Natural Differential Privacy

Micah Altman
Center for Research in Equitable and Open Scholarship
Massachusetts Institute of Technology
Cambridge, MA
<escience@mit.edu>

Aloni Cohen
Department of Computer Science
University of Chicago
Chicago, IL
<aloni@uchicago.edu>

***Abstract*— We introduce "Natural" differential privacy (NDP) -- which utilizes features of existing hardware architecture to implement differentially private computations. We show that NDP both guarantees strong bounds on privacy loss and constitutes a practical exception to no-free-lunch theorems on privacy. We describe how NDP can be efficiently implemented and how it aligns with recognized privacy principles and frameworks.**

Keywords— differential privacy, physical mechanisms, no free lunch, privacy by design, privacy by default

I. Introduction

Differential privacy offers provable privacy guarantees [6] but has been criticized as difficult to integrate into existing data production systems and requiring substantial utility loss. [7]

We introduce *natural differential privacy (NDP)* -- a framework for guaranteeing differential privacy for arbitrary computations by leveraging features of existing hardware architectures and natural sources of entropy.

NDP provides all the advantages of "pure" differential privacy as originally formulated in [35], not resorting to any of the myriad relaxed definitions that provide weaker guarantees [36]. NDP provides a worst-case bound on the privacy loss that is substantially better than some high-profile, large-scale commercial implementations of DP.

Moreover, in contrast to existing implementations, NDP provides privacy by default for *all* computations on a platform. Furthermore, NDP is simple and inexpensive to implement at a large scale and requires no practical reduction in utility or performance.

II. Preliminaries & Definition

We consider the setting of a dataset x consisting of n records, where each record is a bitstring of dimension d. We view each row as containing the data of a single individual. Databases x and x' are *neighboring* if they differ in at most one record. A mechanism M is a randomized mapping from datasets to some set of possible outputs Y.

Definition 2.1 (ε-Differential Privacy (ε-DP) [3]). M is *ε-differentially private* if for all neighboring datasets x and x', and for all sets $S \subseteq Y$:

$$\Pr[M(x) \in S] \leq e^{\varepsilon} \cdot \Pr[M(x') \in S] \qquad (1)$$

where the probabilities are taken over M's coins.

NDP is defined as any system that integrates DP protections directly in a Von Neumann architecture [8] via hardware implementation using persistent sources of entropy for noise infusion. By construction, every internal computation by an NDP system integrates noise infusion guaranteeing (ε, 0)-DP. And because DP preserves privacy under postprocessing -- all outputs from the system are thus (ε, 0)-DP

Implementation: NDP applies to the computation of arbitrary m-bit functions f of the data x, for any m. To evaluate the NDP-version of f, one simply evaluates f on a RAM machine. For best results, the RAM should be operated at or above sea level.

We use as a building block the Randomized Response mechanism. The Randomized Response mechanism is parameterized by a probability $0<p<0.5$, and we denote the corresponding mechanism RR_p. RR_p takes as input a bit $b \in \{0,1\}$, and outputs 1-b with probability p, otherwise outputting b. Results established in [16] provide a formula for the exact equivalence between the probability of randomized response and the epsilon parameter. For any $p < 0.5$, equation 2 expresses this relationship:

$$1-p = \exp(\varepsilon)/(1 + \exp(\varepsilon)) \qquad (2)$$

Privacy parameters: Let $T_{in} > 0$ and $T_{out} > 0$ be the length of time that the input x and output f(x) are stored in RAM over the course of the computation, respectively.[1]

Each computation has a corresponding parameter q that depends on the environment within which the computation is performed. Thus q is the probability of any single bit flip caused by cosmic rays occurring on 1 GB of RAM over the course of 1 day (see Table 1). From q, it is easy to derive the probability that any single bit is flipped in the period T_{in} or T_{out}. Using p and applying equation 1, it is straightforward to solve for ε.

III. Related Work

Current implementations of DP at scale have used artificial (non-natural) DP. Because of the substantial utility tradeoffs that artificial DP often requires -- commercial implementations often use values of epsilon well above 1. Recent large-scale implementations of differential privacy by major corporations (including Google and Apple) have employed effective epsilon levels ranging from dozens to hundreds -- with one major implementation exceeding seven hundred and fifty. [2] [28]

Other variants of DP, such as epsilon-delta DP and concentrated DP have been proposed [4]. However, these variants relax the definition of DP yielding weaker privacy properties. (We refer henceforth to such relaxations as 'artificial.')

Natural sources of entropy for noise diffusion have been studied for over four decades [10]. Their importance in security and privacy has been recognized in related areas:

- Bit flipping induced by radiation or other environmental conditions has been previously used for practical attacks against system security [9,32].
- The importance of high-quality random number generation for all differential privacy methods has recently been recognized. [34] In practice, nearly all implementations rely on pseudo-random sequences seeded from a physical entropy source. The use of physical sources of randomness for direct noise infusion has not been well-examined.
- More recently, the inherent instability of quantum computation has been examined as a theoretical source of protective noise infusion -- although practical implementation remains far off. [33]

IV. Advantages of NDP

Although natural noise infusion has been studied in related work, the use of natural sources directly for differential privacy is novel. The NDP approach offers a number of advantages:

- NDP protects all computations made on a system.
- NDP does not require any relaxation of the formal differential privacy guarantees -- unlike artificial DP variants.
- NDP is simple to implement and inexpensive to deploy.
- NDP provides protections that are substantively equivalent (or better) than the formal guarantees provided by notable commercial implementations -- while maintaining higher utility, substantially reducing implementation cost, and extending protection to a broader range of computations.

Further, DP has additional attractive features:

1. First, NDP encourages privacy by design [15] -- NDP can be integrated into hardware architecture, systems design, and facility deployment, as well as at the application level.
2. Second, NDP provides privacy by default [16] since a floor for protection is provided for all users without requiring any opt-ins [17].
3. Third, NDP aligns well with the widely adopted 'five safes framework' [20]. Specifically, it uses architectural privacy by design to guarantee 'safe outputs.'
4. Fourth, NDP can be implemented either at the time of manufacture or during deployment. This facilitates certification and auditing of secure hardware and facilities.
5. Fifth, NDP provides guaranteed, measurable privacy with zero marginal utility loss --

[1] If the portions of the input (respectively, output) are in RAM for different lengths of time, then we take T_{in} (resp., T_{out} to be the minimum time over every bit of the input (resp., output).]

[2] Estimates are for the effective protection of user information over a month of activity. Because of composition effects, the effective epsilon for protection of a user information in these systems grows geometrically over time. Thus a cumulative epsilons of ten thousands or more is possible for frequent, long-term users of these systems.

theoretical no-free-lunch theorems notwithstanding [5]. [3]

V. Setting Privacy Parameters Within NDP Systems

NDP, when fully integrated into the architecture, operates by default, continuously, and at the hardware level. In modern architectures, which implement random access memory using MOSFET technology, bit-flipping is an ideal mechanism for noise infusion.

At manufacturing time, the infusion of readily available alpha sources into memory chip packaging material requires no increase in manufacturing cost. This approach can be used to ensure a maximum bound for epsilon that is both auditable through inspection and can be verified through hardware certification processes. [4]

More generally, ionizing radiation provides a natural and ubiquitous source for inducing randomized response at the bit level -- acting through the injection of memory faults. Utilizing this physical mechanism for entropy provides true randomness, which provides stronger formal guarantees than pseudorandom number generation [27,34].

Results established in [14] provide a formula for the exact equivalence between the probability of randomized response and the epsilon parameter.

$$p = 1/(1 + \exp(\varepsilon)) \qquad (2)$$

Many sources of radiation are readily available, as shown in Figure 1.

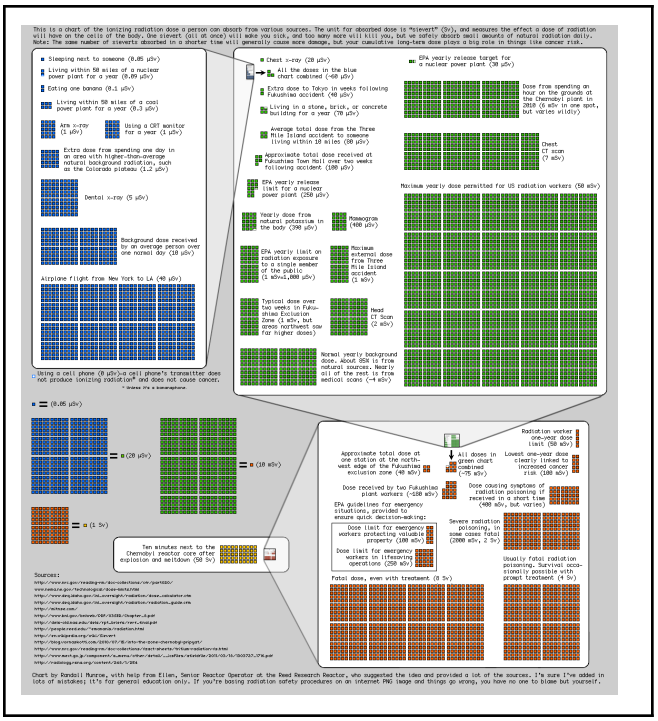

Fig. 1. Comparison of Readily Available Entropy Sources for Setting ε
(Source [2], placed in the public domain)

Post-manufacture, maximum bounds on epsilon can be further reduced at deployment time through the positioning of the hardware. Entire computing facilities may be certified as safe using this method.

Conveniently, cosmic rays cause i.i.d. bit flips at a rate directly related to the level of cosmic radiation (typically measured in micro-sieverts) exposure, which is itself a function of atmospheric density (above sea level) and crustal density (below sea level) at a given altitude. The levels of radiation experienced at specified altitude relative to sea level has been established empirically and fits a geometric distribution within near-earth orbiting distances. Using these empirical results and equation 2, we can derive the effective value of epsilon at exemplar locations, as shown in Table 2.

[3] In theory, theory and practice are the same -- in practice, they differ. See [18,19] for the original formulation by Brewster, and contributions attributed to Berra and Einstein.
[4] Even without changes in manufacturing processes, manufacturers can readily certify that their RAM is not unsafe by design -- in that it does not provide ECC or this feature has been disabled (e.g. at the firmware level).

TABLE I. *Privacy Budget Configuration through Altitudinal Adjustment*[5]

m* > *sea-level	***Exemplar Location***	***μSv/h***	***Error*** */GBxDay*	***Max*** ε
-3840	Mponeng gold mine	0	0	∞
10	Cambridge, MA	0.06	.2	33.70711
1742	Mount Wilson observatory	0.237	2	31.39832
10000	Jet airplane lower cruising	6	60	27.99537
781000	Iridium Satellite Constellation	60	600	25.69307

Note that relative to baseline, very low values of epsilon can be achieved through altitudinal adjustment. Further note that the level of epsilon provided naturally at sea level is more protective than the level provided by the most notable and largest scale production implementations of differential privacy to date. [13] Finally, in practice, the effective epsilon will be statistically indistinguishable from implementations using a theoretically lower value. [6]

At runtime, noise injection can readily and effectively be achieved by altering the thermal operating environment. [32] Further, in high-density computing deployments, simply reducing the level of external cooling will not only increase protective noise infusion but also reduce electricity usage -- benefitting the global environment. Moreover, various external noise injection tools are available and can provide additional protection without affecting the location or manufacturing process [11].

Finally, given the wide availability of (level-2) hypervisor-based system-level virtual machine technologies, simulation-based noise infusion (aka. synthetic natural differential privacy) can be used to produce any desired level of epsilon with a relatively small decrease in runtime performance [12].

VI. Limitations

On occasion, serious points are best conveyed through humor.[7] This article is intended as a work of (serious) humor. While each of the individual technical assertions in the article is true, catastrophic drawbacks are omitted or glossed over. Thus the substantial benefits claimed for the method, particularly in section IIIB, are parodically misleading.

For example, some of the limitations of the above proposal include the following:

- Use of ε above 1, although frequent in practice, requires caution.
- It is rarely the proper objective of law or public policy to protect the privacy of an 'event.' It is usually a more appropriate policy goal to protect the privacy of a persistent actor -- such as a person, organization, or designated group of people.
- Information about an event or other unit of protection is rarely limited to a single bit. Where measurements of the unit of protection comprise multiple bits, composition will multiplicatively increase the effective ε.
- Randomized response does not generally result in an efficient tradeoff between privacy and utility -- especially as the effective information measured grows large.
- Inducing bit flipping at random in main memory causes unwanted side effects -- such as entirely incorrect results, nonsense output, program failures, and system crashes.
- Thus, there is no market for unreliable RAM, despite the low cost of manufacture.
- Further, introducing ionizing radiation to the deployment site to achieve a meaningful level of protection may risk exposure to lethal amounts of radiation. This potentially violates

[5] Derivative-free numerical minimization [26] is used to obtain epsilon corresponding to p, given Equation 2. Bit-level frequency data is provided by [25,29]. Epsilon levels are calculated for protection at the (bit) event-level, which is the most practical unit of protection for streaming systems [24] -- and has been used at scale for public release of large scale Facebook data for scientific research [21.22] To calculate epsilon for other units of protection, it is straightforward to calculate the effective epsilon by using the standard dp composition formula across the maximum number of shared events in the computation [23]. Even under composition, the effective epsilon remains trivially small relative to the baseline.

[6] Assuming a continuous audit period of one hour, and conventional levels of statistical significance (p=.05) is used, the observed value at sea level will not be statistically distinguishable from a theoretical epsilon of 31.91007. (Zero bits flipped will be observed during that period at either level of epsilon, during at least 95% of audits).

[7] See [30] for foundations of this principle, and [31] for both a modern defense and exemplar of the principle.

local regulations, national laws, and international treaties.

These limitations notwithstanding, there is a sense in which cosmic rays formally induce ε-DP. Further, the epsilon values used by some large-scale commercial implementations of DP provide provable worst-case privacy-loss guarantees that are arguably no stronger than those provided by cosmic rays in Cambridge.

When very high values of epsilon are employed claims that such systems provide formal protection are misleading -- but this does not imply that such systems provide no protection: As techniques for measuring privacy-loss continue to evolve, we may come to understand that some systems are more protective than initially proved. Or these implementations may provide useful protection in particular contexts even if such protections are not formally provable.

Most important, this parody illustrates that neither formal privacy guarantees nor compliance with privacy principles are sufficient for adequate protection. Provable guarantees, such as those provided by differential privacy, have force only when the specific level of protection provided by implementation privacy parameters are meaningful and when the formal unit of protection corresponds to real-world entities with meaningful privacy interests. Moreover, when a system embeds a weak implementation of a protection mechanism at its core, compliance with other privacy principles, such as privacy-by-design, may offer little value.

Acknowledgment

We describe the authors' contributions following a standard taxonomy [1], AC and MA authored the first draft of the manuscript and had overall (ir)responsibility for content and revisions. All authors contributed to the conception of the report (including core ideas and statement of research questions), to the methodology, and to the writing through critical review and commentary. All authors contributed to substantiating the paper's arguments, revision, critical review, and commentary.

We thank Stephen Chong and Kobbi Nissim for their helpful comments on prior drafts.

Appendix: Reproducibility

The following R code is sufficient to replicate the results in table 1:

```{r}
#minimum reported failure rate in memory is .2 per errors GB/day

rate_GBday_wikipedia <- 0.2

# assumed quantity that new data remains resident in memory for 1 second

newdata_Gb_sec <-  1E9

# p value for randomized response from Wang et al. 2016
randp <- function(eps) {
  exp(eps)/(1+exp(eps))
}

# predicted bit errors per day
dayerr <- function(eps,
                   flow = newdata_Gb_sec
) {
  res <- (1-randp(eps))  * flow * 60 * 60 * 24
  res
}

# find epsilon with  5% change of observed error in an hour
optim(par=8,
fn=function(x){abs(.05-dayerr(eps=x)/24)}, method="Brent", lower=.1, upper=42)

# find epsilon for a given bits/day error rate
optim(par=8, fn = function(x) {
abs(rate_GBday_wikipedia -
dayerr(eps=x))}, method="Brent",
lower=.1, upper=42)
```

References

[1] Allen, L., Scott, J., Brand, A., Hlava, M. and Altman, M., 2014. Publishing: Credit where credit is due. Nature, 508(7496), pp.312-313.

[2] Munro, R., 2011 Ration Dose Chart, XKCD. < https://xkcd.com/radiation/ >

[3] C. Dwork, F. McSherry, K. Nissim, and A. Smith. Calibrating noise to sensitivity in private data analysis. In TCC, pages 265–284, 2006.

[4] Cynthia Dwork and Guy Rothblum. Concentrated differential privacy. CoRR, abs/1603.01887, 2016

[5] Kifer, D. and Machanavajjhala, A., 2011, June. No free lunch in data privacy. In Proceedings of the 2011 ACM SIGMOD International Conference on Management of data (pp. 193-204).

[6] Wood, A., Altman, M., Bembenek, A., Bun, M., Gaboardi, M., Honaker, J., Nissim, K., O'Brien, D.R., Steinke, T. and Vadhan, S., 2018.

Differential privacy: A primer for a non-technical audience. Vand. J. Ent. & Tech. L., 21, p.209.

[7] Blanco-Justicia, A., Sanchez, D., Domingo-Ferrer, J. and Muralidhar, K., 2022. A Critical Review on the Use (and Misuse) of Differential Privacy in Machine Learning. arXiv preprint arXiv:2206.04621.

[8] von Neumann, John (1945), First Draft of a Report on the EDVAC

[9] Dinaburg, A., 2011. Bitsquatting: DNS hijacking without exploitation. Proceedings of BlackHat Security.

[10] Ziegler, J. F.; Lanford, W. A. (1979). "Effect of Cosmic Rays on Computer Memories." Science. 206 (4420): 776–788.

[11]Hsueh, Mei-Chen, Timothy K. Tsai, and Ravishankar K. Iyer. "Fault injection techniques and tools." Computer 30, no. 4 (1997): 75-82.

[12] Li, Z., Kihl, M., Lu, Q. and Andersson, J.A., 2017, March. Performance overhead comparison between hypervisor and container based virtualization. In 2017 IEEE 31st International Conference on advanced information networking and applications (AINA) (pp. 955-962). IEEE.

[13] A. Greenberg. 2017. How one of Apple's key privacy safeguards falls short. Wired

[14] Y. Wang, X. Wu, and D. Hu. 2016. Using randomized response for differential privacy preserving data collection. In Proceedings of the EDBT/ICDT 2016 Joint Conference. Bordeaux, France

[15] Cavoukian, Ann. "Privacy by design." (2009).

[16] Willis, L.E., 2014. Why not privacy by default? Berkeley Tech. LJ, 29, p.61.

[17] NISO. 2015. Consensus Principles on Users' Digital Privacy in Library, Publisher, and Software-Provider Systems. ISBN 978-1-937522-70-4

[18] 1882 February, The Yale Literary Magazine, Conducted by the Students of Yale College, Volume 47, Number 5, Portfolio: Theory and Practice by Benjamin Brewster, Quote Page 202, New Haven, Connecticut.

[19] QuoteInvestigator, 2018. "In Theory, Theory and Practice are the Same", https://quoteinvestigator.com/2018/04/14/theory/

[20] Desai, Tanvi; Ritchie, Felix; Welpton, Richard (2016). "Five Safes: designing data access for research" (PDF). Bristol Business School Working Papers in Economics: Footnote 1

[21] King, G., & Persily, N. (2019). A new model for industry—Academic partnerships. PS: Political Science & Politics, 53(4), 703–709.

[22] King, G., & Persily, N. (2020). Unprecedented Facebook URLs Dataset now Available for Academic Research through Social Science One; IQSS, Harvard. < https://socialscience.one/blog/unprecedented-facebook-urls-dataset-now-available-research-through-social-science-one>

[23] Kairouz, P., Oh, S. and Viswanath, P., 2015, June. The composition theorem for differential privacy. in International conference on machine learning (pp. 1376-1385). PMLR.

[24] Lécuyer, M., Spahn, R., Vodrahalli, K., Geambasu, R., & Hsu, D. (2019, October). Privacy accounting and quality control in the sage differentially private ML platform. In *Proceedings of the 27th ACM Symposium on Operating Systems Principles* (pp. 181-195).

[25] Soft Error, 2022, Wikipedia. https://en.wikipedia.org/wiki/Soft_error

[26] Brent, R. (1973) Algorithms for Minimization without Derivatives. Englewood Cliffs N.J.: Prentice-Hall

[27] Vadhan, Salil P. "Pseudorandomness." Foundations and Trends® in Theoretical Computer Science 7, no. 1–3 (2012): 1-336.

[28] Ryan Rogers, Subbu Subramaniam, Sean Peng, David Durfee, Seunghyun Lee, Santosh Kumar Kancha, Shraddha Sahay, and Parvez Ahammad, Linkedin's audience engagements API: A privacy-preserving data analytics system at scale, arXiv:2002.05839, 2020

[29] Enyinna, Paschal Ikenna. "Radiological risk assessment of cosmic radiation at aviation altitudes (a trip from Houston Intercontinental Airport to Lagos International Airport)." Journal of Medical Physics/Association of Medical Physicists of India 41, no. 3 (2016): 205.

[30] Janko, Richard, trans. 1987. Poetics with Tractatus Coislinianus, Reconstruction of Poetics II and the Fragments of the On Poets. By Aristotle. Cambridge: Hackett. ISBN 0-87220-033-7.

[31] Van Stempvoot, Stephen J. et. al., (2022) Brief Of The Onion As Amicus Curiae In Support Of Petitioner, Novak v. City of Parma, Ohio

[32] Govindavajhala, S., & Appel, A. W. (2003, May). Using memory errors to attack a virtual machine. In 2003 Symposium on Security and Privacy, 2003. (pp. 154-165). IEEE.

[33] Zhou, Li, and Mingsheng Ying. "Differential privacy in quantum computation." In 2017 IEEE 30th Computer Security Foundations Symposium (CSF), pp. 249-262. IEEE, 2017.

[34] Garfinkel, Simson L., and Philip Leclerc. "Randomness concerns when deploying differential privacy." In Proceedings of the 19th Workshop on Privacy in the Electronic Society, pp. 73-86. 2020.

[35] Dwork, C., McSherry, F., Nissim, K., and Smith, A. Calibrating noise to sensitivity in private data analysis. In Theory of cryptography, pp. 265–284. Springer, 2006.

[36] Pejó, B. and Desfontaines, D., 2022. Guide to Differential Privacy Modifications: A Taxonomy of Variants and Extensions. Springer Nature.