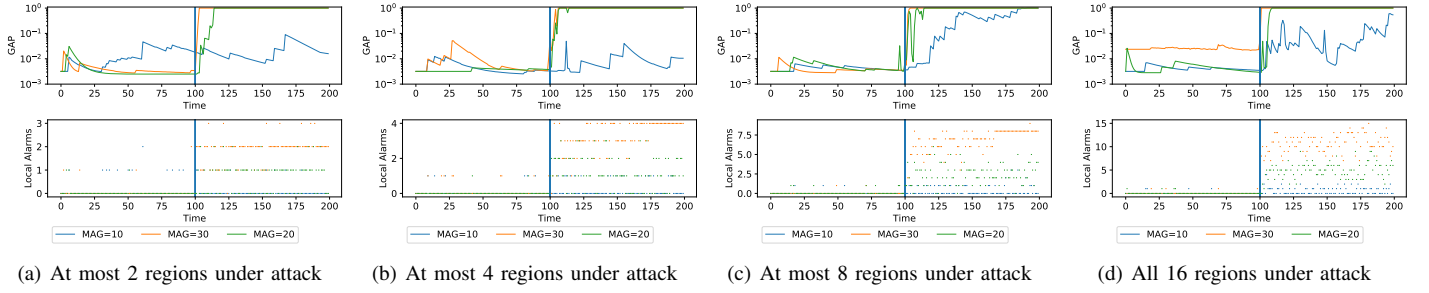
---

[3]https://geth.ethereum.org

(a) At most 2 regions under attack

(b) At most 4 regions under attack

(c) At most 8 regions under attack

(d) All 16 regions under attack

Fig. 3: 16 region decomposition of IEEE 3012 bus case : Global Attack Probability detection with varying number of regions under attack

(a) 1 gossip round per epoch

(b) 10 gossip rounds per epoch

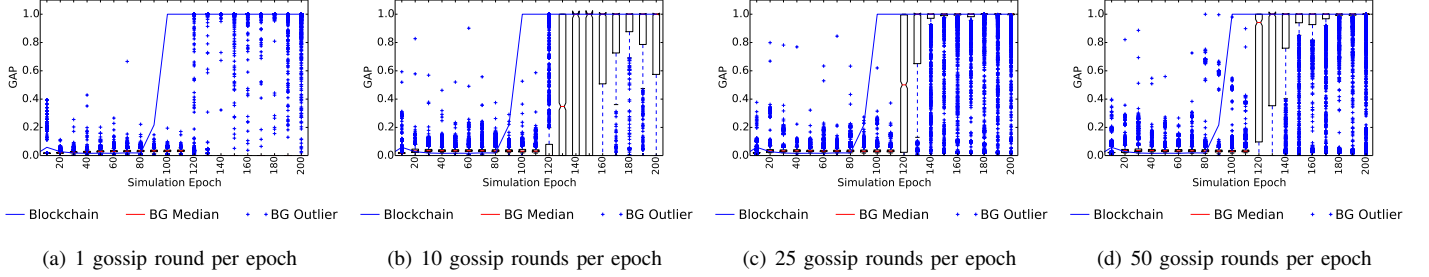(c) 25 gossip rounds per epoch

(d) 50 gossip rounds per epoch

Fig. 4: 40 region decomposition of IEEE 3012 bus case : Comparison with BG with varying local gossip rounds with fixed 2 regions under attack

(a) At most 2 regions under attack

(b) At most 4 regions under attack
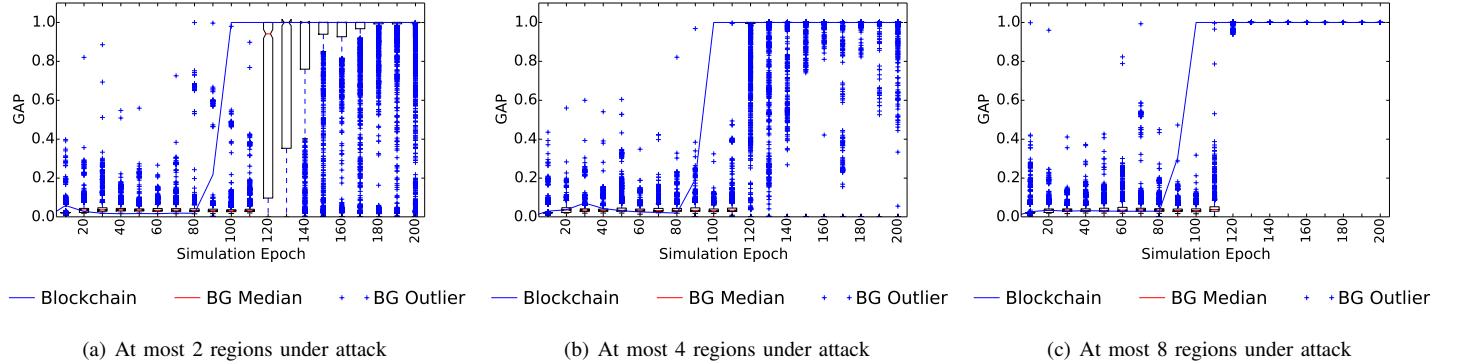
(c) At most 8 regions under attack

Fig. 5: 40 region decomposition of IEEE 3012 bus case : Comparison with BG with varying number of regions under attack with local gossip rounds fixed at 50

processor with 128 cores. We used *Solidity 5.2* to compile the SC and utilized multithreading to simulate the region processes. Each region is assigned one thread each for the Ethereum node and the regional detection model respectively. There is no global synchronization among the regional processes. We used the IEEE 3012 bus transmission network to simulate a large scale transmission network consisting of 150 generators. In our experiments, we set $\alpha = 0.005$ for the regional detection component for all regions. For our blockchain framework, we use a precision factor value of $D = 10^6$. Further, we employ the Proof-of-Authority (PoA) consensus protocol [4] given its suitability for power system applications wherein utilities can act as the authorities. PoA eliminates computationally intensive consensus by shifting the onus of trust to the authorities themselves. Our simulations occur at discrete epochs at which local alarms are reported. In all our

experiments the simulation horizon is 200 epochs.

We leverage the Clique Proof-of-Authority consensus protocol which offers numerous benefits from the security perspective. First, Clique relies on the authority of real world entities that are a part of a permissioned ledger. Based on Table I, a majority (i.e. 50%) of the authorities have to reach consensus before a block is permanently added to the chain. Moreover, Clique can also tolerate an adversarial attack of at most 50% of the validators/authorities. As a result, PoA improves security without relying on computational challenges, since an attacker must hack a majority of authorities in order to rescind all transactions. PoA chains are also known to have low latency, deterministic block creation process as well as faster block creation times which are more important in a permissioned setting. PoA is more suited to consortium or permissioned settings due to its security, easier maintenance as well as accountability of the authorities

[4]https://wiki.parity.io/Proof-of-Authority-Chains

themselves. Therefore, in this paper, we adopt a permissioned blockchain driven approach since our problem setting involves a consortium of real world entities such as utility stakeholders with their own ICSs desirous of detecting global replay attacks.

A. Computational Performance and Accuracy

We evaluate the computational performance of our blockchain framework by considering varied regional decompositions of the IEEE 3012 bus case. Figure 2 presents the incurred time for simulation with varying number of total regions for our blockchain based framework. We observe that a lesser number of regions takes longer to simulate in general. We also notice that the overall simulation time decreases sharply after 4 regions, decreasing gradually and eventually settling at approximately 500secs. This is because as a result of the Clique [5] PoA consensus protocol, lower values of n contribute to a high overall simulation time due to increased block generation time [51]. However, increasing n, results in lesser block generation time but more communication burden. Therefore, we observe that the simulation time eventually stabilizing around a fixed value even as n is increased.

Figure 3 presents results pertaining to experiments involving varied number of regions under attack along with varying degree of attack magnitude. We observe that the framework is robust to varying attack magnitudes and is able to qualitatively detect attacks regardless of attack magnitude. Further, the results also indicate that the framework performs consistently with varying number of regions under attack as well.

B. Evaluation against BG

Figures 4 and 5 depict box and line plots against discrete simulation epochs corresponding to the BG and the blockchain respectively for a magnitude of 20. Each box plot presents the spectrum of global probability values recorded by all region across 100 repeated trials of BG. Similarly, the line plot represents the global probability values determined by the blockchain framework.

Figure 4 compares the performance of the blockchain version with a BG based scheme with varying number of gossip rounds for each simulation epoch by enforcing two regional attacks. We observe that the blockchain based framework detects a global attack faster than the BG by a margin of 30 epochs. Moreover an increasing number of BG rounds per epoch leads to a sharp decrease in the mean, median and the variance of the global attack probability across all regions. Figure 4 demonstrates that despite 50 synchronous gossip rounds per epoch, a global attack detection cannot be successfully disseminated among all regions in BG.

Figure 5 compares the performance of the blockchain based global attack detection paradigm with its BG counterpart with increasing number of regions under attack. We fix the maximum number of gossip rounds to 50 for each simulation case. We observe that the blockchain outperforms the gossip algorithm by a margin of approximately 20 simulation epochs. Figure 5 also indicates that in case of BG the initial disparity of the GAP among the regions decreases with an increase in the number of regions under attack. Such a scenario results in marginal improvement in the global attack detection.

As postulated by Theorem V.1, low initial error is preferred for BG, which only happens when an increasing number of regions are under attack. On the other hand, a high number of gossip rounds are required for BG in order to overcome regional

---

[5] https://github.com/ethereum/EIPs/issues/225

---

connectivity constraints predicted by Theorem V.1. Both scenarios highlight the operational obstacles associated with an information diffusion scheme based on gossip. Figures 4, 5 conclusively show that the blockchain based framework delivers a reliable, timely and accurate detection as compared to its gossip counterpart in diverse operational scenarios.

VII. CONCLUSION AND FUTURE WORK

In this paper, we present a decentralized blockchain based global attack detection mechanism that only uses locally reported alarm and its associated statistics to detect the onset of a global replay attack. We design a novel Bayesian update mechanism requiring one global multiplication and one global addition leading to a scalable and computationally efficient blockchain paradigm. We characterize the performance of the blockchain based global attack detection mechanism against a broadcast gossip based counterpart. In order to do so, we first reformulate the computation of the global multiplication and addition operations to be amenable towards broadcast gossip. We then theoretically analyze the performance of the broadcast gossip with a limited precision blockchain version. Our analysis predicts an overall computational superiority of the blockchain version as opposed to the broadcast gossip. We implement and evaluate the blockchain based approach on a private Ethereum network with the help of Solidity for orchestrating the Smart Contracts. Our experiments demonstrate the accuracy of the decentralized detection mechanism as well as its robustness to increasing number of regions. Moreover, our results also corroborate our theoretical claim of computational superiority over the state-of-the-art, decentralized broadcast gossip paradigm by a significant margin. For our future work we plan to focus on local attacks that could be mutually correlated. We also aim to investigate a blockchain driven approach to distinguish between multiple failure modes and an attack.

REFERENCES

[1] M. Rosas-Casals, S. Valverde, and R. V. Solé, "Topological vulnerability of the european power grid under errors and attacks," *International Journal of Bifurcation and Chaos*, vol. 17, no. 07, pp. 2465–2475, 2007.

[2] R. Kinney, P. Crucitti, R. Albert, and V. Latora, "Modeling cascading failures in the north american power grid," *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 46, no. 1, pp. 101–107, 2005.

[3] R. L. Krutz, *Securing SCADA Systems*. Hungry Minds Inc, 2005.

[4] D. Kushner, "The real story of stuxnet," *IEEE Spectrum*, vol. 3, no. 50, pp. 48–53, 2013.

[5] D. Albright, P. Brannan, and C. Walrond, *Did Stuxnet take out 1,000 centrifuges at the Natanz enrichment plant?* Institute for Science and International Security, 2010.

[6] J. Weiss, "Aurora generator test," *Handbook of SCADA/Control Systems Security*, p. 107, 2016.

[7] T. Rid and P. McBurney, "Cyber-weapons," *The RUSI Journal*, vol. 157, no. 1, pp. 6–13, 2012.

[8] Y. Liu, P. Ning, and M. K. Reiter, "False data injection attacks against state estimation in electric power grids," *ACM Transactions on Information and System Security (TISSEC)*, vol. 14, no. 1, p. 13, 2011.

[9] C. Peng, H. Sun, M. Yang, and Y.-L. Wang, "A survey on security communication and control for smart grids under malicious cyber attacks," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 8, pp. 1554–1569, 2019.

[10] Y. Mo, R. Chabukswar, and B. Sinopoli, "Detecting integrity attacks on scada systems," *IEEE Transactions on Control Systems Technology*, vol. 22, no. 4, pp. 1396–1407, 2014.

[11] R. S. Smith, "A decoupled feedback structure for covertly appropriating networked control systems," *IFAC Proceedings Volumes*, vol. 44, no. 1, pp. 90–95, 2011.

[12] Y. Benjamini and D. Yekutieli, "The control of the false discovery rate in multiple testing under dependency," *Annals of Statistics*, pp. 1165–1188, 2001.

[13] "Multiyear plan for energy sector cybersecurity," https://www.energy.gov/sites/prod/files/2018/05/f51/DOE%20Multiyear%20Plan%20for%20Energy%20Sector%20Cybersecurity%20_0.pdf.

[14] G. Fisk, C. Ardi, N. Pickett, J. Heidemann, M. Fisk, and C. Papadopoulos, "Privacy principles for sharing cyber security data," in *2015 IEEE Security and Privacy Workshops*. IEEE, 2015, pp. 193–197.

[15] P. Ramanan and K. Nakayama, "Baffle: Blockchain based aggregator free federated learning," *arXiv preprint arXiv:1909.07452*, 2019.

[16] S. Wang, L. Ouyang, Y. Yuan, X. Ni, X. Han, and F.-Y. Wang, "Blockchain-enabled smart contracts: architecture, applications, and future trends," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 49, no. 11, pp. 2266–2277, 2019.

[17] S. Boyd, A. Ghosh, B. Prabhakar, and D. Shah, "Randomized gossip algorithms," *IEEE Transactions on Information Theory*, vol. 52, no. 6, pp. 2508–2530, 2006.

[18] T. C. Aysal, M. E. Yildiz, A. D. Sarwate, and A. Scaglione, "Broadcast gossip algorithms for consensus," *IEEE Transactions on Signal Processing*, vol. 57, no. 7, pp. 2748–2761, 2009.

[19] A. Cook, H. Janicke, R. Smith, and L. Maglaras, "The industrial control system cyber defence triage process," *Computers & Security*, vol. 70, pp. 467–481, 2017.

[20] L. D. Sturm, C. B. Williams, J. A. Camelio, J. White, and R. Parker, "Cyber-physical vulnerabilities in additive manufacturing systems: A case study attack on the. stl file with human subjects," *Journal of Manufacturing Systems*, vol. 44, pp. 154–164, 2017.

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

[22] C. Copeland and B. Cody, "Terrorism and security issues facing the water infrastructure sector," 2010.

[23] A. Lazarevic, L. Ertoz, V. Kumar, A. Ozgur, and J. Srivastava, "A comparative study of anomaly detection schemes in network intrusion detection," in *Proceedings of the 2003 SIAM International Conference on Data Mining*. SIAM, 2003, pp. 25–36.

[24] M. Caselli, E. Zambon, J. Amann, R. Sommer, and F. Kargl, "Specification mining for intrusion detection in networked control systems." in *USENIX Security Symposium*, 2016, pp. 791–806.

[25] N. Ye, Y. Zhang, and C. M. Borror, "Robustness of the markov-chain model for cyber-attack detection," *IEEE Transactions on Reliability*, vol. 53, no. 1, pp. 116–123, 2004.

[26] D. I. Urbina, J. A. Giraldo, A. A. Cardenas, N. O. Tippenhauer, J. Valente, M. Faisal, J. Ruths, R. Candell, and H. Sandberg, "Limiting the impact of stealthy attacks on industrial control systems," in *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security*. ACM, 2016, pp. 1092–1105.

[27] D.-H. Kang, B.-K. Kim, and J.-C. Na, "Cyber threats and defence approaches in scada systems," in *16th International Conference on Advanced Communication Technology*. IEEE, 2014, pp. 324–327.

[28] Z. Drias, A. Serhrouchni, and O. Vogel, "Taxonomy of attacks on industrial control protocols," in *2015 International Conference on Protocol Engineering (ICPE) and International Conference on New Technologies of Distributed Systems (NTDS)*. IEEE, 2015, pp. 1–6.

[29] G. Dan and H. Sandberg, "Stealth attacks and protection schemes for state estimators in power systems," in *Smart Grid Communications (SmartGridComm), 2010 First IEEE International Conference on*. IEEE, 2010, pp. 214–219.

[30] A. Teixeira, S. Amin, H. Sandberg, K. H. Johansson, and S. S. Sastry, "Cyber security analysis of state estimators in electric power systems," in *49th IEEE Conference on Decision and Control (CDC). Atlanta, GA. DEC 15-17, 2010*, 2010, pp. 5991–5998.

[31] Z.-H. Yu and W.-L. Chin, "Blind false data injection attack using pca approximation method in smart grid," *IEEE Transactions on Smart Grid*, vol. 6, no. 3, pp. 1219–1226, 2015.

[32] A. A. Cárdenas, S. Amin, Z.-S. Lin, Y.-L. Huang, C.-Y. Huang, and S. Sastry, "Attacks against process control systems: risk assessment, detection, and response," in *Proceedings of the 6th ACM symposium on information, computer and communications security*. ACM, 2011, pp. 355–366.

[33] G. Chaojun, P. Jirutitijaroen, and M. Motani, "Detecting false data injection attacks in ac state estimation," *IEEE Transactions on Smart Grid*, vol. 6, no. 5, pp. 2476–2483, 2015.

[34] T. Huang, B. Satchidanandan, P. Kumar, and L. Xie, "An online detection framework for cyber attacks on automatic generation control," *IEEE Transactions on Power Systems*, vol. 33, no. 6, pp. 6816–6827, 2018.

[35] D. Van Long, L. Fillatre, and I. Nikiforov, "Sequential monitoring of scada systems against cyber/physical attacks," *IFAC Papers Online*, vol. 48, no. 21, pp. 746–753, 2015.

[36] A. Hoehn and P. Zhang, "Detection of covert attacks and zero dynamics attacks in cyber-physical systems," in *American Control Conference (ACC), 2016*. IEEE, 2016, pp. 302–307.

[37] S. Weerakkody and B. Sinopoli, "Detecting integrity attacks on control systems using a moving target approach," in *Decision and Control (CDC), 2015 IEEE 54th Annual Conference on*. IEEE, 2015, pp. 5820–5826.

[38] S. Pan, T. Morris, and U. Adhikari, "Classification of disturbances and cyber-attacks in power systems using heterogeneous time-synchronized data," *IEEE Transactions on Industrial Informatics*, vol. 11, no. 3, pp. 650–662, 2015.

[39] M. S. Rahman, M. A. Mahmud, A. M. T. Oo, and H. R. Pota, "Multi-agent approach for enhancing security of protection schemes in cyber-physical energy systems," *IEEE Transactions on Industrial Informatics*, vol. 13, no. 2, pp. 436–447, 2017.

[40] R. Romagnoli, S. Weerakkody, and B. Sinopoli, "A model inversion based watermark for replay attack detection with output tracking," in *2019 American Control Conference (ACC)*. IEEE, 2019, pp. 384–390.

[41] M. Porter, P. Hespanhol, A. Aswani, M. Johnson-Roberson, and R. Vasudevan, "Detecting generalized replay attacks via time-varying dynamic watermarking," *IEEE Transactions on Automatic Control*, 2020.

[42] P. Hespanhol, M. Porter, R. Vasudevan, and A. Aswani, "Sensor switching control under attacks detectable by finite sample dynamic watermarking tests," *IEEE Transactions on Automatic Control*, 2020.

[43] Y. Mo, S. Weerakkody, and B. Sinopoli, "Physical authentication of control systems: Designing watermarked control inputs to detect counterfeit sensor outputs," *IEEE Control Systems Magazine*, vol. 35, no. 1, pp. 93–109, 2015.

[44] T. Irita and T. Namerikawa, "Detection of replay attack on smart grid with code signal and bargaining game," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 2112–2117.

[45] D. Li, N. Gebraeel, and K. Paynabar, "Detection and differentiation of replay attack and equipment faults in scada systems," *IEEE Transactions on Automation Science and Engineering*, 2020.

[46] J. Wan, J. Li, M. Imran, D. Li, and Fazal-e-Amin, "A blockchain-based solution for enhancing security and privacy in smart factory," *IEEE Transactions on Industrial Informatics*, vol. 15, no. 6, pp. 3652–3660, June 2019.

[47] Z. Zhou, B. Wang, M. Dong, and K. Ota, "Secure and efficient vehicle-to-grid energy trading in cyber physical systems: Integration of blockchain and edge computing," *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 50, no. 1, pp. 43–57, 2019.

[48] M. Mylrea and S. N. G. Gourisetti, "Blockchain: A path to grid modernization and cyber resiliency," in *2017 North American Power Symposium (NAPS)*, Sep. 2017, pp. 1–5.

[49] M. L. Di Silvestre, P. Gallo, M. G. Ippolito, E. R. Sanseverino, and G. Zizzo, "A technical approach to the energy blockchain in microgrids," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 11, pp. 4792–4803, Nov 2018.

[50] N. Alexopoulos, E. Vasilomanolakis, N. R. Ivánkó, and M. Mühlhäuser, "Towards blockchain-based collaborative intrusion detection systems," in *International Conference on Critical Information Infrastructures Security*. Springer, 2017, pp. 107–118.

[51] S. De Angelis, L. Aniello, R. Baldoni, F. Lombardi, A. Margheri, and V. Sassone, "Pbft vs proof-of-authority: applying the cap theorem to permissioned blockchain," 2018.

[52] M. Salimitari and M. Chatterjee, "A survey on consensus protocols in blockchain for iot networks," *arXiv preprint arXiv:1809.05613*, 2018.

[53] "Proof-of-Authority Chains," https://openethereum.github.io/Proof-of-Authority-Chains, [Online; accessed 22-April-2021].

[54] R. Saltini and D. Hyland-Wood, "Ibft 2.0: A safe and live variation of the ibft blockchain consensus protocol for eventually synchronous networks," *arXiv preprint arXiv:1909.10194*, 2019.

[55] B. D. Anderson and J. B. Moore, "Optimal filtering," *Englewood Cliffs*, vol. 21, pp. 22–95, 1979.

[56] M. Athans, "The role and use of the stochastic linear-quadratic-gaussian problem in control system design," *IEEE Transactions on Automatic Control*, vol. 16, no. 6, pp. 529–552, 1971.

[57] Z. Zheng, S. Xie, H.-N. Dai, W. Chen, X. Chen, J. Weng, and M. Imran, "An overview on smart contracts: Challenges, advances and platforms," *Future Generation Computer Systems*, vol. 105, pp. 475–491, 2020.

APPENDIX A

We provide a description of all the characteristic features listed in Table I for permissioned consensus mechanisms. The features and their descriptions are as follows:

- *Validators* refers to the minimum number of entities required for consensus to take place.
- *Decentralization* aspect characterizes the potential of a few entities dominating over block creation.
- *Fault tolerance* describes the maximum number of entities that must fail to jeopardize the blockchain operation.

- *Consistency* refers to the final state of the distributed ledger among multiple entities post consensus.
- *Block latency* characterizes the degree of communication required in order to mint a block.

Decentralization potential of Clique and Aura are deemed high due to the fact that all participating entities are provided an opportunity to propose blocks during run time [51]. In IBFT, validators once chosen need to be added and removed explicitly by a proposal followed by voting [54]. In Raft, if the leader is malicious, then the entire blockchain can be compromised, unless a change is effected using voting of the verifiers [52]. Therefore, the decentralization potential for IBFT and Raft is deemed medium and low respectively.

With respect to consistency, Clique relies on the Ethereum GHOST protocol to eventual resolve forks and achieve consistency of the chain. However, Aura makes no such guarantees [51]. On the other hand, IBFT emphasizes consistency over speed resulting in a stalling in cases where a fork has emerged [51]. Raft meanwhile assumes that no forking can emerge on account of there being only one leader [52]