ON PRIVACY AND CONFIDENTIALITY OF COMMUNICATIONS IN ORGANIZATIONAL GRAPHS

Masoumeh Shafieinejad
University of Waterloo
Waterloo, Canada
masoumeh@uwaterloo.ca

Huseyin Inan
Microsoft Research
Redmond, United States
Huseyin.Inan@microsoft.com

Marcello Hasegawa
Microsoft
Redmond, United States
marcellh@microsoft.com

Robert Sim
Microsoft Research
Redmond, United States
rsim@microsoft.com

ABSTRACT

Machine learned models trained on organizational communication data, such as emails in an enterprise, carry unique risks of breaching confidentiality, even if the model is intended only for internal use. This work shows how confidentiality is distinct from privacy in an enterprise context, and aims to formulate an approach to preserving confidentiality while leveraging principles from differential privacy (DP). Works that apply DP techniques to natural language processing tasks usually assume independently distributed data, and overlook potential correlation among the records. Ignoring this correlation results in a fictional promise of privacy while, conversely, extending DP techniques to include group privacy is over-cautious and severely impacts model utility. We introduce a middle-ground solution, proposing a model that captures the correlation in the social network graph, and incorporates this correlation in the privacy calculations through Pufferfish privacy principles.

1 INTRODUCTION

A number of applications in natural language understanding rely on language models (Chen et al., 2019; Adam et al., 2020). To enable such models it is necessary to process training data that best represents the target application. Such tasks become especially sensitive in the setting of organizational communication, where organizations, individuals, or communities may share secret data, and preserving confidentiality is of utmost importance. Organizational communication often presents a complex underlying structure of interactions which is well modeled by a social graph. Data privacy in graphs have been addressed by a number of previous works (Yuan et al., 2010; Mittal et al., 2012; Cheng et al., 2010; Kearns et al., 2016; Gao et al., 2017; Zhu et al., 2017; Karwa et al., 2014), where most of the works based on differentially private approaches model individuals as nodes and exchanged messages as edges. These proposed methods provide either node level privacy guarantees (Kasiviswanathan et al., 2013; Blocki et al., 2013; Chen & Zhou, 2013; Raskhodnikova & Smith, 2016; Chen et al., 2014; Ghosh & Kleinberg, 2018), or edge level privacy guarantees (Nissim et al., 2007; Sala et al., 2011; Karwa et al., 2014; Hay et al., 2009). In the context of organizational communication we define node level guarantees as individual privacy and edge level guarantees as confidentiality. That is, confidentiality involves protecting information that is shared between two or more individuals in the organization. In this work, we set aside questions of individual privacy and examine problems in ensuring confidentiality in organizational communication.

In considering the case of social graphs, often the properties of an edge can be inferred from the properties of other nearby edges. Liu et al. (2016); Zheng et al. (2018) show that the dependency between instances affects the robustness of differential privacy guarantees. One approach to address this problem is via group differential privacy, which assumes that all edges in a group of participants are fully correlated, and queries on the graph must be invariant to the presence/absence of the entire

group. This can significantly impact the accuracy of the query by being over-protective of edge-edge relationships (Chen et al., 2014; Ghosh & Kleinberg, 2018).

We address the issues presented by edge correlation by employing a generalized version of differential privacy called Pufferfish (Kifer & Machanavajjhala, 2014; Song et al., 2017). In Pufferfish a set $\mathcal{S}_{\text{pairs}} \subseteq \mathcal{S} \times \mathcal{S}$ of complementary secret pairs is defined, and privacy is provided by ensuring the secret pairs are indistinguishable for any data distribution θ (capturing the correlation among the records) known to the adversary. Through this requirement the correlation problem can be addressed while allowing utility to be preserved. In addition, the non-independence between edges provides us with a simple model for confidentiality such that information passing between neighboring edges is more likely to be confidential than information that is randomly distributed in the social graph. From this perspective, we propose a privacy model that accounts for information dependence between edges in the graph and define a notion of what constitutes an edge's neighborhood. Edges may be labeled with a set of zero or more properties. For example, an edge may be labeled with the token "acquisition" if the correspondents discussed the topic "acquisition".

In our work we focus on the task of safely releasing a set of edge properties present in the graph. For this task, we are limited to L-Lipschitz queries which are sufficient to cover counting and frequency queries. For language tasks, this can be viewed as extracting common n-grams from correspondences (Gopi et al., 2020; Durfee & Rogers, 2019).

2 Mechanism Design

We consider a graph representation of the organizational communications, consisting of nodes for individuals and edges for the correspondence among them.

In our neighborhood model, we capture the correlation among the *adjacent edges* as shown in Figure 1. We define the graph as a union of neighborhoods, where each neighborhood is defined as a central edge and its adjacent neighbors. We apply a conditional independence assumption that knowledge of the adjacent edge properties is sufficient to determine the properties of the central edge, independent of the rest of the graph.

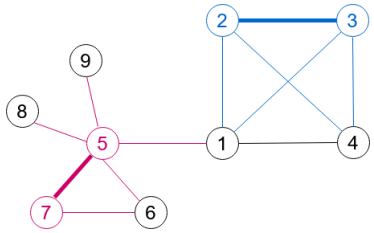

Figure 1: Neighborhood correlation, each edge is correlated with its adjacent edges. The 23 edge is adjacent to edges 21, 24, 31 and 34. Edge 57 is adjacent to edges 51, 56, 58, 59, and 76.

A change in an edge's properties will influence its neighbors. Using Pufferfish privacy, if we can model the effect of this change probabilistically, we can compute the Wasserstein distance between query distributions, providing a sensitivity measure that accounts for an edge's correlation with its neighbors.

2.1 Privacy Definition

We use Pufferfish privacy to design a private mechanism for counting one property in the graph, and extend it to all property counts:

1. The database is a set of records: $X = \{X_1, \cdots, X_N\}$; $X_i = 0$ or $X_i = 1$ corresponding to the events the edge *i* has the property or not, indicating complementary secrets s_i^0 or s_i^1.
2. The Pufferfish parameters: $(\mathcal{S}, \mathcal{S}_{\text{pairs}}, \Theta)$: the set of secrets $\mathcal{S} = \{s_i^0, s_i^1; i = 1, \ldots, n\}$, the secret pairs to be indistinguishable $\mathcal{S}_{\text{pairs}} = \{(s_i^0, s_i^1), i = 1, \ldots, n\}$, and Θ, the set of models describing the correlation. The Pufferfish privacy guarantee is shown in equation 1.

$$\Pr_{M,\theta}(M(X) = w | s_i^0, \theta) \leq e^{\epsilon} \Pr_{M,\theta}(M(X) = w | s_i^1, \theta) \tag{1}$$

   (a) $\mathcal{S}$ consists of the binary values of each X_i, $i = 1, \ldots, n$.
   (b) $\mathcal{S}_{\text{pairs}}$ is (s_i^0, s_i^1), indicating presence/absence of a property on an edge.
   (c) Θ: neighborhood correlation $\theta \in \Theta$. We use the Markov Quilt mechanism from Song et al. (2017). We empirically measure the exact correlation inside the quilt.

(d) Query f: Maps the dataset X into a scalar $f(X) = |i \in \{1, \ldots, n\} : X_i = 1|$, counting the number of edges having the property of interest.

3. Markov Quilt. We assume that each edge X_i is correlated to its adjacent edges X_N and has no correlation with the rest of the graph (neither to X_R, nor to X_Q), i.e. $\delta = 0$ in Song et al. (2017). card(X_N) translates to the maximum number of adjacent edges to an edge, i.e. $2 \times D_{max} - 1$, where D_{max} is the maximum degree of a node in the graph, and we subtract 1 so as not to double-count the central edge itself. $2 \times D_{max}$ group differential privacy (Corollary A.1.1) would be a baseline for our case.

2.2 Correlation Models

To assess the Wasserstein distance between neighboring secret pairs (changing a single property from true to false or vice-versa), we require a model of $\Pr(f(X) = w|s_i, \theta)$ that can be used to estimate by how much a neighborhood's labels might change due to a change in the central edge's label. By applying the Markov assumption, we need only measure the impact of a label change on an edge's immediate neighborhood (i.e. its impact on the Markov quilt)– w is measured for the local neighborhood and the rest of the graph is assumed to be constant. We estimate three models for our experiments:

- **Conditional Model** estimates the probability $\Pr(f(X) = w|s_i, \deg(X_i), \text{freq}(a))$, where $\deg(X_i)$ is the number of edges adjacent to edge X_i, $\text{freq}(a)$ is the attacker's prior on the frequency of property a.
- **Global Model** ignores $\deg(X_i)$ and $\text{freq}(a)$ and empirically measures $\Pr(f(X) = w|s_i^j)$ for secrets s_i^0 and s_i^1– a normalized frequency histogram of how often $f(X) = w$ when the central edge's property is set, s_i^1, and a separate histogram for when it is not set s_i^0.
- **Binomial Model** empirically estimates $p_i = \Pr(s_j|s_i)$, the probability distribution over a randomly selected adjacent edge's secrets, given the label of the central edge, and then estimates $\Pr(f(X) = w|s_i)$ as a Binomial distribution parameterized by p_i and $\deg(X_i)$: $P(f(X) = w|s_i) = \text{Binomial}(\deg(X_i), p_i)$.

The Wasserstein distance $W = max_{X_i \in X} W_\infty(X_i)$ is measured as follows: for each neighborhood in the graph, instantiate distributions $\Pr(f(X)|s_i^0)$ and $\Pr(f(X)|s_i^1)$, and measure W_∞ as the maximum horizontal distance between their respective cumulative distribution functions. W is then the maximal W_∞ over all neighborhoods. Note that W is bounded above by the largest neighborhood size: flipping a single edge property may trigger a flip in at most $\deg(X_i)$ adjacent edges.

3 Experiments

We run our experiments on the Avocado corpus (Oard et al., 2015). The complete graph contains 393 nodes (individuals) and 21312 edges (correspondence). The largest neighborhood in the graph consists of 1883 edges.

We extract unigrams and bi-grams from messages passed between edges and set the edge property X_i^a to "true" for each n-gram a. Thus, an edge with the property "acquisition" set to true indicates that at least one message passed between the connected nodes containing the word "acquisition". Edges with no such property are implicitly "false" for that property.

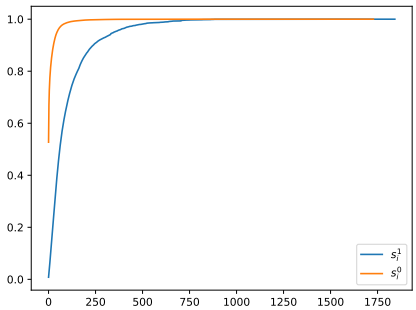

Figure 2: Cumulative distributions of $\Pr(f(X) = w|s_i^j)$ for the **Global** model, conditioned on secret s_i^j.

We construct the three correlation models, as described in section 2.2. Table 1 shows the estimated Wasserstein measures for the various property frequencies and neighborhood size under the **Conditional** correlation model. The maximum influence W_∞ of a bucket is scaled by the maximum neighborhood size for the bucket, up to the largest possible neighborhood in the graph $N_{max} = 1883$.

While the largest W corresponds to large neighborhoods ($\log(\textit{deg}) = 3$), it does not necessarily correspond to high-frequency or low-frequency properties.

The **Global** model measures $\Pr(f(X) = w|s_i)$ directly for secrets s_i^0 and s_i^1. The cumulative distribution functions of these measures are shown in Figure 2. The maximal Wasserstein measure is 866, corresponding to the maximum horizontal distance between these two cumulative distributions.

The **Binomial** model represents the two label distributions by estimating Bernoulli parameters p_0 and p_1 for each label respectively, and measuring the maximal Wasserstein distance between these distributions $\Pr(f(X) = w|s_i^j) = \text{Binomial}(\deg(X_i), p_j)$. Using this approach we emprically measure p_0 to be 0.028 and p_1 to be 0.274 and the maximal Wasserstein measure to be 558. These parameters indicate that an adjacent edge is about ten times more likely to have property a if the central edge has property a.

3.1 Language Task

We apply the privacy mechanism to differentially private set union (DPSU) (Gopi et al., 2020). DPSU aims to identify the union of elements in k input sets (in our setting, sets of edge properties on k edges). Each edge has a contribution limit up to c properties. To account for edge correlation, it is necessary to scale the sensitivity of the property counts by cW, as changing any edge can change as many as c properties and may influence its neighborhood by a factor as large as W. We compare the yield (the number of published n-grams) of the privacy mechanism over ten independent applications of the mechanism, for each of the three correlation models. We also provide baseline yield for node-level, edge-level, and group privacy. We choose $\epsilon = 100$, reflecting the relatively small size of the input graph, and $c = 1000$. The results of this experiment are shown in Table 2.

$\log(\textit{freq})$	$\log(\textit{deg})$	W_∞	W
0	0	1.0	10.0
0	1	0.08	8.0
0	2	0.02	20.0
0	3	0.01	18.83
1	0	1.0	10.0
1	1	0.58	57.0
1	2	0.5	500.0
1	3	0.09	169.47
2	0	0.71	7.1
2	1	0.66	66.0
2	2	0.74	740.0
2	3	0.51	960.33
3	0	0.5	5.0
3	1	0.31	31
3	2	0.37	370.0
3	3	0.29	546.07
4	0	0.36	3.6
4	1	0.16	16.0
4	2	0.21	210.0
4	3	0.13	244.79

Table 1: Wasserstein metrics for edges with neighborhoods of size *deg* and properties with global frequency *freq*. The boxed row represents the highest sensitivity.

The best result corresponds to edge-level privacy, which neglects to account for edge-neighborhood correlation. Of the approaches that address correlation, the binomial model yields the largest set of n-grams. If ϵ is large, and the number of private entities numbers in the tens of thousands, the yield of a DP mechanism can still be very limited. Our results illustrate how a privacy mechanism can be appropriately modified to account for neighborhood correlations in the graph, and we believe algorithmic yields can only improve with larger graphs.

4 Conclusion

In this paper we explored the problem of preserving organizational confidentiality in language tasks, leveraging the Pufferfish privacy framework, while addressing non-IID data among edges in the organization's social network. Our proposed scheme presents a compromise between two extreme measures of privacy, record-level differential privacy and group privacy. By taking record correlation into account, our scheme provides a more meaningful notion of privacy than record-level differential pri-

Description	W	$E[\texttt{yield}]$	σ
Edge-level privacy	1	24814.9	46.3
Node privacy	1	91	3.35
Binomial Model	558	228.7	5.71
Global Model	866	135	6.51
Conditional model	960	116.7	7.29
Group privacy	1883	41.9	3.11

Table 2: Experimental results for DPSU, $\epsilon = 100$. We measure yield- the number of extracted n-grams for each of the correlation models, as well as for node-level, edge-level, and group privacy.

vacy, while improving the utility compared
to group privacy. We summarize the contributions of this work as the following: i) the privacy mechanism protects edges, ii) the mechanism accounts for correlation between neighboring edges, iii) the mechanism protects against changes to edge properties, but not to changes in graph structure. Our mechanism preserves the privacy while assuming: i) the graph structure is known to attackers, and ii) attackers may have access to a data generation model θ that can predict an edge's properties, given its neighbors'.

REFERENCES

Martin Adam, Michael Wessel, and Alexander Benlian. Ai-based chatbots in customer service and their effects on user compliance. *Electronic Markets*, pp. 1–19, 2020.

J. Blocki, A. Blum, A. Datta, and O. Sheffet. Differentially private data analysis of social networks via restricted sensitivity. In *ITCS*, pp. 87–96, 2013.

Mia Xu Chen, Benjamin N Lee, Gagan Bansal, Yuan Cao, Shuyuan Zhang, Justin Lu, Jackie Tsay, Yinan Wang, Andrew M Dai, Zhifeng Chen, et al. Gmail smart compose: Real-time assisted writing. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, KDD'19, pp. 196–206, New York, NY, USA, 2019. Association for Computing Machinery.

Rui Chen, Benjamin C.M.Fung, Philip S.Yu, and Bipin C.Desai. Correlated network data publication via differential privacy. *The VLDB Journal*, 2014.

S. Chen and S. Zhou. Recursive mechanism: Towards node differential privacy and unrestricted joins. In *SIGMOD*, pp. 653–664, 2013.

James Cheng, Ada Fu, and Jia Liu. K-isomorphism: Privacy preserving network publication against structural attacks. pp. 459–470, 01 2010. doi: 10.1145/1807167.1807218.

David Durfee and Ryan Rogers. Practical differentially private top-k selection with pay-what-you-get composition. *CoRR*, abs/1905.04273, 2019. URL http://arxiv.org/abs/1905.04273.

C. Dwork, F. McSherry, K. Nissim, and A. Smith. In *Proceedings of the Third Conference on Theory of Cryptography*, pp. 265–284, 2006.

Tianchong Gao, Feng Li, Yu Chen, and XuKai Zou. Preserving local differential privacy in online social networks. In *International Conference on Wireless Algorithms, Systems, and Applications*, pp. 393–405. Springer, 2017.

Arpita Ghosh and Robert Kleinberg. Inferential privacy guarantees for differentially private mechanisms. abs/1603.01508, 2018.

Sivakanth Gopi, Pankaj Gulhane, Janardhan Kulkarni, Judy Hanwen Shen, Milad Shokouhiand, and Sergey Yekhanin. Differentially private set union. *ICML*, 2020.

Michael Hay, Chao Li, Gerome Miklau, and David Jensen. Accurate estimation of the degree distribution of private networks. In *Proceedings of the 2009 Ninth IEEE International Conference on Data Mining*, ICDM '09, pp. 169–178, USA, 2009. IEEE Computer Society. ISBN 9780769538952. doi: 10.1109/ICDM.2009.11. URL https://doi.org/10.1109/ICDM.2009.11.

X. He, A. Machanavajjhala, and B. Ding. Blowfish privacy: Tuning privacy-utility trade-offs using policies. SIGMOD '14, pp. 1447–1458, New York, NY, USA, 2014. Association for Computing Machinery. ISBN 9781450323765. doi: 10.1145/2588555.2588581. URL https://doi.org/10.1145/2588555.2588581.

F. L. Hitchcock. The distribution of a product from several sources to numerous localities. *Journal of Mathematics and Physics*, 20(1-4):224–230, 1941. doi: https://doi.org/10.1002/sapm1941201224. URL https://onlinelibrary.wiley.com/doi/abs/10.1002/sapm1941201224.

Vishesh Karwa, Sofya Raskhodnikova, Adam Smith, and Grigory Yaroslavtsev. Private analysis of graph structure. *ACM Trans. Database Syst.*, 2014.

S.P. Kasiviswanathan, K. Nissim, and S.Raskhodnikova. Analyzing graphs with node differential privacy. In *TCC*, pp. 81–98, 2013.

Michael Kearns, Aaron Roth, Zhiwei Steven Wu, and Grigory Yaroslavtsev. Private algorithms for the protected in social network search. *Proceedings of the National Academy of Sciences*, 113(4): 913–918, 2016.

D. Kifer and A. Machanavajjhala. Pufferfish: A framework for mathematical privacy definitions. *Association for Computing Machinery*, 1:39, 2014.

Changchang Liu, Supriyo Chakraborty, and Prateek Mittal. Dependence makes you vulnerable: Differential privacy under dependent tuples. In *NDSS*, volume 16, pp. 21–24, 2016.

F. McSherry and K. Talwar. In *Proceedings of the 48th IEEE Symposium on Foundations of Computer Science (FOCS)*, pp. 94–103, 2007.

F. D. McSherry. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In *Proceedings of the 2009 ACM SIGMOD International Conference on Management of Data*, SIGMOD '09, pp. 19–30, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605585512. doi: 10.1145/1559845.1559850. URL https://doi.org/10.1145/1559845.1559850.

Prateek Mittal, Charalampos Papamanthou, and Dawn Song. Preserving link privacy in social network based systems. 08 2012.

Kobbi Nissim, Sofya Raskhodnikova, and Adam Smith. Smooth sensitivity and sampling in private data analysis. In *Proceedings of the Thirty-Ninth Annual ACM Symposium on Theory of Computing*, STOC '07, pp. 75–84, New York, NY, USA, 2007. Association for Computing Machinery. ISBN 9781595936318. doi: 10.1145/1250790.1250803. URL https://doi.org/10.1145/1250790.1250803.

Douglas Oard, William Webber, David Kirsch, and Sergey Golitsynskiy. Avocado research email collection. https://catalog.ldc.upenn.edu/LDC2015T03, 2015.

S. Raskhodnikova and A. Smith. Efficient lipschitz extensions for high-dimensional graph statistics and node private degree distributions. *FOCS*, 2016.

A. Sala, X. Zhao, C. Wilson, H. Zheng, and B.Y. Zhao. Sharing graphs using differentially private graph models. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pp. 81–98, 2011.

S. Song, Y. Wang, and K. Chaudhuri. Pufferfish privacy mechanisms for correlated data. In *Proceedings of the 2017 ACM International Conference on Management of Data*, pp. 1291–1306, 2017.

Mingxuan Yuan, Lei Chen, and Philip S. Yu. Personalized privacy protection in social networks. *Proc. VLDB Endow.*, 4(2):141–150, November 2010. ISSN 2150-8097. doi: 10.14778/1921071.1921080. URL https://doi.org/10.14778/1921071.1921080.

X. Zheng, J. Han, and A. Sun. A survey of location prediction on twitter. *IEEE Transactions on Knowledge and Data Engineering*, 30(9):1652–1671, 2018. doi: 10.1109/TKDE.2018.2807840.

Tianqing Zhu, Gang Li, Wanlei Zhou, and S Yu Philip. Differentially private social network data publishing. In *Differential Privacy and Applications*, pp. 91–105. Springer, 2017.

A APPENDIX

Here, we provide the detailed definitions for the concepts used throughout the paper.

A.1 DIFFERENTIAL PRIVACY

Differential privacy (DP) is a mathematical framework that offers strong and robust guarantees at protecting user privacy under the release of a query function calculated over a statistical database (Dwork et al., 2006). The main idea to achieve DP is perturbing the query function by the introduction of random noise generated according to a carefully chosen distribution. We formally define the notion of ϵ-differential privacy in the following:

Definition A.1 (ϵ-Differential Privacy (Dwork et al., 2006)) *A randomized algorithm $\mathcal{M}$ is said to provide ϵ-differential privacy if for any two databases D, D' differing in only a single entry, and for any set $S \subseteq \text{Range}(M)$,*

$$\frac{\Pr(\mathcal{M}(D) \in S)}{\Pr(\mathcal{M}(D') \in S)} \leq \exp(\epsilon)$$

the probability taken over the randomness of the algorithm $\mathcal{M}$.

A typical way to achieve ϵ-differential privacy is to apply the Laplace mechanism as shown in McSherry & Talwar (2007). The main idea is to apply noise to the output of a query for the sake of perturbation and the amount of noise depends on the global sensitivity of the query and the privacy budget ϵ. Let us first define the global sensitivity.

Definition A.2 (Global sensitivity) *Let d be a positive integer and $\mathcal{D}$ be a collection of datasets. For any function $f : \mathcal{D} \to \mathbb{R}^d$, the global sensitivity of f, denoted by Δf, is defined by*

$$\Delta f = \max_{D_1, D_2} ||f(D_1) - f(D_2)||_1$$

where D_1 and D_2 are datasets in $\mathcal{D}$ differing in at most one element and $||\cdot||_1$ denotes the ℓ_1 norm.

Theorem A.1 *Let d be a positive integer and $\mathcal{D}$ be a collection of datasets. For any $f : \mathcal{D} \to \mathbb{R}^d$, the randomized mechanism $\mathcal{M}$*

$$\mathcal{M}(\mathcal{D}) = f(\mathcal{D}) + \text{Laplace}(0, \Delta f/\epsilon)$$

satisfies ϵ-differential privacy.

Global sensitivity is a bound on the maximum effect of any element in the dataset, which will be helpful to provide privacy to *all* elements in the dataset. Based on this, for a given dataset D, a function f, and the privacy parameter ϵ, the Laplace mechanism adds $\text{Laplace}(0, \lambda)$ noise to the output of f where the parameter λ is determined by both Δf and ϵ.

Note that in the definition of ϵ-differential privacy, the guarantee holds for a single data entry. However, by applying the composability property of differential privacy (McSherry, 2009), the setting can be extended to multiple data entries. If the goal is to protect a group, this can be achieved by setting ϵ to ϵ/k for any $k \in \mathbb{N}$ where k represents the size of the group. In this case, all groups of size k are ϵ-differentially private protected. We summarize this by formally defining group differential privacy in the following corollary:

Corollary A.1.1 (Group Differential Privacy) *A randomized algorithm $\mathcal{M}$ is said to provide ϵ-differential privacy for all groups of size k if for any two databases D, D' differing in at most k entries, and for any set $S \subseteq \text{Range}(M)$,*

$$\frac{\Pr(\mathcal{M}(D) \in S)}{\Pr(\mathcal{M}(D') \in S)} \leq \exp(\epsilon)$$

where the probability is taken over the randomness of the algorithm $\mathcal{M}$. Any ϵ/k-differentially private algorithm is ϵ-differentially private for all groups of size k.

We point out that the scaling of the noise is inversely proportional to the privacy budget ϵ. Therefore, setting ϵ to ϵ/k will in turn change the noise level from $\Delta f/\epsilon$ to $k \cdot \Delta f/\epsilon$, which may significantly decrease the utility of the query. However, this is the price to pay to obtain stronger privacy guarantees with group differential privacy.

A.2 PUFFERFISH PRIVACY

Differential privacy provides robust guarantees for a wide range of database queries. For our scenario, it is useful to consider Pufferfish privacy (Kifer & Machanavajjhala, 2014), which is a Bayesian privacy framework providing rigorous privacy guarantees against many types of attackers. An advantage of the Pufferfish framework is that a domain expert can develop rigorous privacy definitions for their data sharing needs without holding expertise in privacy. This is achieved by specifying three components in the Pufferfish privacy framework: a set $\mathcal{S}$ of potential secrets, a set $\mathcal{S}_{\text{pairs}} \subseteq \mathcal{S} \times \mathcal{S}$ of discriminative secret pairs, and a collection of data distributions Θ. The Pufferfish framework provides a rich class of privacy definitions based on the components specified by a domain expert. We formally define the framework in the following based on Kifer & Machanavajjhala (2014).

Definition A.3 (Pufferfish Privacy) *A randomized algorithm $\mathcal{M}$ is said to provide ϵ-Pufferfish privacy for a domain $(\mathcal{S}, \mathcal{S}_{\text{pairs}}, \Theta)$ if for all distributions $\theta \in \Theta$, for all secret pairs $(s_i, s_j) \in \mathcal{S}_{\text{pairs}}$, and for all possible outputs $w \in \text{Range}(\mathcal{M})$ it satisfies*

$$\left| \frac{\Pr_{\mathcal{M},\theta}(\mathcal{M}(\mathcal{D}) = w | s_i, \theta)}{\Pr_{\mathcal{M},\theta}(\mathcal{M}(\mathcal{D}) = w | s_j, \theta)} \right| \leq \exp(\epsilon)$$

where $\mathcal{D}$ is drawn from the distribution θ and s_i and s_j are such that $\Pr(s_i|\theta) \neq 0$ and $\Pr(s_j|\theta) \neq 0$.

We note that there is an additional source of randomness in the definition of Pufferfish privacy. The dataset $\mathcal{D}$ is itself a random variable that is drawn from a distribution $\theta \in \Theta$.

In words, a domain expert constructs the set S for the potential secrets that are desired to be hidden (e.g. private data of an individual). $\mathcal{S}_{\text{pairs}}$ is simply the pair of such potential secrets that we would like to guarantee are indistinguishable in evaluating $\mathcal{M}$. Finally, Θ is a collection of distributions where each probability distribution $\theta \in \Theta$ corresponds to an attacker to be protected against. Θ can be selected based on the fine grain of how data can be plausibly generated and it also reflects the attackers' beliefs in how the data were generated (incorporating any background knowledge and side information). The whole process gives the domain expert flexibility to customize privacy to the specific set of secrets and data generation scenarios that are typical in their domain.

We further point out that Pufferfish privacy provides a general framework in the sense that it covers ϵ-differential privacy as an instantiation for a particular choice of domain $(\mathcal{S}, \mathcal{S}_{\text{pairs}}, \Theta)$ (see Theorem 6.1 in Kifer & Machanavajjhala (2014)).

A.3 WASSERSTEIN MECHANISM

While there is no efficient general mechanism that applies to any Pufferfish instantiation, there are a number of mechanisms for specific Pufferfish instantiations (Kifer & Machanavajjhala, 2014; He et al., 2014). For general Pufferfish instantiation, Song et al. (2017) introduce a mechanism that achieves Pufferfish privacy, but does not satisfy efficiency in its original form. We introduce their base mechanism here. Later in Section 2, we present our adjustments to their mechanism that makes it efficient to utilize for our use case of enterprise communications.

The main idea of the mechanism in Song et al. (2017) is similar to the Laplace mechanism in differential privacy. Instead of adding noise based on the global sensitivity Δf in differential privacy, Song et al. use the distributions $\Pr(f(\mathcal{D})|s_i, \theta)$ and $\Pr(f(\mathcal{D})|s_j, \theta)$ in the Pufferfish framework, propose a metric quantifying the worst case distance between these two distributions, and inject noise proportional to this distance. They find that the ∞-Wasserstein distance is the right choice for this purpose.

Definition A.4 (∞-Wasserstein distance) *Let μ, ν be two probability distributions on $\mathbb{R}$ and $\tau(\mu, \nu)$ denote the set of all joint distributions with marginals μ and ν. The ∞-Wasserstein distance between μ and ν is defined as*

$$W_\infty(\mu, \nu) = \inf_{\gamma \in \tau(\mu,\nu)} \max_{(x,y) \in \text{support}(\gamma)} | x - y | .$$

Intuitively, W_∞ measures the maximal distance that any probability mass moves while transforming μ to ν in the most optimal way possible. W_∞ is related to the well-known Earth Mover's Distance

in that it accounts for the maximal shift in probability over the domain of τ but not the amount of mass in the shift (Hitchcock, 1941).

Based on the ∞-Wasserstein distance, the Wasserstein mechanism calculates the maximum over $(s_i, s_j) \in \mathcal{S}_{\text{pairs}}$ and $\theta \in \Theta$, analogous to Δf, and applies the Laplace noise proportional to the maximum ∞-Wasserstein distance. It is proven in Song et al. (2017) (see Theorem 3.2) that this mechanism yields ϵ-Pufferfish privacy.

Theorem A.2 (Wasserstein mechanism) *Let $(\mathcal{S}, \mathcal{S}_{\text{pairs}}, \Theta)$ be a domain. For any function $f : \mathcal{D} \to \mathbb{R}$ the randomized mechanism $\mathcal{M}$*

$$\mathcal{M}(\mathcal{D}) = f(\mathcal{D}) + \text{Laplace}(0, W/\epsilon)$$

where $W = \sup_{(s_i,s_j)\in\mathcal{S}_{\text{pairs}},\theta\in\Theta} W_\infty(\mu_{i\theta}, \nu_{j\theta})$ for $\mu_{i,\theta} = \Pr(f(\mathcal{D})|s_i, \theta)$ and $\nu_{j,\theta} = \Pr(f(\mathcal{D})|s_j, \theta)$ satisfies ϵ-Pufferfish privacy.

A.4 Markov Quilt Mechanism

The Wasserstein mechanism can be quite expensive in terms of computational complexity, as it requires modeling the effects of varying all complementary secret pairs on the function f. Song et al. (2017) introduce the Markov Quilt mechanism for the special case where the dependence inside a dataset can be described by a Bayesian network, which fits to our setting of interest. In the case where the dependence is most effective in the "local" neighborhood, the amount of noise can be calibrated with respect to the size of this neighborhood. To this end, max-influence of a variable $\mathcal{D}_i$ on a set of variables $\mathcal{D}_A$ under a distribution class Θ is defined as

$$e_\Theta(\mathcal{D}_A|\mathcal{D}_i) = \sup_{\theta\in\Theta} \max_{a,b\in\mathcal{X}} \max_{d_A\in\mathcal{X}^{|\mathcal{D}_A|}} \log \frac{\Pr(\mathcal{D}_A = d_A|\mathcal{D}_i = a, \theta)}{\Pr(\mathcal{D}_A = d_A|\mathcal{D}_i = b, \theta)}$$

where $\mathcal{X}$ denotes the range of each $\mathcal{D}_i$.

In terms of privacy it is an advantage that the dependence stays as "local" as possible if one can find a large set $\mathcal{D}_A$ such that $\mathcal{D}_i$ has low max-influence on $\mathcal{D}_A$. Especially if one can claim certain conditional independence from a variable towards some part of the dataset it can also simplify the calibration of the noise. The following notion is helpful to show what is described here.

Definition A.5 (Markov Quilt) *A set of variables $\mathcal{D}_Q$ in a dataset is a Markov Quilt for a variable $\mathcal{D}_i$ if there exists a set $\mathcal{D}_i \in \mathcal{D}_N$ such that $\mathcal{D} = \mathcal{D}_N \cup \mathcal{D}_Q \cup \mathcal{D}_R$ and $\mathcal{D}_i$ is conditionally independent from $\mathcal{D}_R$ given $\mathcal{D}_Q$, i.e. $\Pr(\mathcal{D}_R|\mathcal{D}_Q, \mathcal{D}_i) = \Pr(\mathcal{D}_R|\mathcal{D}_Q)$.*

In this formulation Song et al. (2017) choose the subscripts N and R to represent "nearby" and "remote" nodes in the Bayesian network, respectively, with the Q (quilt) nodes separating them and establishing conditional independence.

Based on this notion the Markov Quilt mechanism protects a variable $\mathcal{D}_i$ by adding Laplace noise to a L-Lipschitz query f with scale parameter $L \times |\mathcal{D}_N|/(\epsilon - \delta)$ where δ is an upper bound on the max-influence of $\mathcal{D}_i$ on $\mathcal{D}_Q$. We note that the effect of $\mathcal{D}_i$ is obscured with noise whose amount is based on the cardinality of the local variables ($|\mathcal{D}_N|$) and a correction term to account for the effect of the distant variables (δ). Naturally, the privacy of all variables can be protected by adding noise with the maximum scale parameter over all variables $\mathcal{D}_i \in \mathcal{D}$. It is shown in Song et al. (2017) (see Theorem 4.3) that this mechanism yields ϵ-Pufferfish privacy. It is also proven that Markov Quilt mechanism satisfies sequential composition (see Theorem 4.4 in Song et al. (2017)).

Theorem A.3 (Markov Quilt Mechanism) *Let $(\mathcal{S}, \mathcal{S}_{\text{pairs}}, \Theta)$ be a domain. For any L-Lipschitz function f if each $\mathcal{D}_i \in \mathcal{D}$ has the trivial quilt $\mathcal{D}_Q = \emptyset$ (with $\mathcal{D}_N = \mathcal{D}$, $\mathcal{D}_R = \emptyset$), then the Markov Quilt Mechanism provides ϵ-Pufferfish privacy.*