On the Privacy Afforded by Opaque Identifiers in Traffic Monitoring (Extended Version)

Marcus Gelderie[a]
Aalen University of Applied Sciences, Beethovenstr. 1,73430 Aalen, Germany
{firstname.lastname}@hs-aalen.de

Keywords: Privacy,License Plate Recognition,Differential Privacy,Smart City,Traffic Monitoring

Abstract: We consider traffic monitoring via license plate recognition. Anonymizing license plates by substituting randomized identifiers is a common privacy enhancing strategy in this situation. However, the systematic effect of this anonymization strategy has not been fully explored. We study the information gain of an adversary upon observing such anonymized output. We find the effectiveness of randomized IDs to deteriorate with decreasing popularity of a given route. Moreover, we study the effect differential privacy has on the situation, given that an adversary must be assumed to have prior knowledge about the likelihood of various traffic patterns. We find that travel participants with a very strong preference for a given route are put most at risk.

1 INTRODUCTION

In the recent past, smart *traffic monitoring systems (TMS)* have received considerable attention from both the academic community as well as from commercial vendors (for example (Jain et al., 2019; Biswas et al., 2016; Bisio et al., 2022; Rana et al., 2021; swa, 2024; Djahel et al., 2015; Gade, 2019; Bhardwaj et al., 2022)). The aim of smart traffic monitoring systems is to automatically collect data that are then used to facilitate other use-cases, such as traffic management or city planning. There are many different ways in which a TMS can be built. The kind of data that it records depends, in part, on that architecture (Rana et al., 2021; Jain et al., 2019). Examples include systems based on *digital image processing (DIP)* (Krishnamoorthy and Manickam, 2018), vehicle-to-X networks (Du et al., 2015), probe vehicles (Feng et al., 2014), and even drone-based approaches (Bisio et al., 2022). DIP-based approaches do not require any cooperation from the vehicles and are thus compatible with existing vehicle technology. When combined with *license-plate recognition (LPR)* software (Jain et al., 2019; Du et al., 2013), they offer the possibility to track vehicles through a city and build statistics from these data.

Privacy is a particular concern when LPR is used. First, the license plate clearly is a piece of personally identifiable information (PII) — unlike, for example, a birds-view snapshot of a busy intersection. Second, indiscriminate scanning of license plates is difficult to base on explicit consent (which is relatively straightforward to do when using vehicle-to-X networks, for instance). As a result, solutions relying on LPR usually employ some form of pseudo-anonymization (see e.g. (Gao et al., 2019)). Existing commercial solutions rely on anonymization techniques (swa, 2024), replacing the license plate by an opaque, pseudo-random ID for each vehicle that is used to correlate individual locations at a central (e.g. cloud) service. Using pseudo-random IDs in this way is a relatively straightforward technique that is simple to implement. This approach can also be augmented with differential privacy (Gelderie. et al., 2024). Yet its security properties are not fully understood (see also *Related Work* below).

We study the question: "What information can be inferred from the anonymized data collected in this way, if the attacker has prior knowledge on the traffic patterns?" More specifically, we study this question in two settings: In its most basic nature, the data is simply anonymized and transmitted to a server. We call this the *anonymization scenario*. This scenario corresponds to what we have observed in commercial products and literature (see Related Work for details). By contrast, the *differential privacy scenario* considers the case where the obfuscated data that reaches the server satisfies the requirements of *differential privacy* (Dwork, 2006; Dwork et al., 2006). Our main contribution is a precise mathematical treatment of

[a] https://orcid.org/0009-0003-0291-3911

these scenarios in the framework of information theory.

In the anonymization scenario, we ask: Can the server de-anonymize the data given prior knowledge about a victim's route preferences? This is the well-known *de-anonymization attack*. Given an opaque ID and the route which it travels (both are part of the data-set such a server consumes as part of its operation), we show that an adversarial server learns information inversely proportional to the routes popularity (the number of cars expected to drive on it at a given time).

In the differential privacy scenario, the server sees data that is anonymized *and* subjected to randomized noise. In accordance with the typical definition of differential privacy, the server receives a given data set with roughly equal probability regardless of whether some specific individual drove along some route or not. We ask: What can the server ascertain about the probability of a specific individual driving along a given route under these circumstances, given that the server has prior knowledge about the overall traffic patterns (i.e. the underlying probability distribution). We show that an attacker can infer information about a target vehicle, if said vehicle shows a strong preference for a specific route. Since this is a fairly common situation, the privacy of individuals is at increased risk, even if differential privacy is guaranteed.

The role of prior knowledge in the study of privacy is both well-known and easy to ignore. In the context of differential privacy, for example, it is already discussed in some detail in (Dwork et al., 2014) to motivate the formal definition of differential privacy). Indeed, the definition of differential privacy is intended to quantify privacy by excluding prior knowledge from the equation (see again (Dwork et al., 2014)). This is a sensible approach from a formal point of view. But in practice, prior knowledge exists and needs to be taken into account when the *de-facto* privacy of a system is to be quantified. We think that our results show that simply anonymizing data and adding differential privacy does not sufficiently protect users under all circumstances. Care must be taken to restrict monitoring to sufficiently popular routes. We hope this paper can serve as a first step in studying the actual privacy guarantees that can be offered to a travel participant by modern license plate monitoring systems in practice.

1.1 Related Work

Traffic Monitoring The area of traffic monitoring is a wide field and only a subset of the existing research is pertinent to this paper. Generally, traffic monitoring systems fall into three categories (Jain et al., 2019): i) in situ (e.g. sensors embedded into the road surface) ii) vehicular (e.g. probe vehicles or vehicular networks) iii) digital image processing (DIP). Many papers presenting specific technical solutions exist (Biswas et al., 2016; Jain et al., 2019; Feng et al., 2014; Du et al., 2015; Baran et al., 2014; Krishnamoorthy and Manickam, 2018; Bhardwaj et al., 2022; Khanna et al., 2019; Rana et al., 2021; Rizwan et al., 2016).

Of particular relevance to this paper is the third category. Again there are many works on DIP based systems (Du et al., 2013; Baran et al., 2014; Krishnamoorthy and Manickam, 2018) from the conventional pole mounted camera to even drone-based traffic monitoring systems (Bisio et al., 2022).

Altogether, these prior works highlight the relevance of analyzing and addressing gaps in privacy guarantees offered by proposed anonymization techniques.

Differential Privacy Differential Privacy (DP) was introduced by Dwork, Nissim, McSherry, and Smith in (Dwork et al., 2006; Dwork, 2006). DP has since seen sustained and intensive research activity, resulting in a slew of research studying various application domains (works pertaining to traffic monitoring and smart cities are discussed below). Of general interest are observations on the limits of DP, particularly if the adversary is assumed to have *prior knowledge* (Dwork et al., 2014).

Some works on DP investigate continual release of statistics that are built from *streams* of events (Dwork et al., 2010; Shi et al., 2011; Kellaris et al., 2014; Chan et al., 2011; Jain et al., 2023). Those works lay algorithmic foundations and provide lower bounds on the noise that is required to achieve DP. In particular, some works study the aggregation of statistics from multiply locations or agents (Bonawitz et al., 2017; Bittau et al., 2017; Cheu et al., 2019; Corrigan-Gibbs and Boneh, 2017; Chan et al., 2012). One important aspect of these works is that they classify DP algorithms as either working in a *central model* or a *local model*. In the central model, DP is implemented by a trusted party (the *curator*) that sees all data-points in the clear. This curator then outputs a perturbed version of this statistic that meets the definition of DP. In the local model, on the other hand, one ensures that even subsets of agents involved in the collection of data (so-called *coalitions*) see only DP data. In particular, they usually ensure that the curator sees only a perturbed version of the raw data.

While the decentralized nature and focus on time-series data considered in these works is relevant to

traffic monitoring, their focus is on algorithmic foundations and security guarantees in the context of DP. Prior knowledge of an adversary is not taken into account and, in some cases (e.g. smart metering) would be different from the use-case of traffic monitoring.

Privacy in Traffic Monitoring Many papers have tackled the problem of privacy in traffic monitoring from a solutions perspective (e.g. (Li et al., 2018; Qu et al., 2019; Jain et al., 2019; Sun et al., 2021; Gelderie. et al., 2024)). Those papers typically propose a specific privacy measure and analyze its security in the underlying model. For example, works with a focus on differential privacy (Gelderie. et al., 2024; Sun et al., 2021) prove that the definition of DP is satisfied. But by itself, DP says very little about the situation when an attacker has prior knowledge. Unfortunately, in traffic monitoring it is particularly easy for the attacker to acquire at least partial knowledge about the probabilities of certain traffic patterns (e.g. via products like Google Maps or because many cities publish such data). We must assume that the adversary has intimate knowledge of the overall traffic movement patterns and their probability distribution.

In this vein, there are efforts quantifying privacy loss (Gao et al., 2019). Gao et. al. study the privacy of LPR in a somewhat similar setting to ours. Their study is of an empirical nature and is based on a large LPR dataset. We seek to augment these results by performing a systematic mathematical analysis in the framework of information theory.

A related line of research is concerned with the privacy of vehicle data, such as trajectory information (Ma et al., 2019; Zhou et al., 2018). In these works, the vehicles themselves actively participate in the data gathering. This means that those architectures have a wealth of privacy enhancing options at their disposal that cannot be easily ported to LPR settings.

In the larger context of *smart cities*, even more related works on privacy exist (e.g. (Husnoo et al., 2021; Hassan et al., 2019; Yao et al., 2023; Kumar et al., 2022; Qu et al., 2019; Gracias et al., 2023; Jain et al., 2019)). The field is very diverse, but we are not aware of any works that investigate the effect of prior knowledge on the privacy afforded by license plate obfuscation or similar use-cases.

2 SETTING: ANONYMIZED TRAFFIC MONITORING

As mentioned, we study traffic monitoring via license plate recognition. It is assumed that each vehicle is

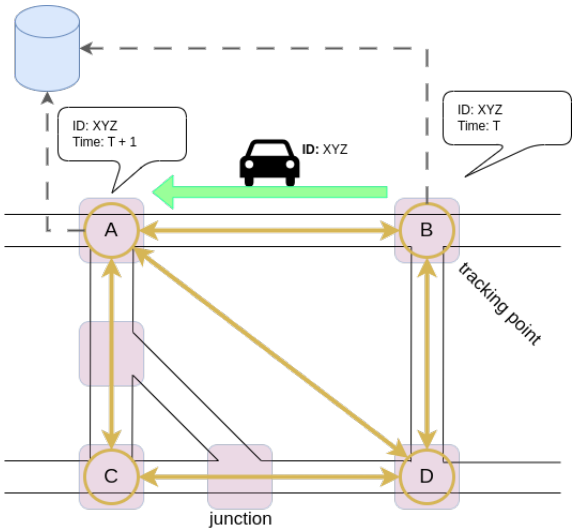

Figure 1: Architecture Overview

equipped with a unique and readable *license plate* l from some finite set $\mathcal{L}$ that is recorded at certain points in the city. We call such locations *tracking points (TPs)*. At each TP, a camera, or group of cameras, will record the license plate of every vehicle that passes by. This is depicted in fig. 1.

Tracking points (depicted as circles in fig. 1) are physical locations in the city at which data about current traffic is gathered. They reside alongside roads, usually junctions (depicted as gray boxes). TPs report the recorded vehicles to a central *statistics server* (depicted in blue). This server then builds a central statistic about the number of vehicles per route form this data.

Tracking points form a graph structure that we call the *city graph*. If one assumes that every junction on the city is a tracking point, the city graph is a faithful representation of the street map in terms of the roads and how they are connected. The relative distances between junctures are not reflected in this representation. There is an obvious trade-off between the number of TPs and the accuracy with which the city graph models the street map, but we do not explore that trade-off in this paper.

Definition 1 (City Graph)**.** The set V of all tracking points forms a graph $\mathcal{G} = (V, E)$, where $E \subseteq V^2$ is the directed edge relation defined by the rule that $(v, v') \in E$ whenever v' can be reached from v without visiting a tracking point in between. We call $\mathcal{G}$ the *city graph*.

Let $\mathcal{G} = (V, E)$ be the city graph as described above. We define the set $\mathcal{R}$ of *routes* as

$$\mathcal{R} = \{v_1 v_2 \cdots v_l \mid (v_i, v_{i+1}) \in E,\ 1 \leq i < l, l \in \mathbb{N}\}$$

Note that if every junction is a tracking point, there is exactly one edge per road and direction. In

this case $\mathcal{G}$ is an exact model of the city street map from a graph theoretic point of view (of course, distances and other non-graph theoretic properties are not represented in this model).

When a vehicle passes through a tracking point, its license plate is recorded. The TP then transmits a report to a central *statistics server* for every vehicle that it records. The server can compute the target statistic $S\colon \mathcal{R} \to \mathbb{N}_0$, which simply counts cars on a given route, from these reports. The statistic S is implicitly a function of time: $S(r)$ can vary over time for each $r \in \mathcal{R}$. In this paper we will study snapshots of S at some unspecified, yet fixed point in time t. Notably, all probability distributions that are studied depend on t. The fact that we consider snapshots is no limitation: The attacker's advantage is then simply the maximum over her per-time-step advantages.

The information about the traffic statistic at previous times is relevant to an attacker: The victim might be known to normally drive route r_1 during rush hours, but divert to r_2 if a certain intersection is congested. In that situation, knowing whether or not said intersection was congested 20 minutes ago has an impact on the adversaries knowledge on the situation right now. However, the analysis conducted in this paper extends to situations where an attacker has knowledge about the last T time-steps for some fixed T. It merely affects the probability of vehicles colliding on IDs (see below).

The reports to the central server are of interest. If the reports of tracking points to the server include licence plate information, they expose sensitive data to a central location. This is undesirable from both a privacy as well as a security perspective. As a result, one usually anonymizes this data in some way.

A typical pseudo-anonymization strategy ((swa, 2024; Gao et al., 2019)) is to replace the license plate with some random identifier. Here, TPs transmits an opaque ID to the server instead of the license plate. As long as the same vehicle is always assigned the same opaque ID, the server can compute the same statistic from this data. There are various ways in which such a mapping can be implemented (for example (swa, 2024) uses a hash function). We do not consider those approaches in this paper. Instead, we assume that there exists a binding of IDs, drawn from a set $\mathcal{U}$ to license plates.

The mapping of license plates to IDs defined in this way is of a temporary nature: Each time an ID is assigned to a license plate, this binding has a time to live (TTL). The TTL determines for how many hops a license plate is tracked. It is always initialized to constant value $T \in \mathbb{N}$ and decremented for every vertex on the path that the given vehicle visits[1]. In particular, the routes $r \in \mathcal{R}$ that will be recorded have length at most T. Because of this, we will (abusing notation) assume that $\mathcal{R}$ consists only of sequences of length at most T and is, in particular, finite.

In this paper, we gloss over the fact that some vehicles may not observed correctly: Their license plate may be missing or otherwise unreadable, or the system might fail to recognize them as vehicles due to some unknown error. Note also that we do not consider traffic participants that cannot be identified via a license plate (cyclist and pedestrians).

The anonymization provided by the use of opaque IDs hinges on the assumption that the knowledge which license plate corresponds to a given opaque ID is not known to an adversary. As mentioned before, there are numerous ways in which such an assignment could be implemented. If an approach involves cryptography, for instance, the resulting security notion is usually defined in terms of computational indistinguishability. By contrast, we define the security of the system as follows:

Definition 2. A binding $B\colon \mathcal{L} \to \mathcal{U}$ is *secure*, if $B(l)$ is a uniform random variable for each $l \in \mathcal{L}$ and $B(l)$ is independent of $B(l')$ for all $l \neq l'$.

Note that our security notion does not take the challenges in implementing such a binding into account. For example, consider a system in which an identifier $U_l \in \mathcal{U}$ is chosen uniformly at random for each $l \in \mathcal{L}$. To ensure that all TPs report the same value U_l for l, this binding needs to be shared between TPs. This leads to a number of significant cryptographic properties that the system must ensure, such as (forward) secrecy and post-compromise security. The way these systems ensure those properties is another important aspect of the security of the system. However, in this paper we focus purely on the stochastic properties of the binding itself and leave an investigation of the various technical options and security notions of secure bindings to future work.

Example 1. As an example, one might choose random UUIDs (specifically variant 2 UUIDs in version 4; see also (Leach et al., 2005)). Such a UUID contains 122 random bits. If N UUIDs are chosen uniformly at random, this constitutes a secure binding.

Naturally, there may be collisions, though they are rare: It is well-known (e.g. (Katz and Lindell, 2020)) that the probability of a collision is at most

$$p(N) \leq \frac{N(N-1)}{2^{122}} \leq \frac{N^2}{2^{122}}$$

[1] In practice, one will likely add another deprecation mechanism based on elapsed time to account for situations where a vehicle stops before the TTL expires.

So we have, say, $p(N) \leq 10^{-6}$, whenever at most $N \leq 2^{51} \approx 10^{15}$ distinct vehicles traverse the city. Since we can ignore collisions that do not occur simultaneously (i.e. where the two colliding occurrences of a UUID overlap in lifetime), we can lower the number N of vehicles further to the maximum number of vehicles that traverse the city within any time-period.

3 SCENARIO 1: OBFUSCATED TRAFFIC MONITORING

We now state formal privacy guarantees that are afforded by TMS using randomized bindings as outlined in section 2. Our main statement in theorem 2 is reminiscent of the well-known notions from cryptography (see e.g. (Katz and Lindell, 2020)). Of note is the role the expected value of cars per roads plays in the privacy that the system affords.

In this section, we assume opaque identifiers are chosen from the set $\mathcal{U} = \mathbb{B}^\lambda$, where $\lambda \in \mathbb{N}$ is a security parameter. We assume *secure* bindings, as defined in definition 2. However, we first study an idealized setting, where we assume that collisions on IDs do not occur: $U_l \neq U_{l'}$ for all $l \neq l'$. Such a binding is neither practical, not secure in the sense of definition 2 (the variables U_l are clearly not independent). We subsequently lift the result to the general case in theorem 2.

For notational convenience, we write $[\cdot]$ for events defined by some function on the output of one or more random variables. For example, if $X_1, \ldots, X_n$ are random variables with identical range, we may write $[\exists i \neq j \colon X_i \neq X_j]$ for the event that not all X_i are equal.

Formally, let $(\Omega, \Pr)$ and $(\Omega_i, \Pr_i)_{i=1}^n$ be probability spaces. Let $X_i \colon \Omega \to \Omega_i$ be a random variables for $1 \leq i \leq n$ and Φ a predicate on $\Omega_1 \times \cdots \times \Omega_n$. We write $[\Phi(X_1, \ldots, X_n)] = \{\omega \in \Omega \mid \Phi(X_1(\omega), \ldots, X_n(\omega))\} \subseteq \Omega$ for the event that Φ is true.

As mentioned, we first study an idealized setting: We assume that IDs are assigned via an injective function $f \colon \mathcal{L} \to \mathcal{U}$ that is chosen uniformly at random. Write $\mathcal{F} = \{f \colon \mathcal{U} \to \mathcal{L} \mid f \text{ injective}\}$. We choose $f^* \leftarrow \mathcal{F}$ uniformly at random and assign to each $l \in \mathcal{L}$ the ID $f^*(l)$. For notational convenience, we use a random variable $U_l = f^*(l)$ for $l \in \mathcal{L}$.

We use some additional random variables for notational convenience. Let $l \in \mathcal{L}$ and write R_l for the random variable that assigns a route to l. If l does not currently drive at all, then R_l and U_l take the special value $\perp \notin \mathcal{U} \cup \mathcal{L}$. Finally, we write $D = \{l \in \mathcal{L} \mid U_l \neq \perp\}$ for the random variable producing the set of cars that drive.

Remark 1. Note that even if l does not drive, the value $f^*(l)$ is defined. However, we deliberately choose to not define $U_l = f^*(l)$ in the case where l does not drive. This is because in the model discussed in section 2 IDs are only assigned to cars that drive. In that situation, U_l and R_l *cannot be independent*. They are only conditionally independent on the events "l drives"$= [l \in D]$ or "l does not drive"$= [l \notin D]$.

So, with a view to the more general situation of theorem 2, we chose to adhere to this behavior even though it would be possible to work around it in our simplified setting.

It is sometimes useful to associate a route $r \in \mathcal{R}$ with an ID $u \in \mathcal{U}$. For $u \in \mathcal{U}$, we write $R_u = \{w \in V^* \mid w \text{ consistent with } u\}$ for the random variable that denotes the set of all walks in $\mathcal{G}$ that can be associated with u. If u is not currently assigned to any license plate, then $R_u = \emptyset$. If there is a collision on u, then R_u *may* contain more than one route (or even sequences of vertices that are not valid walks in $\mathcal{G}$, though this turns out to be irrelevant for the purposes of this analysis). Since we assume, for the moment, that IDs are assigned using an injective function f^*, R_u is either empty, or a singleton containing one valid route from $\mathcal{R} \subsetneq V^*$. We therefore write $R_u = r$ whenever it is ensured that R_u is a singleton.

For $r \in \mathcal{R}$, write $N_r = |\{l \in \mathcal{L} \mid R_l = r\}|$ for the random variable denoting the number of vehicles on route r. Note that while $f^* \leftarrow \mathcal{F}$ is uniform by assumption, we make *no* assumption about the distribution of R_l and U_l. However, $\Pr[U_l = u \mid l \in D] = \Pr[f^*(l) = u] = |\mathcal{U}|^{-1} = 2^{-\lambda}$.

Lemma 1. *For every $r \in \mathcal{R}$, $u \in \mathcal{U}$ and every $l \in \mathcal{L}$ with $\Pr[R_u = r] \neq 0$ it holds that*

$$\Pr[U_l = u \mid R_u = r] = \frac{\Pr[R_l = r]}{\mathbb{E}[N_r]}$$

where the probabilities are taken over the uniform choice of $f^* \leftarrow \mathcal{F}$ and over the randomness of R_l.

Proof. Observe that if $\Pr[U_l = u] = 0$, then $\Pr[U_l = \perp] = 1$ and $\Pr[R_l = r] = 0$: *If* a ID is chosen, it is chosen uniformly; so if $\Pr[U_l = u] = 0$ for one $u \in \mathcal{U}$ can only mean that the license plate in question does not drive at all. In this case, both sides if the equation are equal to zero. We may therefore assume $\Pr[U_l = u] \neq 0$.

If $\Pr[U_l = u] \neq 0$, we have:

$$\Pr[U_l = u \mid R_u = r] = \frac{\Pr[R_u = r \mid U_l = u]}{\Pr[R_u = r]} \cdot \Pr[U_l = u]$$

Note that $\Pr[R_u = r \mid U_l = u] = \Pr[R_l = r \mid U_l = u]$. This is because there are no collisions on IDs by construction (f^* is injective). Thus:

$$\Pr[U_l = u \mid R_u = r] = \frac{\Pr[R_l = r, U_l = u]}{\Pr[R_u = r]}$$

As noted before, U_l and R_l are not independent: $\Pr[R_l = r, U_l = \bot] = 0$ for all $r \in \mathcal{R}$, and $l \in \mathcal{L}$, even though we may have both $\Pr[U_l = \bot] \neq 0$ and $\Pr[R_l = r] \neq 0$ for appropriate choices of l and r. However, conditioned on the event $[l \in D]$, the two variables *are* independent:

$$\begin{aligned}&\Pr[U_l = u, R_l = r \mid l \in D]\\ &\qquad = \Pr[U_l = u \mid l \in D] \cdot \Pr[R_l = r \mid l \in D]\end{aligned}$$

This is because the route l takes and the ID it is assigned stem from independent random sources. Indeed, as noted above, $\Pr[U_l = u \mid l \in D] = |\mathcal{U}|^{-1}$.

Since moreover the event $[l \in D]$ is a superset of the event $[U_l = u \in \mathcal{U}]$ (whereby $u \neq \bot$) and is likewise a superset of the event $[R_l = r \in \mathcal{R}]$ (and thus $r \neq \bot$), this gives for all $u \in \mathcal{U}$ and all $r \in \mathcal{R}$:

$$\begin{aligned}&\Pr[U_l = u, R_l = r]\\ =&\Pr[U_l = u, R_l = r \mid l \in D] \cdot \Pr[l \in D]\\ =&\Pr[U_l = u \mid l \in D] \cdot \Pr[R_l = r \mid l \in D] \cdot \Pr[l \in D]\\ =&|\mathcal{U}|^{-1} \cdot \Pr[R_l = r]\end{aligned}$$

Taken together, we get:

$$\Pr[U_l = u \mid R_u = r] = \frac{\Pr[R_l = r] \cdot |\mathcal{U}|^{-1}}{\Pr[R_u = r]}$$

We consider the denominator next:

$$\Pr[R_u = r] = \sum_{l \in \mathcal{L}} \Pr[U_l = u, R_l = r]$$

since the events on the right side are all disjoint (again because there are no collisions on IDs). And by our observation about independence:

$$\Pr[R_u = r] = |\mathcal{U}|^{-1} \cdot \underbrace{\sum_{l \in \mathcal{L}} \Pr[R_l = r]}_{\chi_r}$$

and so it suffices to show that $\chi_r = \mathbb{E}[N_r]$.

Denote by $\mathbb{1}_{r,l}$ the indicator random variable for the event $[R_l = r]$. Then $\mathbb{E}[\mathbb{1}_{r,l}] = \Pr[R_l = l]$ and by linearity of expectation:

$$\begin{aligned}\chi_r = \sum_{l \in \mathcal{L}} \Pr[R_l = r] &= \sum_{l \in \mathcal{L}} \mathbb{E}[\mathbb{1}_{r,l}]\\ &= \mathbb{E}\left[\sum_{l \in \mathcal{L}} \mathbb{1}_{r,l}\right] = \mathbb{E}[N_r]\end{aligned}$$

□

The assumption that f^* is chosen uniformly from the set $\mathcal{F}$ of all injective functions from $\mathcal{L}$ to $\mathcal{U}$ is, of course, impractical. It is more natural to chose on $u \in \mathcal{U}$ per $l \in \mathcal{L}$ uniformly at random (as is done whenever we work with Variant 2, Version 4 UUIDs, for example; see (Leach et al., 2005)).

If we draw IDs from the set $\mathcal{U} = \mathbb{B}^\lambda$ uniformly at random, we may deal with collisions. This case can be dealt with in the usual way using well-known Birthday Paradox probability bounds. This is the content of the theorem below.

In the following, we recall that $R_u \subseteq V^*$ is defined to be the set of all sequences of vertices that are consistent with $u \in \mathcal{U}$. This set can now contain more than one element.

Theorem 2. *Let $\lambda \in \mathbb{N}$. For every $r \in \mathcal{R}$, $u \in \mathcal{U}$ and every $l \in \mathcal{L}$ with $\Pr[r \in R_u] \neq 0$ it holds that*

$$\left|\Pr[U_l = u \mid r \in R_u] - \frac{\Pr[R_l = r]}{\mathbb{E}[N_r]}\right| \leq \frac{|\mathcal{L}|^2}{2^\lambda}$$

where the probabilities are taken over the uniform choices of $U_l = u$ and over the assignment of routes (if any) to $l \in \mathcal{L}$.

Proof. We condition on the complementary events NoColl and Coll defined as $\mathsf{NoColl} = \overline{\mathsf{Coll}}$ and $\mathsf{Coll} = [\exists l, k \in D\colon U_l = U_k \wedge l \neq k]$.

We have:

$$\begin{aligned}&\Pr[U_l = u \mid r \in R_u]\\ =&\Pr[U_l = u \mid r \in R_u, \mathsf{NoColl}] \cdot \Pr[\mathsf{NoColl} \mid r \in R_u]\\ &\quad + \Pr[U_l = u \mid r \in R_u, \mathsf{Coll}] \cdot \Pr[\mathsf{Coll} \mid r \in R_u]\\ =&\Pr[U_l = u \mid r \in R_u, \mathsf{NoColl}] \qquad (*)\\ &\quad + \Pr[\mathsf{Coll} \mid r \in R_u] \cdot (\Pr[U_l = u \mid r \in R_u, \mathsf{Coll}]\\ &\quad - \Pr[U_l = r \mid r \in R_u, \mathsf{NoColl}])\end{aligned}$$

For the term $(*)$, we note that the condition NoColl puts us into the situation of previous theorem. We can apply the corresponding proof and obtain:

$$\Pr[U_l = u \mid r \in R_u, \mathsf{NoColl}] = \frac{\Pr[R_l = r \mid \mathsf{NoColl}]}{\mathbb{E}[N_r \mid \mathsf{NoColl}]}$$

But since NoColl and $[R_l = r]$ are independent for all $r \in \mathcal{R}$ and $l \in \mathcal{L}$ (the assignment of IDs is done uniformly at random independently of what route a car drives, if it drives), this simplifies to the familiar

$$\Pr[R_l = r] \cdot \frac{1}{\mathbb{E}[N_r]}$$

It suffices to show that the remaining term falls in the interval $[-|\mathcal{L}|^2 \cdot 2^{-\lambda}, |\mathcal{L}|^2 \cdot 2^{-\lambda}]$. Since $(p - p') \cdot q \in [-q, q]$ for any probabilities p, p', q, it suffices to show $\Pr[\mathsf{Coll} \mid r \in R_u] \leq |\mathcal{L}|^2 \cdot 2^{-\lambda}$.

Note that

$$
\begin{aligned}
&\Pr[\mathsf{NoColl} \mid r \in R_u] \\
&= \sum_{k=0}^{|\mathcal{L}|} \Pr[\mathsf{NoColl} \mid r \in R_u, |D| = k] \cdot \Pr[|D| = k \mid R_u] \\
&= \sum_{k=0}^{|\mathcal{L}|} \prod_{i=1}^{k-1} \frac{2^\lambda - i}{2^\lambda} \cdot \Pr[|D| = k \mid r \in R_u] \qquad (*) \\
&\geq \sum_{k=0}^{|\mathcal{L}|} \prod_{i=0}^{|\mathcal{L}|-1} \frac{2^\lambda - i}{2^\lambda} \cdot \Pr[|D| = k \mid r \in R_u] \\
&= \prod_{i=0}^{|\mathcal{L}|-1} \frac{2^\lambda - i}{2^\lambda} \cdot \sum_{k=0}^{|\mathcal{L}|} \Pr[|D| = k \mid r \in R_u] \\
&= \prod_{i=0}^{|\mathcal{L}|-1} \frac{2^\lambda - i}{2^\lambda}
\end{aligned}
$$

where $(*)$ follows because $k-1$ IDs are assigned collision free an uniformly from the set $\mathcal{U} \setminus \{u\}$. We get:

$$
\Pr[\mathsf{Coll} \mid r \in R_u] = 1 - \Pr[\mathsf{NoColl} \mid r \in R_u] \leq 1 - \prod_{i=0}^{|\mathcal{L}|-1} \frac{2^\lambda - i}{2^\lambda} \leq \frac{|\mathcal{L}| \cdot (|\mathcal{L}| - 1)}{2^\lambda}
$$

where the last inequality is well known (see e.g. discussion on the Birthday Paradox in (Katz and Lindell, 2020)). □

Remark 2. This result formalizes the intuitive notion that driving anonymously along a very popular route does not leak much information about the vehicle l. Conversely, if the route is unpopular, then the system leaks information consistent with the probability of vehicle l driving that route.

In particular, we capture intuitively obvious observations, such as: If vehicle l is parked at a remote location every night, then the expected number of cars on all routes leading to that location is close to 1. In this event the driver can be de-anonymized: $\Pr[R_{l'} = r] \approx 0$ for all $l' \neq l$.

4 SCENARIO 2: OBFUSCATION & DIFFERENTIAL PRIVACY

It is well-known (see e.g. (Dwork et al., 2014)) that the security guarantees afforded by DP make no statement about the knowledge an adversary might draw from *prior knowledge*. If a traffic monitoring system provides (ε, δ)-DP, this just says that any two possible adjacent inputs produce the same output with almost equal probability. But of course, those two adjacent inputs could reasonably have very different prior probabilities, meaning that an attacker can infer significantly more information about the nature of the input from observing the output than one might reasonably expect from the notion of DP.

We recall the definition of differential privacy (Dwork et al., 2006; Dwork, 2006). Note that *adjacency* has not yet been defined; we give a use-case specific definition below in definition 4.

Definition 3. Let $\varepsilon, \delta > 0$. Let $\mathcal{M}$ be a randomized algorithm. Then $\mathcal{M}$ is said to have (ε, δ)-differential privacy, if for all adjacent inputs $x, x' \in \mathsf{dom}(\mathcal{M})$ and all subsets $S \subseteq \mathsf{range}(\mathcal{M})$ it holds that

$$\Pr[\mathcal{M}(x) \in S] \leq \exp(\varepsilon) \cdot \Pr[\mathcal{M}(x') \in S] + \delta$$

The algorithms $\mathcal{A}$ is referred to as the *curator*.

In this section, we investigate the effect that an (ε, δ)-DP curator $\mathcal{M}$ has on the conclusions an adversary can rationally draw from the observed outputs. Recall that we assign routes to vehicles via the random variable $R_l \in \mathcal{R} \uplus \{\perp\}$ for each $l \in \mathcal{L}$, where $R_l = \perp$ means that the vehicle with license plate l is not driving at all.

First, we introduce some notation. In what follows, it is convenient to interpret $\perp$ as just another route. We define $\hat{\mathcal{R}} = \mathcal{R} \uplus \{\perp\}$. We denote by $X = (R_l)_{l \in \mathcal{L}}$ the random vector that assigns a route (or $\perp$) to every vehicle. Clearly X is a complete representation of the input to a curator in the central DP model (cf. section 1.1). We recognize values of x of X as functions $x \colon \mathcal{L} \to \hat{\mathcal{R}}$. Write $\mathfrak{X} = \hat{\mathcal{R}}^{\mathcal{L}}$ for the set of all possible values X might take on. Note that since $\mathcal{L}$ and $\mathcal{R}$ are finite (the latter because of the length-bound enforced by the TTL), the set $\mathfrak{X}$ is also finite. In what follows, let $p \colon \mathfrak{X} \to [0,1]$ the distribution of X.

To reason about DP, we need to clarify adjacency in our context:

Definition 4 (l-Adjacency)**.** Let $l \in \mathcal{L}$ and let $x, x' \in \mathfrak{X}$. We call x and x' *a-adjacent* (or simply *adjacent*), if $x(s) = x'(s)$ for all $s \neq l$, and $x'(l) \neq x(l)$. We write $x \bowtie_l x'$.

Note that our definition of adjacency is such that no x is adjacent to itself: $x \not\bowtie_l x$ for all $x \in \mathfrak{X}$ and $l \in \mathcal{L}$. This technicality will simplify the notation in some of the proofs below. It has no other significance and, in particular, the main theorem of this paper holds even if we allow self-adjacency.

We consider all $x \in \mathfrak{X}$ that map l to some particular route $r \in \hat{\mathcal{R}}$: $A_{l,r} = \{x \in \mathfrak{X} \mid x(l) = r\}$. If l is clear from context, we write A_r. An *l-surrounding* is an element $x \in \mathfrak{X}_{-l} = \hat{\mathcal{R}}^{\mathcal{L} \setminus \{l\}}$. Surroundings give a complete description of the movements of every participant except for l. Write $x[l/r]$ for the function defined by $x[l/r](t) = x(t)$ for all $t \neq l$ and $x[l/r](l) = r$.

Note that this definition is applicable to both $x \in \mathfrak{X}$ and $x \in \mathfrak{X}_{-l}$.

Observe that two distinct elements $x \neq x'$ are l-adjacent iff they have the same l-surrounding. Note $A_{l,r} = \{x[l/r] \mid x \in \mathfrak{X}_{-l}\}$. If $x \in \mathfrak{X}$, we write x_{-l} for the corresponding l-surrounding.

Recall that the *information content of x* is defined as $I(x) \stackrel{\text{def}}{=} -\log(p(x))$. We can now define:

Definition 5 (Preference Gap)**.** Let $l \in \mathcal{L}$ and $r \in \hat{\mathcal{R}}$. The quantity

$$\sigma(l,r) = \sup_{x \in A_r} \sup_{x' \bowtie_l x} I(x') - I(x)$$

is called the *preference gap of l and r*. The quantity

$$\sigma(l) = \sup_{r \in \hat{\mathcal{R}}} \sigma(l,r)$$

is called the *preference gap of l*.

The preference gap measures the greatest difference of the information contents of two l-adjacent values.

Proposition 3. *Let $l \in \mathcal{L}$ and $r \in \hat{\mathcal{R}}$.*

1. *By definition, we have:*
$$\sigma(l,r) = \sup_{x \in A_{l,r}} \sup_{x' \bowtie_l x} \log\left(\frac{p(x)}{p(x')}\right)$$
2. *As a consequence of the previous item, we have for any $x \in A_{l,r}$ and all $x' \bowtie_l x$ with $p(x) \neq 0$ and $p(x') \neq 0$ that*
$$\frac{p(x)}{p(x')} \leq \exp(\sigma(l,r)) \leq \exp(\sigma(l))$$
3. *$\sigma(l,r) = \infty$ if and only if $\Pr[R_l = r] = 1$, and $\sigma(l,r) = -\infty$ if and only if $\Pr[R_l = r] = 0$.*
4. *In particular, $\sigma(l) \neq -\infty$ for all $l \in \mathcal{L}$.*

With the convention that $\exp(\infty) = \infty$, we have:

Theorem 4. *Let $\mathcal{M}$ be any (ε,δ)-DP curator defined on $\mathfrak{X}$.*

For every $l \in \mathcal{L}$, $r \in \hat{\mathcal{R}}$ and $y \in \mathsf{range}(\mathcal{M})$ with $\Pr[Y = y] > 0$ it holds that:

$$\Pr[R_l = r \mid Y = y] \leq \left(\exp(\varepsilon)\Pr[R_l \neq r \mid Y = y] + \frac{\delta}{\Pr[Y = y]}\right) \cdot \exp(\sigma(l))$$

Let $n_r = |\mathcal{R}|$. If $\Pr[R_l = r] \neq 0$, then moreover:

$$\Pr[R_l \neq r \mid Y = y] \leq n_r \cdot \left(\exp(\varepsilon)\Pr[R_l = r \mid Y = y] + \frac{\delta}{\Pr[Y = y]}\right) \cdot \exp(\sigma(l))$$

Proof. Note that if $\sigma(l) = \infty$, the statement is trivially true. We may therefore assume $\sigma(l) \in \mathbb{R}$.

Pick an arbitrary $f\colon A_r \to \overline{A_r}$ with the property that $x \bowtie_l f(x)$ for all $x \in A_r$. For any $x \in \mathfrak{X}$ and $y \in \mathsf{range}(\mathcal{M})$ write $q(x,y) = \Pr[\mathcal{M}(x) = y]$. Now note that

$$\begin{aligned}
&\Pr[Y = y, R_l = r] = \sum_{x \in A_r} p(x) \cdot q(x,y) \\
&\leq \exp(\sigma(l)) \sum_{x \in A_r} p(f(x)) \cdot \big(\\
&\qquad \exp(\varepsilon) \cdot q(f(x),y) + \delta\big) \\
&\leq \exp(\sigma(l)) \cdot \Bigg(\exp(\varepsilon) \sum_{x \in \overline{A_r}} p(x) \cdot q(x,y) \\
&\qquad + \delta \cdot \Pr[R_r \neq r] \Bigg) \\
&\leq \exp(\sigma(l)) \cdot (\exp(\varepsilon)\Pr[Y = y, R_l \neq r] + \delta)
\end{aligned}$$

For the second claim, let $\Pr[R_l = r] \neq 0$. Then

$$\begin{aligned}
&\Pr[Y = y, R_l \neq r] = \sum_{x \notin A_r} p(x) \cdot q(x,y) \\
&= \sum_{x \in \mathfrak{X}_{-l}} \sum_{r' \neq r} p(x[l/r']) \cdot q(x[l/r'],y) \\
&\leq \sum_{x \in \mathfrak{X}_{-l}} \sum_{r' \neq r} p(x[l/r]) \cdot \exp(\sigma(l)) \cdot q(x[l/r'],y) \\
&\leq \sum_{x \in \mathfrak{X}_{-l}} \exp(\sigma(l)) \sum_{r' \neq r} \big(\exp(\varepsilon) \\
&\qquad \cdot p(x[l/r]) \cdot q(x[l/r],y) + \delta\big) \\
&= \exp(\sigma(l)) n_r (\exp(\varepsilon)\Pr[R_l = r, Y = y] + \delta)
\end{aligned}$$

For the last equality, note that $|\hat{\mathcal{R}}| = n_r + 1$. □

5 CONCLUSION

We have considered traffic monitoring using anonymized license plates. In this context, we have studied the question, how prior knowledge about the overall probability distributions of the general driving behavior or individual participants affects the privacy guarantees of such traffic monitoring systems. We extended this study to systems that provide differential privacy.

When no DP is involved, the knowledge an adversary has about the probabilities of individual driving behavior can greatly increase the confidence in unmasking attacks, where an adversary tries to identify the individual behind a certain opaque ID. Specifically, if the route in question is very unpopular, the risk of unmasking is high.

We then studied how the guarantees of DP, stated in terms of neighboring data-sets, generalize to guarantees about the likelihood of a particular individual driving on a certain route. We found that the unevenness of the underlying probability distribution of traffic patterns can degrade the assurances made by the DP mechanism significantly.

An interesting open question for future work is to what extent the picture changes when the adversary has only *partial* knowledge of the probability distributions involved. For instance, an adversary may have information about the probability of a given number of vehicles per route at a given time, but not the probabilities of individual vehicles being on that route.

REFERENCES

(2024). Swarm analytics technical documentation — technical concept. https://docs.swarm-analytics.com/technical-documentation/use-cases/advanced-traffic-insights/traffic-counting/technical-concept. Accessed: 2024-04-10.

Baran, R., Ruść, T., and Rychlik, M. (2014). A smart camera for traffic surveillance. In *Multimedia Communications, Services and Security: 7th International Conference, MCSS 2014, Krakow, Poland, June 11-12, 2014. Proceedings 7*, pages 1–15. Springer.

Bhardwaj, V., Rasamsetti, Y., and Valsan, V. (2022). Traffic control system for smart city using image processing. *AI and IoT for Smart City applications*, pages 83–99.

Bisio, I., Garibotto, C., Haleem, H., Lavagetto, F., and Sciarrone, A. (2022). A systematic review of drone based road traffic monitoring system. *IEEE Access*, 10:101537–101555.

Biswas, S. P., Roy, P., Patra, N., Mukherjee, A., and Dey, N. (2016). Intelligent traffic monitoring system. In *Proceedings of the Second International Conference on Computer and Communication Technologies: IC3T 2015, Volume 2*, pages 535–545. Springer.

Bittau, A., Erlingsson, U., Maniatis, P., Mironov, I., Raghunathan, A., Lie, D., Rudominer, M., Kode, U., Tinnes, J., and Seefeld, B. (2017). Prochlo: Strong privacy for analytics in the crowd. In *Proceedings of the 26th Symposium on Operating Systems Principles*, SOSP '17, page 441–459, New York, NY, USA. Association for Computing Machinery.

Bonawitz, K., Ivanov, V., Kreuter, B., Marcedone, A., McMahan, H. B., Patel, S., Ramage, D., Segal, A., and Seth, K. (2017). Practical secure aggregation for privacy-preserving machine learning. In *proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1175–1191.

Chan, T.-H. H., Shi, E., and Song, D. (2011). Private and continual release of statistics. *ACM Transactions on Information and System Security (TISSEC)*, 14(3):1–24.

Chan, T.-H. H., Shi, E., Song, D., and Song, D. (2012). Optimal lower bound for differentially private multi-party aggregation. *Embedded Systems and Applications*.

Cheu, A., Smith, A., Ullman, J., Zeber, D., and Zhilyaev, M. (2019). Distributed differential privacy via shuffling. In *Advances in Cryptology–EUROCRYPT 2019: 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19–23, 2019, Proceedings, Part I 38*, pages 375–403. Springer.

Corrigan-Gibbs, H. and Boneh, D. (2017). Prio: Private, robust, and scalable computation of aggregate statistics. In *14th USENIX symposium on networked systems design and implementation (NSDI 17)*, pages 259–282.

Djahel, S., Doolan, R., Muntean, G.-M., and Murphy, J. (2015). A communications-oriented perspective on traffic management systems for smart cities: Challenges and innovative approaches. *IEEE Communications Surveys & Tutorials*, 17(1):125–151.

Du, R., Chen, C., Yang, B., Lu, N., Guan, X., and Shen, X. (2015). Effective urban traffic monitoring by vehicular sensor networks. *IEEE Transactions on Vehicular Technology*, 64(1):273–286.

Du, S., Ibrahim, M., Shehata, M., and Badawy, W. (2013). Automatic license plate recognition (alpr): A state-of-the-art review. *IEEE Transactions on Circuits and Systems for Video Technology*, 23(2):311–325.

Dwork, C. (2006). Differential privacy. In *International colloquium on automata, languages, and programming*, pages 1–12. Springer.

Dwork, C., McSherry, F., Nissim, K., and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. In *Theory of Cryptography*, pages 265–284. Springer.

Dwork, C., Naor, M., Pitassi, T., and Rothblum, G. N. (2010). Differential privacy under continual observation. In *Proceedings of the Forty-Second ACM Symposium on Theory of Computing*, STOC '10, page 715–724, New York, NY, USA. Association for Computing Machinery.

Dwork, C., Roth, A., et al. (2014). The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science*, 9(3–4):211–407.

Feng, Y., Hourdos, J., and Davis, G. A. (2014). Probe vehicle based real-time traffic monitoring on urban roadways. *Transportation Research Part C: Emerging Technologies*, 40:160–178.

Gade, D. (2019). Ict based smart traffic management system "ismart" for smart cities. *International Journal of Recent Technology and Engineering*, 8(3):1000–1006.

Gao, J., Sun, L., and Cai, M. (2019). Quantifying privacy vulnerability of individual mobility traces: A case study of license plate recognition data. *Transportation Research Part C: Emerging Technologies*, 104:78–94.

Gelderie., M., Luff., M., and Brodschelm., L. (2024). Differential privacy for distributed traffic monitoring in smart cities. In *Proceedings of the 10th International Conference on Information Systems Security and Privacy - ICISSP*, pages 758–765. INSTICC, SciTePress.

Gracias, J. S., Parnell, G. S., Specking, E., Pohl, E. A., and Buchanan, R. (2023). Smart cities—a structured literature review. *Smart Cities*, 6(4):1719–1743.

Hassan, M. U., Rehmani, M. H., and Chen, J. (2019). Differential privacy techniques for cyber physical systems: a survey. *IEEE Communications Surveys & Tutorials*, 22(1):746–789.

Husnoo, M. A., Anwar, A., Chakrabortty, R. K., Doss, R., and Ryan, M. J. (2021). Differential privacy for iot-enabled critical infrastructure: A comprehensive survey. *IEEE Access*, 9:153276–153304.

Jain, N. K., Saini, R., and Mittal, P. (2019). A review on traffic monitoring system techniques. *Soft computing: Theories and applications: Proceedings of SoCTA 2017*, pages 569–577.

Jain, P., Raskhodnikova, S., Sivakumar, S., and Smith, A. (2023). The price of differential privacy under continual observation. In *International Conference on Machine Learning*, pages 14654–14678. PMLR.

Katz, J. and Lindell, Y. (2020). *Introduction to Modern Cryptography*. Chapman & Hall/CRC Cryptography and Network Security Series. CRC Press, 6000 Broken Sound Parkway NW, Suite 300, Boca Raton, FL 33487-2742.

Kellaris, G., Papadopoulos, S., Xiao, X., and Papadias, D. (2014). Differentially private event sequences over infinite streams. *Proceedings of the VLDB Endowment*, 7(12):1155–1166.

Khanna, A., Goyal, R., Verma, M., and Joshi, D. (2019). Intelligent traffic management system for smart cities. In Singh, P. K., Paprzycki, M., Bhargava, B., Chhabra, J. K., Kaushal, N. C., and Kumar, Y., editors, *Futuristic Trends in Network and Communication Technologies*, pages 152–164, Singapore. Springer Singapore.

Krishnamoorthy, R. and Manickam, S. (2018). Automated traffic monitoring using image vision. In *2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT)*, pages 741–745.

Kumar, A., Upadhyay, A., Mishra, N., Nath, S., Yadav, K. R., and Sharma, G. (2022). Privacy and security concerns in edge computing-based smart cities. In *Robotics and AI for Cybersecurity and Critical Infrastructure in Smart Cities*, pages 89–110. Springer.

Leach, P., Mealling, M., and Salz, R. (2005). A universally unique identifier (uuid) urn namespace. Technical Report RFC 4122, Internet Engineering Taskforce.

Li, Y., Zhang, P., and Wang, Y. (2018). The location privacy protection of electric vehicles with differential privacy in v2g networks. *Energies*, 11(10):2625.

Ma, Z., Zhang, T., Liu, X., Li, X., and Ren, K. (2019). Real-time privacy-preserving data release over vehicle trajectory. *IEEE transactions on vehicular technology*, 68(8):8091–8102.

Qu, Y., Nosouhi, M. R., Cui, L., and Yu, S. (2019). Privacy preservation in smart cities. In *Smart cities cybersecurity and privacy*, pages 75–88. Elsevier.

Rana, S., Rana, S., Garg, U., Garg, U., Gupta, N., and Gupta, N. (2021). Intelligent traffic monitoring system based on internet of things. *2021 International Conference on Computational Performance Evaluation (ComPE)*.

Rizwan, P., Suresh, K., and Babu, M. R. (2016). Real-time smart traffic management system for smart cities by using internet of things and big data. In *2016 International Conference on Emerging Technological Trends (ICETT)*, pages 1–7.

Shi, E., Chan, T.-H. H., Rieffel, E., Chow, R., and Song, D. (2011). Privacy-preserving aggregation of time-series data. In *Proceedings of the 18th Annual Network & Distributed System Security Symposium (NDSS'11)*, volume 2.4. Internet Society.

Sun, Y.-E., Huang, H., Yang, W., Chen, S., and Du, Y. (2021). Toward differential privacy for traffic measurement in vehicular cyber-physical systems. *IEEE Transactions on Industrial Informatics*, 18(6):4078–4087.

Yao, A., Li, G., Li, X., Jiang, F., Xu, J., and Liu, X. (2023). Differential privacy in edge computing-based smart city applications: Security issues, solutions and future directions. *Array*, page 100293.

Zhou, Z., Qiao, Y., Zhu, L., Guan, J., Liu, Y., and Xu, C. (2018). Differential privacy-guaranteed trajectory community identification over vehicle ad-hoc networks. *Internet Technology Letters*, 1(3):e9.

ACKNOWLEDGEMENTS

This work was done as part of a project funded by the German Federal Office for Information Security under project funding reference number 01MO23006.