EPIC: A Differential Privacy Framework to Defend Smart Homes Against Internet Traffic Analysis

Jianqing Liu, *Graduate Student Member, IEEE*, Chi Zhang, *Member, IEEE*, and Yuguang Fang, *Fellow, IEEE*

***Abstract*—The Internet of Things (IoT) becomes a novel paradigm as more and more devices are connected to the Internet, enabling several innovative applications such as smart home, industrial automation, and connected health. However, the cyber-attack to these applications is a big issue and countermeasures are in dire need to provide system security and user privacy. In this paper, we address the traffic analysis attack to smart homes, where adversaries intercept the Internet traffic from/to the smart home gateway and profile residents' behaviors through digital traces. Traditional cryptographic tools may not work well due to the effectiveness of adversaries' machine learning algorithms in classifying encrypted traffic, so here we propose a privacy-preserving traffic obfuscation framework to achieve the goal. To be specific, we leverage the smart community network of wirelessly connected smart homes and intentionally direct each smart home's traffic to another home gateway before entering the Internet. The design jointly considers the network energy consumption and the resource constraints in IoT devices, while achieving strong differential privacy guarantee so that adversaries cannot link any traffic flow to a specific smart home. Besides, we consider a hostile smart community network and develop secure multihop routing protocols to guarantee the source/destination unlinkability and satisfy each user's personalized privacy requirement. To evaluate the effectiveness of our framework in protecting privacy and reducing network energy consumption, extensive simulations are conducted and the results demonstrate that our design outperforms other differential privacy mechanism in preserving privacy and minimizing network utility cost.**

***Index Terms*—Bayesian inference, differential privacy, energy efficiency, Internet of Things (IoT), secure routing, traffic analysis attack.**

I. Introduction

INTERNET of Things (IoT) is a novel paradigm that encompasses interconnected smart devices, such as sensors, actuators, displays, vehicles, home appliances, etc., which are enabled to communicate with one another and collaboratively accomplish certain goals. It is estimated that there will be more than 50 billion network connected devices by 2020. Such a rapid development of IoT will foster a variety of novel applications, such as industrial automation, mobile healthcare, smart grids, intelligent transportation, and many others [1]. Among these applications, smart home is one of the promising IoT practices where home appliances such as surveillance cameras, medical sensor devices, thermostat, etc. provide pervasive sensing and can be remotely controlled by home owners or caregivers, providing residents with the most convenience, comfort and security [2].

Recently, smart home is also becoming a locus for the health care innovations, such as the Intel's Health Guide [3] and GE's QuietCare [4], as this new paradigm gives the patients higher freedom and reduces the societal costs. On the one hand, studies show that in 2003, U.S. alone spent $1.7 trillion on health care, with 75% of these costs directed toward the treatments of chronic diseases [5], such as heart disease, diabetes, HIV, and cancer. Whereas using the smart home-based remote health systems, caregivers can access patients' health status in a timely manner and provide them with preventive instructions, helping to avoid emergency care and hospital admissions, which reduces a huge amount of societal costs. On the other hand, studies show that people, especially the elderly, regard their home as a sanctuary and therefore prefer to stay at home for medical treatments of the chronic diseases [6], which necessitates the development of the remote health monitoring systems in smart homes.

However, for many individuals, home is a foundational area with the highest level of privacy, but the remote health systems require the collection, disclosure, and usage of the personal health data, which breeds serious privacy concerns. For example, a glucometer measuring the blood sugar level, a spirometer tracking the air entering/leaving the lungs, and a sleep monitoring sensor recording the sleep conditions can potentially reveal whether the resident carries diabetes, seasonal allergy-induced asthma, or depressive disorder, respectively. For privacy concerns, patients are inclined to restrict the access of these data to a limited group of people like their personal doctors.

To enforce the data confidentiality, integrity and access control, some protection mechanisms [1]—lightweight cryptography, secure protocols, and privacy assurance—have already been proposed in current literatures. For instance, Intel Health Guide [3] applies 128 bit secure sockets layer technology to encrypt the sensory and control data. However, the adversarial side channel attacks like Internet traffic analysis (e.g., through analyzing packet sizes and timing) over the encrypted digital traces that the smart home generates can reveal surprising information about the traffic's contents [7], [8]. For instance,

Manuscript received November 8, 2017; revised January 2, 2018; accepted January 23, 2018. Date of publication February 1, 2018; date of current version April 10, 2018. This work was supported by the National Science Foundation under Grant IIS-1722791, Grant CNS-1343356, and Grant CNS-1409797. *(Corresponding author: Yuguang Fang.)*

J. Liu and Y. Fang are with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL 32611 USA (e-mail: jianqingliu@ufl.edu; fang@ece.ufl.edu).

C. Zhang is with the Key Laboratory of Electromagnetic Space Information, Chinese Academy of Sciences, Beijing 100864, China, and also with the School of Information Science and Technology, University of Science and Technology of China, Hefei 230027, China (e-mail: chizhang@ustc.edu.cn).

Digital Object Identifier 10.1109/JIOT.2018.2799820

2327-4662 © 2018 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

Meidan *et al.* [9] showed the effectiveness of identifying smart home devices by applying machine learning algorithm to the encrypted network traffic. By doing so, the attacker could identify what specific health monitoring or actuating devices are utilized in the smart homes. Furthermore, through continuous observation of the Internet traffic, such as destination address of the smart homes traffic and the periodicity of the interactions, the attacker may deduce what type of disease or health issues the residents have. For instance, if the sleep monitoring sensor shows intermittent sleeping patterns (i.e., traffic exhibiting intermittent bursts) over the night and this traffic is directed to a phycological clinic, after observing this pattern for a long period of time, the adversary has a high confidence to infer that the resident may have the depressive disorder. Unfortunately, if this health condition is disclosed to the resident's colleagues or friends, it may exacerbate his/her health condition and ruins his/her social life. Similarly, for other chronic diseases like HIV or other maternal health issues like miscarriage [10], their unique serving health devices and periodicity of sensory/medical instruction data transmissions could reveal many useful information to the adversaries, which causes great impacts on the residents' privacy.

In the smart home-based health systems, all devices communicate with the health care service providers through the Internet via a home wireless router (i.e., home gateway), so the adversaries could just intercept the network traffic remotely from the Internet, and extract many useful health information about the resident. However, the current countermeasure techniques to defend against traffic analysis (e.g., onion routing and traffic morphing) on one hand may not be effective, and on the other hand could result in huge network resource consumption. A detailed survey for the common solution tactics are discussed in Section VI. To this end, it seems that it is difficult for the smart home standalone to effectively defend against the traffic analysis attacks given the limited resources it has.

In this paper, we resort the newly introduced concept called *smart community* [11] to address the traffic analysis attack from the network perspective. We intend to explore how we can manage or maneuver the network resources to preserve the privacy of each smart home residents in a better way. This network-level solution features one of the mostly accepted approaches to manage the IoT security and privacy [12], especially considering IoT devices are resource-limited. Here, the smart community is a network of connected smart homes located in a geographical region. Home gateways, representing their smart homes, are interconnected via wireless multihop transmissions using any radio access technology (e.g., WiFi). The home gateway is a critical component in this context and it should not only be considered merely as a communications device but a local computing platform which executes certain cryptographic operations for security purposes. The basic idea behind this approach is that the smart home traffic are intentionally directed toward and aggregated at certain home gateways (i.e., "proxy gateways" or "outlet gateways") before entering the Internet. This local traffic obfuscation (or "shuffle") leveraging other smart devices in wireless environment is expected to make adversaries unable to link the network traffic observed from the core Internet to a specific physical smart home.

Selecting the appropriate proxy gateways is a nontrivial task when it comes to provide strong privacy guarantees so that adversaries are incapable of distinguishing the source of a specific traffic flow. Meanwhile, the proxy gateway selection strategy should consider network resource consumption (e.g., energy, computing, and communications resources), so that a good balance between privacy and utility is achieved. Furthermore, when we consider a hostile environment where some smart home gateways are compromised by adversaries, the multihop routing design between the source and destination is in dire need to ensure any intermediate or proxy gateway is unaware of where a traffic flow originates from. To cope with these challenges, in this paper, we propose an efficient and privacy-preserving traffic obfuscation (EPIC) framework for connected smart homes. Our contribution can be summarized as follows.

1) We develop a differentially private (DP) mechanism for the selection of proxy gateways. The DP mechanism aims to minimize the network energy consumption due to multihop transmissions while providing differential privacy guarantee for the smart homes, and satisfying round-trip delay and computing resource constraints in home gateways. The DP mechanism design is modeled as a linear optimization problem, which is solved to obtain each smart home's selection strategy.
2) To ensure unlinkability between the source and destination home gateways, we propose a directed random walk (DRW) scheme for uplink transmissions and a DRW and flooding hybrid routing scheme for downlink transmissions. In particular, the downlink routing protocol is designed to limit the capability of the intermediate gateway in inferring the source gateway under a specified level. The overall design is further coupled with the prior DP mechanism design and an iterative algorithm is developed to search for solutions.
3) Extensive simulations are conducted based on the real community topology. We apply the exponential DP mechanism as the benchmark and numerically demonstrate that our framework has the advantage in reducing the network resource consumption and ensuring the privacy level of smart homes.

The rest of this paper is organized as follows. Section II describes the network model. Section III presents the adversarial model and gives necessary preliminaries. The EPIC framework design is presented in Section IV and the performance evaluation is followed in Section V. We give a brief survey of related work in Section VI, and draw a conclusion in Section VII.

II. Network Model and Design Objective

A. Network Model

We consider a smart community consisting of M wirelessly connected smart homes,[1] as shown in Fig. 1, which

[1]Throughout this paper, we use "smart home" and "home gateway" interchangeably to represent the same meaning.

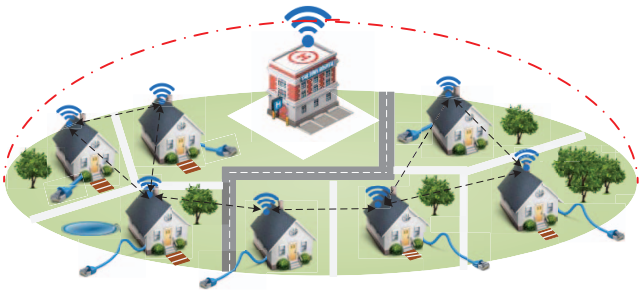

Fig. 1. Smart community: a network of connected smart homes.

are indexed as $\mathcal{M} = \{1, 2, \ldots, m, \ldots, M\}$ with the physical location being represented as $\mathcal{X} = \{x_1, x_2, \ldots, x_m, \ldots, x_M\}$. Each home gateway has a wired connection to the Internet and wireless interfaces to connect with other home gateways. Smart homes in the community form a multihop mesh network with communications resources (e.g., transmission power and frequency channel) being controlled by the smart community center. In this paper, we assume the communications resources are well coordinated and each smart home is aware of the global topology of the wireless multihop community network. The home gateway is not merely a communications device but rather assumed to carry computing capabilities to perform computations such as encryption/decryption and pseudonym generations. We assume a heterogeneous scenario where the computing capabilities (e.g., in unit of CPU cycles) of smart homes are denoted as $\mathcal{C} = \{c_1, c_2, \ldots, c_m, \ldots, c_M\}$.

For the private and secure data communications, a multiple pseudonym techniques can be applied for each home gateway to generate pseudonyms $\{\text{pid}_m\}$, and we assume home gateways be distributed public/private key pairs (pk_m, sk_m) at the system initialization. The communications in smart home IoT environment is a two-way process, where home gateways aggregate and send the intrahome IoT sensory traffic to the Internet or clouds (i.e., uplink transmissions) while the control messages are streamed to the IoT devices via home gateways (i.e., downlink transmissions). All the traffic here could be encrypted to ensure data confidentiality, authenticity and access control, but this design is not the focus of this paper.

B. Design Objective

To protect smart homes from attackers conducting traffic analysis, we aim at designing for each smart home an obfuscation mechanism $\mathcal{A}$, following which a proxy gateway is selected as its traffic outlet to the Internet. Upon designing obfuscation $\mathcal{A}$, we are concerned about the incurred QoS and network resource costs, and here in this paper the evaluation metrics are delay, energy and computing resources consumption, respectively.

To move one step forward, we develop a secure multihop routing scheme to ensure the unlinkability between the source smart home gateway and the proxy gateway, when considering some smart homes in the community may be curious about the traffic origination.

III. Adversarial Model and Preliminaries

A. Adversarial Model

In this paper, we focus on the side channel attacks where adversaries monitor the incoming/outgoing network traffic to/from smart homes, employ classification algorithms to infer activities of smart home IoT devices, and then create residents' profiles to gain advantages in conducting subsequent severe attacks. In assessing the effectiveness of our approach, we consider the informed adversaries who are aware of the protection mechanism, i.e., how it works and the exact obfuscation (i.e., proxy gateway selection) strategy $\mathcal{A}$. Suppose the adversary employs the Bayesian inference attack, where for each proxy gateway z being observed by the adversary, the posterior probability distribution over all community smart homes is used to invert the noise/randomness added by our mechanism $\mathcal{A}$, and thus, to estimate the actual source smart home as follows:

$$h(x|z) = \frac{\Pr(x, z)}{\Pr(z)} = \frac{\Pr(z|x)\psi(x)}{\sum_{x'} \Pr(z|x')\psi(x')} = \frac{\mathcal{A}_x(z)\psi(x)}{\sum_{x'} \mathcal{A}_{x'}(z)\psi(x')} \tag{1}$$

where $\psi(x)$ denotes the adversary's prior knowledge of where the source smart home is. This information could come from traffic analysis: for instance, some unique IoT devices or traffic pattern may help adversaries correlate it with their side information about some residents, thus giving adversaries prior information to aid the inference attack.

Besides, we follow the definition in [13] and quantify residents' privacy as the adversary's expected error in her Bayesian inference attack as in (1). The calculation is as follows:

$$\text{privacy}(x) = \sum_{x' \in \mathcal{X}} \sum_{z \in \mathcal{X}} \mathcal{A}_x(z) h\big(x'|z\big) d_{\text{err}}\big(x, x'\big) \tag{2}$$

where $d_{\text{err}}(x, x')$ could be the Euclidean or Hamming distance measure between x and x', which is the adversary's error in guessing x.

Inside the smart community, we consider a wireless environment where some smart homes are honest-but-curious (HbC), in the sense that they honestly follow the data transmission protocol but are curious where the traffic originates from.

B. Preliminary on Differential Location Privacy

Here, we define the $\epsilon d_{\mathcal{X}}$-differential privacy on a discretized location set [14], with the intuition that observed point z will not help adversaries to differentiate any instance inside this location set who actually initiates the point z.

Definition 1: A randomized mechanism $\mathcal{A}$ satisfies $\epsilon d_{\mathcal{X}}$-differential privacy on location set $\mathcal{X}$ if for any released location z and any two locations x and x', the following holds:

$$\mathcal{A}_x(z) \leq e^{\varepsilon d_{\mathcal{X}}(x,x')} \mathcal{A}_{x'}(z) \tag{3}$$

where $d_{\mathcal{X}}(\cdot, \cdot)$ is a distance metric (e.g., Hamming distance or Euclidean distance), which expresses the *distinguishability level* between x and x'. Given the location set, obfuscation can be performed to satisfy differential privacy following approaches such as 2-D Laplace distribution [14], utility-optimal mechanism [15], exponential mechanism [16], and so on.

IV. EPIC: Framework to Protect Smart Homes in IoT Environments

A. Utility-Aware DP Proxy Gateway Selection

Conventional obfuscation mechanisms (e.g., K-anonymity) only rely on syntactic privacy models and lack strong privacy guarantee. In our design, we apply the $\epsilon d_{\mathcal{X}}$-differential privacy, which limits privacy leakage by bounding the relative information (between prior and posterior) gain of the adversary, regardless of what kind of prior information the adversary may have. However, in a resource-constrained IoT environment, here the smart homes, traditional DP mechanisms (e.g., exponential or 2-D Laplacian approach) may not be cost-effective or even workable. Therefore, in this paper, instead of applying existing schemes, we develop a novel DP obfuscation mechanism in our smart home IoT context, to achieve minimum utility cost while providing DP guarantee. The design is cast into solving a convex optimization problem and we will elaborate it in detail as follows.

1) Delay Constraint: Suppose $\mathcal{A}$ is the DP mechanism we aim to obtain. When smart home x follows the distribution $\mathcal{A}_x$ to decide the proxy gateway z as its traffic outlet from/to the Internet, certain delay is incurred due to multihop transmissions. Denote t_{prop} and t_{proc} as the propagation and processing time, respectively. If the traffic has a delay requirement, such as the urgent health instruction data, the following constraints must be satisfied:

$$\begin{aligned}&\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\cdot\left[nt_{\text{prop}}+(n+1)t_{\text{proc}}\right]\le T_{\text{ul}}\quad\forall x\in\mathcal{X}\\&\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\cdot\left[kt_{\text{prop}}+(k+1)t_{\text{proc}}\right]\le T_{\text{dl}}\quad\forall x\in\mathcal{X}\end{aligned}\tag{4}$$

where n and k represent the number of hops for uplink and downlink transmissions, respectively, which are dependent on the source/destination pair (x, z) and the corresponding routing protocol [i.e., $f:(x,z)\to n$ and f is the routing protocol]. In later part of this paper, the value of n and k will be elaborated. Besides, note that the processing time includes the computation of encryption/decryption (that will be discussed later) at both source and destination gateways, which accounts for $(n+1)$ and $(k+1)$ times of processing unit. T_{ul} and T_{dl} are delay requirements for the uplink sensory traffic and the downlink control data, respectively. The constraints in (4) describes that the expected delay for any smart home $x\in\mathcal{X}$ by following the obfuscation distribution $\mathcal{A}$ should be less than the required delay. On the other hand, we could also apply the concept of value-at-risk (VaR) [17] and model it as a chance constraint of β-confidence level

$$\Pr\left\{\left[nt_{\text{prop}}+(n+1)t_{\text{proc}}\right]\ge T_{\text{uplink}}\right\}\le\beta$$

which can be simplified as $\Pr\{n\ge(T_{\text{uplink}}-t_{\text{proc}}/t_{\text{prop}}+t_{\text{proc}})\}\le\beta$. Here, the n is a random variable (r.v.) following distribution of $\mathcal{A}_x(z)$. Suppose the cumulative distribution function of r.v. n is $\mathcal{C}_n$, which can be obtained from $\mathcal{A}_x(z)$. Then, the above chance constraint can be reformulated as follows:

$$\mathcal{C}_n^{-1}(1-\beta)\le\frac{T_{\text{uplink}}-t_{\text{proc}}}{t_{\text{prop}}+t_{\text{proc}}}.\tag{5}$$

The downlink delay constraint can be characterized in a similar manner.

2) Computing Resource Constraint: Given that the proxy gateway z could be selected by several smart homes and we shall see later that certain cryptographic calculations are executed at z to ensure secure multihop routing, we should guarantee that the aggregated traffic at z will not cause overloading at the resource-limited gateway. This requirement can be captured by the following constraint:

$$\sum_{x\in\mathcal{X}}\mathcal{A}_x(z)\cdot r(x)\le c_z\quad\forall z\in\mathcal{X}\tag{6}$$

where $r(x)$ represents the encrypted traffic generated from source smart home gateway x and the c_z is the computing capability of proxy gateway z measured in CPU cycles. For instance, if the gateway carries 2.8-GHz Core 2 family CPU, the speed for CTR mode of encryption is 230 Mb/s (interested readers are referred to [18] for details), which means the aggregated incoming traffic rate should be smaller than that. Note that the constraint (6) could also be reformulated using the concept of VaR similar to inequality (5).

3) Problem Formulation: To this end, we can construct the utility-aware DP proxy gateway selection mechanism as an optimization problem, which minimizes the expected utility cost while satisfying $\epsilon d_{\mathcal{X}}$-differential privacy

$$\begin{aligned}\text{Min}\quad&\sum_{x\in\mathcal{X}}\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\cdot U(x,z)\\\text{s.t.}\quad&\mathcal{A}_x(z)\le e^{\varepsilon d_{\mathcal{X}}(x,x')}\mathcal{A}_{x'}(z)\quad\forall x,x';z\in\mathcal{X}\\&\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)=1\quad\forall x\in\mathcal{X}\\&0\le\mathcal{A}_x(z)\le1\quad\forall x;z\in\mathcal{X}\\&\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\cdot\left[nt_{\text{prop}}+(n+1)t_{\text{proc}}\right]\le T_{\text{ul}}\\&\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\cdot\left[kt_{\text{prop}}+(k+1)t_{\text{proc}}\right]\le T_{\text{dl}}\\&\sum_{x\in\mathcal{X}}\mathcal{A}_x(z)\cdot r(x)\le c_z.\end{aligned}\tag{7}$$

To convey our basic design philosophy, we utilize (4) and (6) as the optimization constraints and leave the chance constraints formulation to the future exploration. Here in (7), $U(x,z)$ represents the utility function measured in energy consumption of unit J/bit. $U(x,z)$ is not explicitly given here as it is related to the secure routing mechanism design and we will elaborate it later. $d_{\mathcal{X}}(x,x')$ is a measure of Euclidean distance between smart gateways x and x', whose implication is that if the distance is small, then the "secrets" (here two smart gateways) should remain indistinguishable; while if the distance is large, then the adversary is able to distinguish the secrets from each other. In the optimization problem (7), we aim to obtain the obfuscation mechanism $\mathcal{A}_x(z)$ for each smart home x and the first three constraints enforce the $\epsilon d_{\mathcal{X}}$-differential privacy.

When the secure uplink and downlink multihop routing protocols are developed, which will be elaborated in Section IV-B, the problem (7) becomes an convex

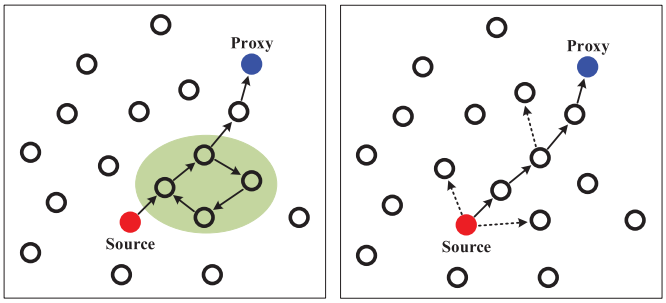

Fig. 2. (a) Inefficiency of the pure random walk scheme due to the routing loop. (b) How the hop-based DRW mechanism works, where the dashed arrow-line indicates the possible next-hop routes.

optimization problem, specifically a linear optimization problem. It consists of $|\mathcal{X}|^2$ decisional variables and $(|\mathcal{X}|^3 + |\mathcal{X}|^2 + 4|\mathcal{X}|)$ constraints, which could be solved via tools like CPLEX.

B. Secure Multihop Routing

In this section, we discuss the routing protocol design for round-trip transmissions, namely, the uplink sensory data transmissions from source smart home gateway to the proxy gateway, and the downlink control data transmissions from proxy gateway to the source gateway. The design goal aims at preventing proxy or any intermediate gateways from knowing the origination of any particular network traffic flow.

1) Uplink Routing Design: There exists several secure routing protocols to preserve source node location privacy/anonymity, such as the onion routing [19] and the random walk [20]. However, these protocols introduce large computational cost and energy consumption, which are not suitable in the resource-constraint smart home IoT environments. In light of this, we leverage the DRW mechanism [21], which is developed from the pure random walk, as our uplink routing scheme to preserve the privacy of source gateways.

Comparing to the pure random walk scheme that may generate a routing loop as shown in Fig. 2, the hop-based DRW is more energy efficient as it directs the traffic less randomly toward the destination. The hop-based DRW is feasible in our scenario because each smart home stores the whole network topology and the scheme works as follows. For each smart home gateway x, it checks its hop distance to the destination (i.e., proxy gateway z) and also its neighbors' hop distance to the destination. Then, the gateway x divides its neighbors into two sets, one of which stores the neighbors of equal or larger hop distance to the destination, while the other records the neighbors of smaller hop distance to the destination. The gateway x randomly selects a node x' from the latter set with equal probability as its next-hop node. The algorithm runs on every intermediate node till the destination proxy gateway z. At the source gateway x, the following encrypted message bundle should be created and transmitted following the above routing protocol:

$$\gamma_{x\to z} = \left(\mathrm{pid}_x|\mathrm{ul\_msg}|\mathrm{rn}|\mathrm{pvt}|(\mathrm{ttl})_{pk_{\mathrm{pvt}}}\right)_{pk_z}|z.$$

The pid_x indicates the source of sensory traffic so that the returned control message could be sent back correspondingly. ul_msg is the encrypted sensory traffic from x; rn, pvt, and ttl represent a random number, the pivot gateway address, and the time-to-live (TTL) value, respectively. These three parameters are utilized in the downlink transmissions and we will elaborate later in this section. Here, all these data are concatenated and encrypted using the proxy gateway's public key pk_z to provide confidentiality along the multihop transmissions, while the address of proxy gateway is in plaintext for the intermediate gateways to perform routing.

Following the hop-based DRW routing protocol, the sensory traffic reaches the proxy gateway along the minimum-hop path. For any source-destination pair (x, z), suppose the number of hops along this path is n, then the energy consumption for uplink transmissions can be calculated as

$$U_{\mathrm{ul}}(x, z) = n(E_{rx} + E_{\mathrm{tx}}) \tag{8}$$

where E_{re} and E_{tx} represent the receiving and transmitting energy consumption measured in J/bit, respectively.

2) Downlink Routing Design: In contrary to the uplink routing design, the reverse transmission should provide the location privacy/anonymity for the destination, which is the previous source smart home gateway x. With this in mind, the prior hop-based DRW mechanism is clearly infeasible as the destination address is in plaintext. Therefore, in this section, we introduce a hybrid of the hop-based DRW and the flooding mechanism to realize the secure multihop routing.

Compared to the pure flooding scheme [22] where the proxy gateway floods the downlink control messages to the whole network, our design is more energy efficient in the sense we initiate the flooding at an intermediate node (also known as the pivot gateway) which is in close proximity to the source smart home so that the flooding mechanism only affects a subset of the network nodes. However, we shall analyze later that despite energy efficiency, this design could let the pivot gateway confine the source gateway in a smaller geographical area which reduces the privacy level. Clearly, there is a tradeoff between utility and privacy and therefore, we will investigate the pivot gateway selection problem in later part of this section so as to achieve a sound balance between energy efficiency and privacy.

First, we discuss the basics of the downlink secure routing protocol as follows. Upon receiving the control message $\beta = \mathrm{pid}_x|\mathrm{dl\_msg}$ from the cloud or remote homeowners, where dl_msg could be encrypted, the proxy gateway z obtains rn and pvt by decrypting the prior uplink traffic using its private key sk_z. The pvt indicates the address of the pivot gateway, from where the data will be flooded. The proxy gateway then creates the following bundle and sends it to the pvt following the prior hop-based DRW routing protocol:

$$\gamma_{z\to\mathrm{pvt}} = (\mathrm{pid}_x|\mathrm{dl\_msg})_{\mathrm{rn}}|(\mathrm{ttl})_{pk_{\mathrm{pvt}}}|\mathrm{pvt}.$$

When the pivot gateway receives $\gamma_{z\to\mathrm{pvt}}$, it decrypts ttl using its private key sk_{pvt} and then floods the message $(\mathrm{pid}_x|\mathrm{dl\_msg})_{\mathrm{rn}}$ into the smart community network following the TTL requirement which was initially specified by the source gateway. For the victim nodes in the flooding area,

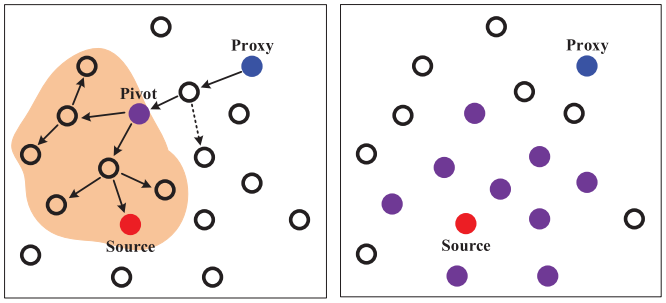

Fig. 3. (a) Mechanism of hybrid DRW and flooding with the design parameter $<s, v> = <2, 2>$. (b) All the candidate pivot gateways in purple circles.

only the source smart home gateway x can decrypt dl_msg using its rn. The downlink routing protocol is also described in Fig. 3(a).

It is important to note that selecting the pivot gateway is a nontrivial task. To be specific, suppose the source gateway x selects a pivot gateway pvt which is s-hop away from the proxy gateway and v-hop away from itself. On one hand, it is obvious that the tuple $<s, v>$ has a great impact on the network energy consumption. On the other hand, since the TTL information is known at the pivot gateway, the location of the source gateway x could be confined in a small area. For instance, if the TTL is measured in hop count, the pivot gateway is confident that the source gateway is one of her TTL-hop neighbors. Clearly, there exist a tradeoff between utility and privacy, but the DP mechanism standalone is insufficient in defending against the inference attack from the pivot node, which is illustrated as the following theorem.

Theorem 1: With the adversaries's improved prior knowledge, the DP mechanism can only bound the adversaries' knowledge gain, but is incapable of providing absolute protection against Bayesian inference attacks.

Proof: See the Appendix. ■

To give a insight into this tradeoff and facilitate our following design, we generate a random connected graph of 20 nodes in a geographic area of 280×220 m^2 as shown in Fig. 4(a), execute the Bayesian inference attack at two candidate pivot nodes, and examine how the selection of pivots and the design of the flooding mechanism could jointly impact the privacy level of the source node and the network energy consumption. The DP obfuscation distributions are obtained through solving problem (7), and the privacy level and energy consumption are displayed in Fig. 4(b) and (c), respectively. Clearly, the pure flooding over the whole network gives the source node the highest privacy level, but also generates the largest energy consumption. In contrary, flooding in a confined area saves energy but compromising the privacy of the source node. It is also interesting to point out that selecting pivot 2 for flooding gives higher privacy than selecting pivot 1, since the density of the neighborhood nodes is higher at pivot 2, which is captured by the concept *privacy mass* in [23].

To this end, we propose a term called *personalized privacy bound* ρ_x, which is explicitly chosen as the lower bound by each smart home $x \in \mathcal{X}$ to defend against inferences from

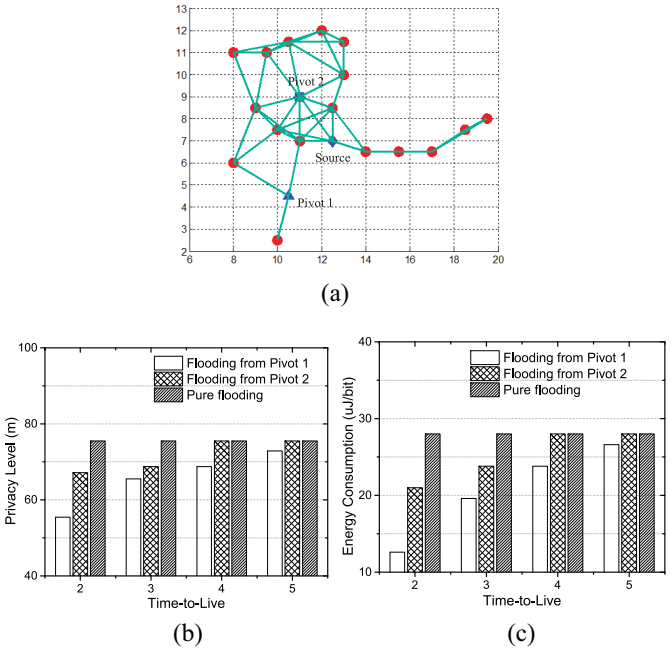

Fig. 4. Demonstrative example showing how the selection of pivot node and design of flooding scheme jointly impact the privacy and energy consumption. (a) Randomly generated connected graph of 20 nodes. (b) Privacy level of the source node. (c) Network-level energy consumption.

pivot gateways. Here, we select the pivot gateway and design the flooding mechanism in such a way that the network energy consumption is minimized while each source gateway's personalized privacy bound is satisfied. Suppose for each $x \in \mathcal{X}$, the set of candidate pivot gateways Φ_x is determined from the connected network graph using x's adjacency matrix, for instance as shown in Fig. 3(b). Then for each candidate pivot gateway $\text{pvt}_i^x \in \Phi_x$, we associate it with a decisional variable $\omega_i^x \in \{0, 1\}$, where $\omega_i^x = 1$ indicates pvt_i^x is selected and 0 if not selected. For each candidate pivot gateway pvt_i^x, denote the set of its infected network nodes due to flooding as $\mathcal{K}(\text{pvt}_i^x, \text{ttl})$. Then the pivot selection and flooding design problem can be cast into the following optimization problem:

$$
\begin{aligned}
\text{Min} \quad & \sum_{i=1}^{|\Phi_x|} \omega_i^x \cdot |\mathcal{K}(\text{pvt}_i^x, \text{ttl})| \cdot (E_{\text{tx}} + E_{rx}) \\
\text{s.t.} \quad & \sum_{i=1}^{|\Phi_x|} \omega_i^x \sum_{z\in\mathcal{X}} \mathcal{A}_x(z) \sum_{x'\in\mathcal{K}(\text{pvt}_i^x,\text{ttl})} h(x'|z)\cdot d(x, x') \geq \rho_x \\
& \sum_{i=1}^{|\Phi_x|} \omega_i^x = 1 \quad \forall x \in \mathcal{X} \\
& \omega_i^x \in \{0, 1\} \quad \forall x \in \mathcal{X} \\
& x \in \mathcal{K}(\text{pvt}_i^x, \text{ttl}) \quad \forall \text{ttl} \in \mathbb{Z}^+
\end{aligned} \tag{9}
$$

where $h(x'|z) = ([\mathcal{A}_{x'}(z)\psi(x')]/[\sum_{y\in\mathcal{K}(\text{pvt}_i^x,\text{ttl})} \mathcal{A}_y(z)\psi(y)])$ denotes the posterior probability derived at the pivot node, and $\mathbb{Z}^+$ represents the set of positive integers. Here, each victim node may receive the same packets from its several neighbors but we assume it only transmits the packets once. Besides, for simplicity, we utilize $|\mathcal{K}(\text{pvt}_i^x, \text{ttl})|E_{rx}$ to denote

Algorithm 1 Heuristic Algorithm for Solving Problem (9)

Input: Network topology of the smart community, v, obfuscation distribution $\mathcal{A}_x(z)$, E_{tx}, E_{rx}, and personalized privacy level ρ_x.

Output: The design tuple (pvt^x, ttl^x).

for i = 1 : $|\mathcal{X}|$ **do**
  Construct x_i's v-hop neighbour nodes in set Φ_{x_i};
  for j = 1 : $|\Phi_{x_i}|$ **do**
    $Utility_j^{x_i} \leftarrow 0$; $ttl_j^{x_i} \leftarrow 1$;
    while $Utility_j^{x_i} == 0$ **do**
      Construct $pvt_j^{x_i}$'s $ttl_j^{x_i}$-hop neighbour nodes in set $\mathcal{K}\left(pvt_j^{x_i}, ttl_j^{x_i}\right)$;
      if $x_i \in \mathcal{K}\left(pvt_j^{x_i}, ttl_j^{x_i}\right)$ & $privacy_j^{x_i} \geq \rho_{x_i}$ **then**
        $Utility_j^{x_i} \leftarrow |\mathcal{K}\left(pvt_j^{x_i}, ttl_j^{x_i}\right)| \cdot (E_{tx} + E_{rx})$;
      else
        $ttl_j^{x_i} \leftarrow ttl_j^{x_i} + 1$;
      end if
    end while
  end for
  $(pvt^{x_i}, ttl^{x_i}) = \arg \min_{\left(pvt_j^{x_i}, ttl_j^{x_i}\right)} \left\{Utility_j^{x_i}\right\}$;
end for

the total receiving power consumption. Due to the integer nature of variable ω_i^x and the discrete nature of $\mathcal{K}(\text{pvt}_i^x, \text{ttl})$, the optimization problem (9) is not easily tractable. However, observing the non-decreasing property of $\mathcal{K}(\text{pvt}_i^x, \text{ttl})$ with respect to ttl for each pvt_i^x, we propose a heuristic searching algorithm for problem (9) in a cost-effective manner. The algorithm is described in Algorithm 1.

The basic idea of the Algorithm 1 is to search the candidate pivot gateways for the ones satisfying the first and last constraint in problem (9). Specifically, the $\text{privacy}_j^{x_i}$ equals to $\sum_{z\in\mathcal{X}} \mathcal{A}_x(z) \sum_{x'\in\mathcal{K}} h(x'|z)d(x, x')$ while $x_i \in \mathcal{K}$ enforces the source gateway being in the flooding region. The worst-case complexity of addressing problem (9) using Algorithm 1 is $\mathcal{O}(N^2M)$, where N and M represent the number of nodes in the network and the maximum hop distance between any two nodes, respectively. Note that problem (9) may not have solutions as the achievable maximum privacy level is bounded by the size of the network, which means the pure flooding mechanism provides the privacy upper bound. Therefore, ρ_x should be selected lower than that. On the other hand, for the case that the DP obfuscation gives the strategy where proxy gateway is the source gateway itself, the prior design for uplink/downlink secure multihop routing is not necessary.

Upon selecting the pivot gateway for each source gateway x, we can calculate its hop distance to any potential proxy gateway $d(\text{pvt}^x, z)$ using for instance Dijkstra algorithm. Thus, the network energy consumption for the downlink multihop transmissions can be calculated as

$$U_{\text{dl}}(x, z) = d\left(\text{pvt}^x, z\right) \cdot (E_{rx} + E_{\text{tx}}) + \text{Utility}^x. \tag{10}$$

Therefore, the utility loss function, which indicates the overall round-trip energy consumption, in optimization problem (7) can be obtained as follows:

$$U(x, z) = U_{\text{ul}}(x, z) + U_{\text{dl}}(x, z). \tag{11}$$

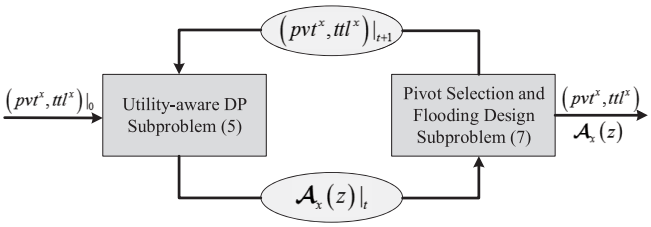

Fig. 5. Diagram for the iterative algorithm.

Given all the utility measures are quantified, we could solve problem (7) and obtain the DP obfuscation $\mathcal{A}$ for each smart home. However, it should be noted that an iterative algorithm is needed as the problem (7) and (9) are mutually dependent on each other. In light of this, we propose an iterative algorithm as shown in Fig. 5, where a DP obfuscation distribution is obtained given a certain system initialization and the algorithm terminates when the selected pivot gateway from the prior step is the same as the one in the current step. In other words, the solution converges as $(\text{pvt}^x, \text{ttl}^x)|_t = (\text{pvt}^x, \text{ttl}^x)|_{t+1}$ for any $x \in \mathcal{X}$.

V. Performance Evaluation

A. Privacy Analysis

First of all, to assess the effectiveness of our DP mechanism in protecting smart homes, we utilize (2) to quantitatively show how the privacy of each smart home can be protected against the Bayesian inference attacks in the later section.

As for the HbC smart homes in the wireless environment, the uplink DRW scheme only allows any intermediate node and the proxy gateway to learn about their preceding gateway, but being unaware of the number of hops between source and destination, they cannot find the source smart home. Furthermore, without any collusion between smart homes, the randomness grows rapidly for the latter gateways, especially the proxy gateway, due to the multiple possible paths resulted from the random selection of next-hop node in the DRW scheme. On the other hand, the downlink hybrid mechanism of DRW and flooding also achieves unlinkability between source and destination gateways. Any intermediate node including the pivot gateway is unaware of the proxy gateway, while they can neither know the source smart home due to the flooding mechanism. However, the pivot gateway could confine the location of the source gateway in a small area. Later, we will quantitatively examine how our design could protect smart homes from pivot gateway's inference attacks.

B. Numerical Evaluation

1) Simulation Setup: In this section, we evaluate the performance of proposed framework based on a community in Gainesville, FL, USA. The geographical map is shown in Fig. 6, where a total of 77 homes are distributed in a 640 × 300 m^2 area. We adopt the widely accepted channel

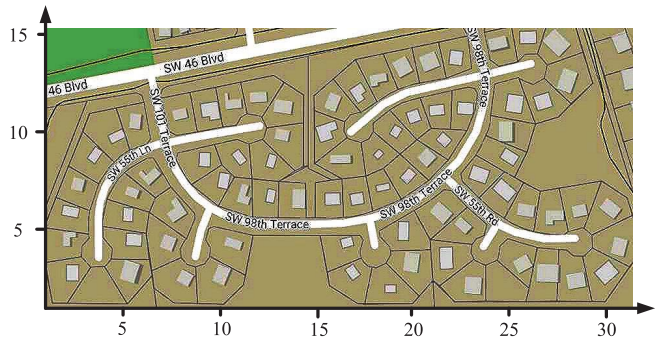

Fig. 6. Geographical view of a smart community.

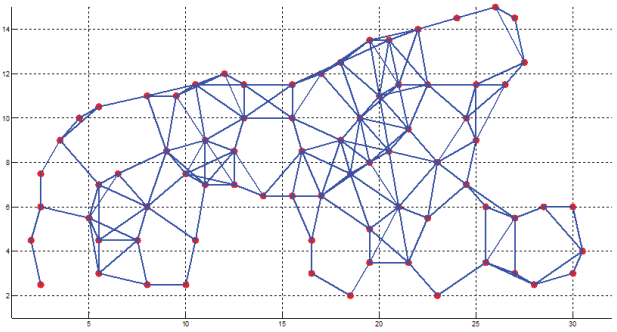

Fig. 7. Connected graph of the smart community.

model [24] $P_{rx} = \eta d(x, y)^{-\zeta} \cdot P_{\text{tx}}$ and assume transmit power P_{tx} of 5 W, antenna gain η of 5.78 and path loss factor ζ of 4. A communication link exists between two smart home routers if the received power exceeds the threshold P_{rx}^{th}. Thus, by setting $P_{rx}^{\text{th}} = 2 \times 10^{-6}$ W, we can obtain that the transmission range of any smart home router is 61 m. Given this setup, the connected graph can be generated accordingly and it is shown in Fig. 7. Furthermore, we consider each smart home gateway generates uplink encrypted sensory traffic of rate 15 Mb/s, and the computing capability for each smart home gateway is 100 Mb/s. The energy consumption for transmitting and receiving data are 0.8 and 0.6 μJ/bit, respectively. The delay requirement for each smart home is assume to be 15 μs for both uplink and downlink transmissions, and the processing and transmission delays are 3 and 0.2 μs, respectively. Moreover, we set the source gateway selects the pivot gateway from its two-hop neighbor nodes. In other words, $v = 2$.

2) Benchmark Mechanism: To compare the effectiveness of our mechanism in protecting user privacy and preserving network utility, we put forward a benchmark differential privacy mechanism, which is constructed using exponential distribution [16]. Specifically, for any smart home $x \in \mathcal{X}$, the probability of selecting $z \in \mathcal{X}$ as its proxy gateway equals to

$$\mathcal{J}_x(z) = e^{\left(\frac{-\varepsilon d(x,z)}{2\Delta u}\right)} \Big/ \sum_{y \in \mathcal{X}} e^{\left(\frac{-\varepsilon d(x,y)}{2\Delta u}\right)} \tag{12}$$

where Δu is the sensitivity of the utility function, which captures the intuition that the longer the distance, the larger the energy consumption (i.e., utility loss). Δu is calculated as follows:

$$\Delta u = \max_{z \in \mathcal{X}} \max_{x,y \in \mathcal{K}} |d(x, z) - d(y, z)|. \tag{13}$$

According to the triangle inequality, for any two nodes $x, y \in \mathcal{K}$ in the confined region, $|d(x, z) - d(y, z)| \leq D(\mathcal{K})$, where $D(\mathcal{K})$ is the diameter or the longest distance between two nodes of any particular region $\mathcal{K}$, so here the sensitivity Δu is set as $D(\mathcal{K})$.

3) Performance Analysis: First, we conduct simulations to examine the performance of our pivot selection and flooding design in problem (9). As shown in Fig. 8(a), the incurred network energy consumption by each smart home is given. We can see that with the increase of ρ_x, the network energy consumption increases, which is for the reason that the flooding area becomes larger in order to satisfy the higher personalized privacy bound. In addition, we shall see that for the same ρ_x, the network energy consumption slightly decreases by reducing the DP parameter ε. The rationale comes from Theorem 1 and the derivations in the Appendix, where we have seen that the smaller the ε, the tighter the DP mechanism bounds the prior and posterior probability. Therefore, with the smaller ε, the Bayesian inference attack using the corresponding posterior probability distribution $h(\cdot)$ causes larger inference error compared with the one with the larger ε. In light of this, to achieve a certain privacy bound ρ_x in our design (9), a small ε helps alleviate the necessity to generate a large flooding area, which as a result reduces the network energy consumption.

On the other hand, Fig. 8(b) demonstrates the convergence performance when we run the iterative algorithm as shown in Fig. 5. Here, ρ_x is set as 40 m. As we can see, the iterative algorithm converges in a few steps and the convergence rate is dependent on the design parameter ε. First of all, the reason that the converged solution gives a higher energy consumption than the initial value is that the design in problem (9) aims to provide the bound on user's privacy level, which as a result sacrifices the network utility. Second, the larger the ε, the larger variation the DP obfuscation distribution $\mathcal{A}_x(z)$ as the bounding effect of the DP property (3). Therefore, the iterative algorithm takes more steps to converge given a larger value of ε.

Next, we compare our design with the benchmark mechanism and evaluate how these two designs impact the network performances, such as energy consumption, round-trip delay, and user privacy. First of all, the ratio of the energy consumption incurred by exponential DP mechanism over the one caused by our proposed utility-aware DP mechanism is shown in Fig. 9(a). It is obvious that the ratio is always greater than 1 regardless of the design parameter we select, meaning that the exponential DP mechanism is (around 10%) less energy-efficient than our utility-aware DP mechanism. On the other hand, with the increase of ρ_x or ε, the advantage of our mechanism over the exponential DP mechanism slightly decreases for the similar reason we demonstrated before. Second, we show in Fig. 9(b) the user's privacy level in terms of the Bayesian inference error measured in meters. Clearly, our utility-aware DP mechanism gives higher privacy level than

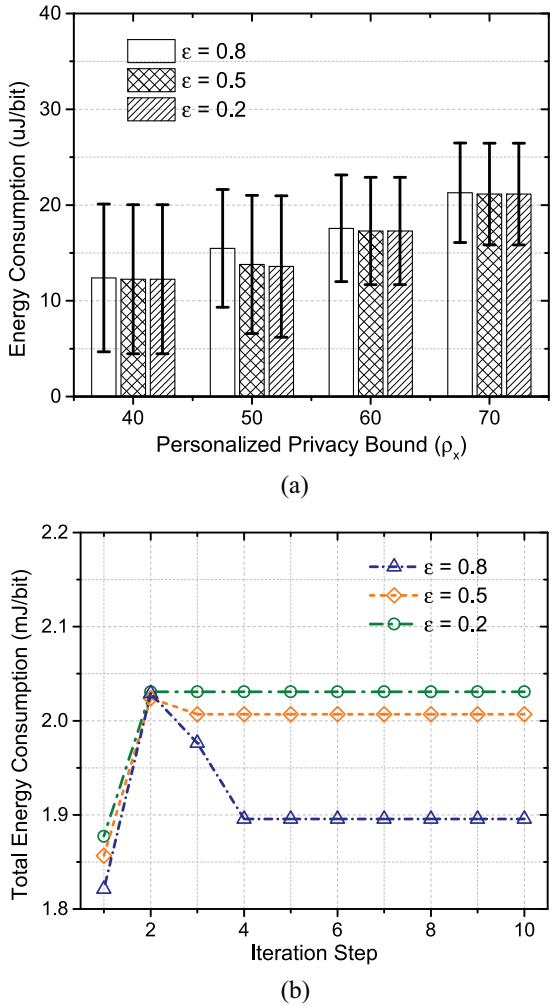

Fig. 8. Performance of the pivot selection and flooding mechanism design. (a) Per-user utility cost. (b) Convergence analysis.

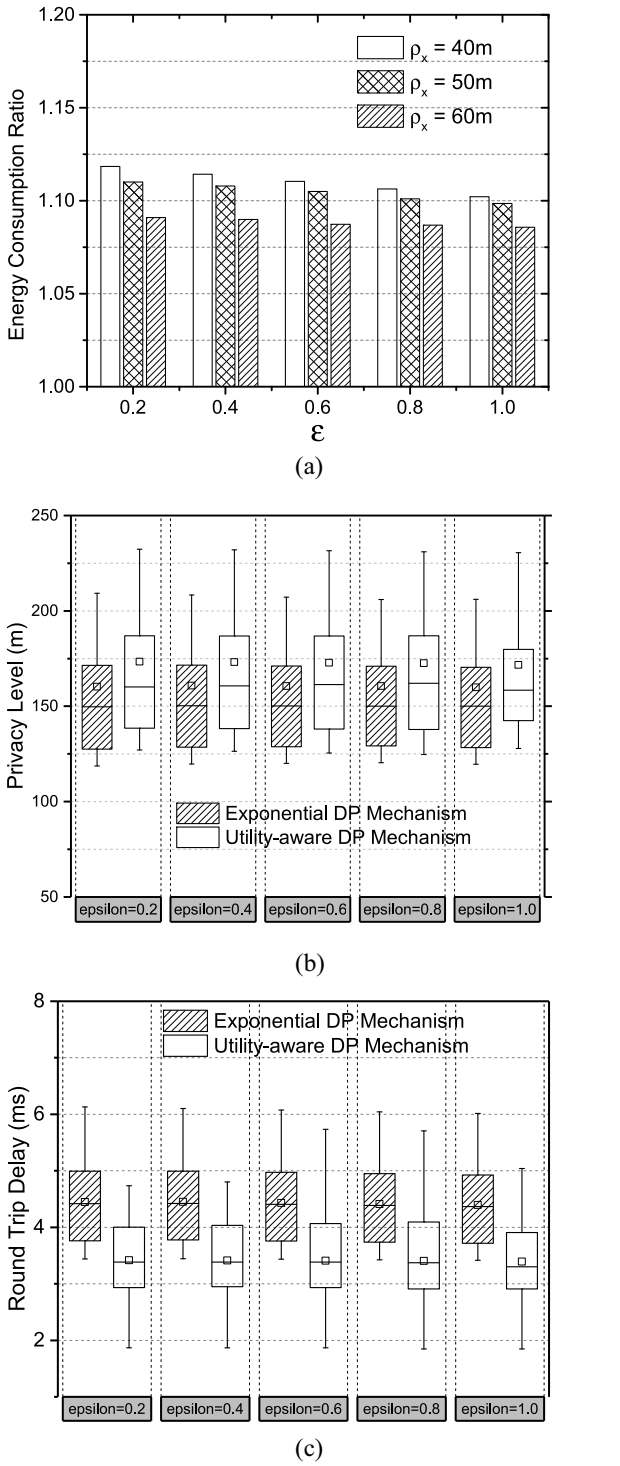

Fig. 9. Performance comparison between two DP mechanisms. (a) Network energy consumption ratio. (b) Per-user privacy level. (c) Round-trip delay for each user.

the exponential mechanism. Besides, with the increase of ε, the privacy level slightly reduces for both mechanisms because of the loose bound of the obfuscation probability distribution. Third, as for the round-trip delay, it can be seen that the exponential DP mechanism incurs a higher value than our utility-aware DP mechanism, the reason is easily obtainable by observing Fig. 10, where the exponential DP mechanism generates a roughly uniform distribution meaning a smart home could possibly select a proxy gateway that is far away from itself which results in huge round-trip delay. Moreover, the delay decreases with the increase of ε for the same reason we presented before.

To give a deep insight of these two mechanisms, we randomly select 20 out of 77 smart homes and show how they differ in generating obfuscation probability distributions. The result is shown in Fig. 10, where nodes 1 and 2 are geographically in close proximity but are far from node 20. We can see that $\mathcal{A}_x$ for nodes 1 and 2 are more similar to each other than $\mathcal{A}_x$ for nodes 1 and 20 or nodes 2 and 20. Besides, there are only five candidate proxy gateways (nodes 1, 9, 11, 13, and 16) for them to select from, which is due to the optimization in (7) for utility concern. On the other hand, the exponential mechanism applied to the same set of smart homes gives a relatively uniform probability distribution $\mathcal{J}_x$, which inevitably results in a higher utility cost and longer round-trip delay.

Last but not least, we conduct simulations to examine how the system parameter could impact the performance of our framework. The result is shown in Fig. 11, where we can

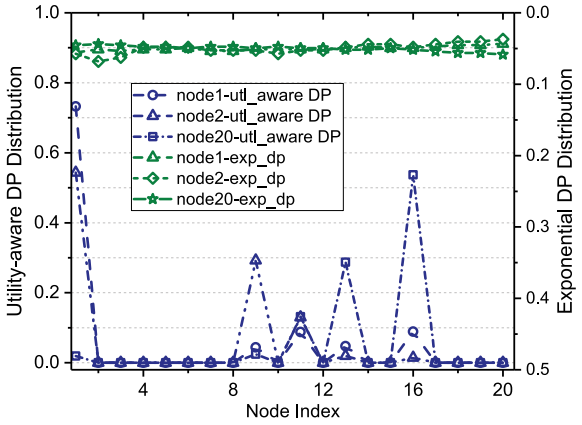

Fig. 10. Comparison of the obfuscation distribution generated by two DP mechanisms.

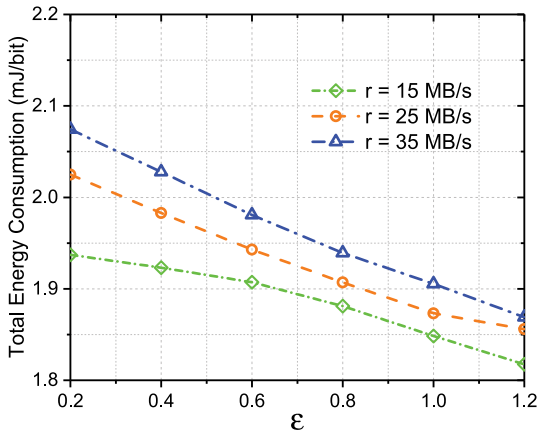

Fig. 11. Energy consumption with respect to the system parameter.

observe that with the increase of smart home data rate r, the total network utility cost increases due to the computing resource constraint at smart home gateways, which negatively impacts the selection of proxy gateways; whereas the total utility loss reduces as the increase of ε, meaning a higher ε reduces the utility cost (in terms of both energy consumption and delay), which coincides with our observation at Fig. 9(b) and (c). However, it is interesting to note that a higher ε causes privacy loss as observed in Fig. 9(b), which sheds the light on the fact that the tradeoff between utility and privacy should be neatly designed via the tuning knob ε.

VI. Related Work

Traffic analysis is by no means a new area of research, and indeed there have been many research efforts in this domain. For instance, by observing encrypted Internet traffic, some valuable information like the language of a VoIP call [25] and the passwords in secure shell logins [26] could be leaked to the network eavesdropper. In this realm, without directly accessing the data contents, adversaries leverage the protocol level information (also known as side channel information), such as packet lengths, IP address/TCP port numbers, and packet interarrival timings, to recover a certain level information about the data traffic or devices in the network.

As more and more connected devices feature in everyday objects, the IoT has become a reality giving users the most convenience, but the traffic analysis attack is still applicable in most IoT applications [7], in particular the smart homes where WiFi is used for connectivity [9]. At first glance, it may seem that the existing countermeasures for traffic analysis could be applied in the smart home IoT environments. However, there are nuances of this application scenario where the countermeasures should be properly designed. In what follows, we survey recent work in this area.

Wright *et al.* [27] proposed a technique called *traffic morphing* where one class of traffic is intentionally modified (via chopping and padding) to look like another class. An optimal morphing strategy is derived by solving a convex optimization problem, which aims to reduce the accuracy of traffic classifier while incurring less overhead. Iacovazzi and Baiocchi [28] proposes a similar technique named *traffic masking* via padding, fragmentation or dummy messages to camouflage the original traffic pattern (e.g., traffic burst). Feghhi and Leith [29] presented several traffic flow obfuscation mechanisms such as injecting dummy request packets or delaying response packets to defend against timing analysis attack. In short, the common tactic for mitigating the threat of traffic analysis is to obfuscate the traffic patterns like packet sizes, timing and other statistical features of the traffic flow. However, a recent study by Dyer *et al.* [30] demonstrated that these countermeasures are vulnerable to simple attacks using naïve Bayes-based classifier that use coarse features of traffic (e.g., total time and bandwidth). Unfortunately, with the advance of machine learning algorithms and more traffic features being exploited, it is getting even harder to defend traffic analysis attacks by purely relying on traffic pattern obfuscation approaches. Moreover, obfuscating a wide range of traffic patterns also incurs significant overhead in terms of delay and bandwidth, which is undesirable in resource-constrained IoT environments.

Based on the aforementioned research work, it is clear to see that the smart home standalone is incapable of defending the traffic analysis attacks due to its limited resources and the inefficacy of the existing countermeasure techniques. In light of this, leveraging the network level approach via collaborating with other devices seems to be the viable way to pursue.

VII. Conclusion

In this paper, we presented an EPIC framework to defend smart homes against the traffic analysis attack. We focused on a resource-constrained IoT environment and exploited the smart community network of wirelessly connected smart homes to perform local traffic obfuscation for each individual smart home. In particular, we designed a utility-optimal differential privacy mechanism to obfuscate the source of traffic flows. A hostile wireless environment was also considered so we developed a secure and privacy-preserving multihop routing scheme to guarantee the source/destination unlinkability and to satisfy the user's personalized privacy bound. Extensive simulations were conducted and our framework showed advantages over the benchmark mechanism in

protecting smart home's privacy and in reducing the network energy consumption.

APPENDIX
PROOF OF THEOREM 1

Suppose $\mathcal{X}$ is the initial protection region, and $\mathcal{R}$ is the region after adversaries excludes a node $\Delta \in \mathcal{X}$ from $\mathcal{X}$, in other words $\mathcal{R} = \mathcal{X}\backslash\Delta$, representing the adversaries' improved knowledge. Thus, for any observed output $z \in \mathcal{X}$, the posterior probability can be constructed to invert the input $x \in \mathcal{R}$ as follows:

$$\begin{aligned} h(x|z) &= \frac{\mathcal{A}_x(z)\frac{\psi(x)}{\sum_{y\in\mathcal{R}}\psi(y)}}{\sum_{x'\in\mathcal{R}}\mathcal{A}_{x'}(z)\frac{\psi(x')}{\sum_{y\in\mathcal{R}}\psi(y)}} \\ &= \frac{\mathcal{A}_x(z)\psi(x)}{\sum_{x'\in\mathcal{R}}\mathcal{A}_{x'}(z)\psi(x')} \\ &= \frac{\psi(x)}{\sum_{x'\in\mathcal{R}}\frac{\mathcal{A}_{x'}(z)}{\mathcal{A}_x(z)}\psi(x')} \\ &\le \frac{\psi(x)}{\psi(x)+\sum_{x'\in\mathcal{R},x\neq x'}e^{-\varepsilon d(x,x')}\psi(x')} \\ &\le \frac{\psi(x)}{e^{-\varepsilon D_{\mathcal{R}}}\sum_{x'\in\mathcal{R}}\psi(x')} \quad \text{or} \quad \frac{\psi(x)}{e^{-\varepsilon}\sum_{x'\in\mathcal{R}}\psi(x')}. \end{aligned} \tag{14}$$

The first inequality is due to the definition of differential privacy (3). The last inequality, respectively, represents the Euclidean and Hamming distance measures, and $D_{\mathcal{R}}$ is the diameter or longest distance between any two nodes in region $\mathcal{R}$. From (14), we see that the DP mechanism bounds the multiplicative distance between the prior and posterior probability regardless of what specific prior knowledge the adversaries may have.

Next, we examine how the improved prior knowledge facilitates the Bayesian inference attack in reducing user's absolute privacy level. Following (2), for any $x \in \mathcal{R}$, we prove the privacy metric is monotonically decreasing with respect to the size of the protection region $\mathcal{R}$. First of all, let $h(\Delta|z) = \alpha$ for presentation clarity

$$\begin{aligned} &\text{privacy}(x;\mathcal{R}) \\ &= \sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\sum_{x'\in\mathcal{R}}\frac{\mathcal{A}_{x'}(z)\psi(x')}{\sum_{y\in\mathcal{R}}\mathcal{A}_y(z)\psi(y)}d(x,x') \\ &= \sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\sum_{x'\in\mathcal{X}}\frac{\mathcal{A}_{x'}(z)\psi(x')}{\sum_{y\in\mathcal{R}}\mathcal{A}_y(z)\psi(y)}d(x,x') \\ &\quad -\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\frac{\mathcal{A}_\Delta(z)\psi(\Delta)}{\sum_{y\in\mathcal{R}}\mathcal{A}_y(z)\psi(y)}d(x,\Delta) \\ &= \sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\sum_{x'\in\mathcal{X}}\left[\frac{\sum_{y\in\mathcal{X}}\mathcal{A}_y(z)\psi(y)}{\sum_{y\in\mathcal{R}}\mathcal{A}_y(z)\psi(y)}\right]\frac{\mathcal{A}_{x'}(z)\psi(x')}{\sum_{y\in\mathcal{X}}\mathcal{A}_y(z)\psi(y)} \\ &\quad \times d(x,x') - \sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\frac{\mathcal{A}_\Delta(z)\psi(\Delta)}{\sum_{y\in\mathcal{R}}\mathcal{A}_y(z)\psi(y)}d(x,\Delta) \\ &= \sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\sum_{x'\in\mathcal{X}}[1+\alpha]\frac{\mathcal{A}_{x'}(z)\psi(x')}{\sum_{y\in\mathcal{X}}\mathcal{A}_y(z)\psi(y)}d(x,x') \\ &\quad -\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\cdot\alpha\cdot d(x,\Delta). \end{aligned} \tag{15}$$

Therefore, the privacy loss can be calculated as follows:

$$\begin{aligned} &\text{privacy}(x;\mathcal{X}) - \text{privacy}(x;\mathcal{R}) \\ &= \sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\left\{\alpha\cdot\left[d(x,\Delta) - \sum_{x'\in\mathcal{X}}h(x'|z)d(x,x')\right]\right\}. \end{aligned} \tag{16}$$

1) *Case 1:* When $d(\cdot,\cdot)$ is in measured in Hamming distance, (16) is rewritten as $\sum_{z\in\mathcal{X}}\mathcal{A}_x(z)\{\alpha\cdot[1-(1-\alpha)]\}$, which is greater than zero.
2) *Case 2:* When $d(\cdot,\cdot)$ is in measured in Euclidean distance, $\sum_{x'\in\mathcal{X}}h(x'|z)d(x,x')$ is the weighted geometric median of region $\mathcal{R}$. Under the mild assumption that $\mathcal{R}$ and $\mathcal{X}$ are both convex regions, $d(x,\Delta) > \sum_{x'\in\mathcal{X}}h(x'|z)d(x,x')$.

Thus, $\text{privacy}(x;\mathcal{X}) > \text{privacy}(x;\mathcal{R})$, meaning the inference error decreases as the protection region shrinks. On the other hand, $\text{privacy}(x;\mathcal{R}) = 0$ if $\mathcal{R} = \{x\}$ as $d(x,x) = 0$ for both distance measures. Therefore, we can claim that privacy level is monotonically decreasing with the minimum privacy level being zero.

To this end, Theorem 1 is proven.

ACKNOWLEDGMENT

The authors would like to thank the anonymous reviewers and the editor in providing constructive suggestions.

REFERENCES

[1] L. Atzori, A. Iera, and G. Morabito, "The Internet of Things: A survey," *Comput. Netw.*, vol. 54, no. 15, pp. 2787–2805, 2010.
[2] T. Denning, T. Kohno, and H. M. Levy, "Computer security and the modern home," *Commun. ACM*, vol. 56, no. 1, pp. 94–103, 2013.
[3] iHealth. (2007). *Chronic Care at the Crossroads*. [Online]. Available: http://www.intel.com/healthcare/
[4] A. Cavoukian, A. Fisher, S. Killen, and D. A. Hoffman, "Remote home health care technologies: How to ensure privacy? Build it in: Privacy by design," *Identity Inf. Soc.*, vol. 3, no. 2, pp. 363–378, 2010.
[5] NC Health Care. (2003). *Health Insurance Cost*. [Online]. Available: http://www.nchc.org/facts/cost.shtml
[6] D. J. Cook and S. K. Das, "How smart are our environments? An updated look at the state of the art," *Pervasive Mobile Comput.*, vol. 3, no. 2, pp. 53–73, 2007.
[7] Y. Yan, E. Oswald, and T. Tryfonas, "Exploring potential 6lowpan traffic side channels," *IACR Cryptol. ePrint Archive*, vol. 2017, p. 316, Apr. 2017.
[8] A. Jacobsson, M. Boldt, and B. Carlsson, "A risk analysis of a smart home automation system," *Future Gener. Comput. Syst.*, vol. 56, pp. 719–733, Mar. 2016.
[9] Y. Meidan *et al.*, "ProfilioT: A machine learning approach for IoT device identification based on network traffic analysis," in *Proc. Symp. Appl. Comput.*, Marrakesh, Morocco, 2017, pp. 506–509.
[10] J. Penders, M. Altini, C. Van Hoof, and E. Dy, "Wearable sensors for healthier pregnancies," *Proc. IEEE*, vol. 103, no. 2, pp. 179–191, Feb. 2015.
[11] X. Li *et al.*, "Smart community: An Internet of Things application," *IEEE Commun. Mag.*, vol. 48, no. 11, pp. 68–75, Nov. 2011.
[12] A. M. Rahmani *et al.*, "Exploiting smart e-health gateways at the edge of healthcare Internet-of-Things: A fog computing approach," *Future Gener. Comput. Syst.*, vol. 78, pp. 641–658, Jan. 2018.
[13] R. Shokri, G. Theodorakopoulos, J.-Y. Le Boudec, and J.-P. Hubaux, "Quantifying location privacy," in *Proc. IEEE Symp. Security Privacy (SP)*, Berkeley, CA, USA, 2011, pp. 247–262.
[14] M. E. Andrés, N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Geo-indistinguishability: Differential privacy for location-based systems," in *Proc. 20th ACM SIGSAC Conf. Comput. Commun. Security (CCS)*, Berlin, Germany, 2013, pp. 901–914.
[15] N. E. Bordenabe, K. Chatzikokolakis, and C. Palamidessi, "Optimal geo-indistinguishable mechanisms for location privacy," *in Proc. ACM SIGSAC Conf. Comput. Commun. Security*, Scottsdale, AZ, USA, 2014, pp. 251–262.

[16] L. Yu, L. Liu, and C. Pu, "Dynamic differential location privacy with personalized error bounds," in *Proc. Netw. Distrib. Syst. Security Symp. (NDSS)*, 2017.
[17] G. A. Holton, *Value-at-Risk: Theory and Practice*, vol. 39. New York, NY, USA: Academic Press, 2003.
[18] CALOMEL. (2017). *AES-NI SSL Performance*. [Online]. Available: https://calomel.org/aesni_ssl_performance.html
[19] D. Goldschlag, M. Reed, and P. Syverson, "Onion routing," *Commun. ACM*, vol. 42, no. 2, pp. 39–41, 1999.
[20] C. Ozturk, Y. Zhang, and W. Trappe, "Source-location privacy in energy-constrained sensor network routing," in *Proc. 2nd ACM Workshop Security Ad Hoc Sensor Netw.*, Washington, DC, USA, 2004, pp. 88–93.
[21] J. Yao and G. Wen, "Preserving source-location privacy in energy-constrained wireless sensor networks," in *Proc. 28th Int. Conf. Distrib. Comput. Syst. Workshops (ICDCS)*, Beijing, China, 2008, pp. 412–416.
[22] A. Vahdat and D. Becker, "Epidemic routing for partially connected ad hoc networks," Duke Univ., Durham, NC, USA, Rep. CS-2000-06, 2000.
[23] K. Chatzikokolakis, C. Palamidessi, and M. Stronati, "Constructing elastic distinguishability metrics for location privacy," *Proc. Privacy Enhancing Technol.*, vol. 2015, no. 2, pp. 156–170, 2015.
[24] J. Liu *et al.*, "An energy-efficient strategy for secondary users in cooperative cognitive radio networks for green communications," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3195–3207, Dec. 2016.
[25] C. V. Wright, L. Ballard, F. Monrose, and G. M. Masson, "Language identification of encrypted VoIP traffic: Alejandra y Roberto or Alice and bob?" in *Proc. USENIX Security Symp.*, vol. 3, 2007, pp. 43–54.
[26] D. X. Song, D. Wagner, and X. Tian, "Timing analysis of keystrokes and timing attacks on SSH," in *Proc. USENIX Security Symp.*, vol. 2001. Washington, DC, USA, 2001, Art. no. 25.
[27] C. V. Wright, S. E. Coull, and F. Monrose, "Traffic morphing: An efficient defense against statistical traffic analysis," in *Proc. NDSS*, vol. 9, 2009.
[28] A. Iacovazzi and A. Baiocchi, "Internet traffic privacy enhancement with masking: Optimization and tradeoffs," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 2, pp. 353–362, Feb. 2014.
[29] S. Feghhi and D. J. Leith, "A Web traffic analysis attack using only timing information," *IEEE Trans. Inf. Forensics Security*, vol. 11, no. 8, pp. 1747–1759, Aug. 2016.
[30] K. P. Dyer, S. E. Coull, T. Ristenpart, and T. Shrimpton, "Peek-a-boo, i still see you: Why efficient traffic analysis countermeasures fail," in *Proc. IEEE Symp. Security Privacy (SP)*, San Francisco, CA, USA, 2012, pp. 332–346.

Jianqing Liu (GS'14) received the B.Eng. degree from the University of Electronic Science and Technology of China, Chengdu, China, in 2013. He is a currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA.

His current research interests include wireless networking and network security in cyber-physical systems.

Chi Zhang (S'06–M'06) received the B.E. and M.E. degrees in electrical and information engineering from the Huazhong University of Science and Technology, Wuhan, China, in 1999 and 2002, respectively, and the Ph.D. degree in electrical and computer engineering from the University of Florida, Gainesville, FL, USA, in 2011.

He joined the School of Information Science and Technology, University of Science and Technology of China, Hefei, China, as an Associate Professor in 2011. His current research interests include network protocol design and performance analysis and network security particularly for wireless networks and social networks.

Yuguang Fang (F'08) received the M.S. degree from Qufu Normal University, Jining, China, in 1987, the Ph.D. degree from Case Western Reserve University, Cleveland, OH, USA, in 1994, and the Ph.D. degree from Boston University, Boston, MA, USA, in 1997.

He joined the Department of Electrical and Computer Engineering, University of Florida, Gainesville, FL, USA, in 2000, where he has been a Full Professor since 2005. He held a University of Florida Research Foundation professorship from 2006 to 2009, a Changjiang Scholar chair professorship with Xidian University, Xi'an, China, from 2008 to 2011, and also with Dalian Maritime University, Dalian, China, since 2015, and a guest chair professorship with Tsinghua University, Beijing, China, from 2009 to 2012.

Dr. Fang was the Editor-in-Chief of IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY from 2013 to 2017 and the *IEEE Wireless Communications* from 2009 to 2012. He is a Fellow of the AAAS.