Differentially Private Data to Support Best Practices in Agricultural Operations

TODD STEINBRUECK
BECS Technology, Inc., St. Louis, Missouri

Email: todd@becs.com

1. INTRODUCTION

The Internet of Things (IoT) is ushering in an era where significant numbers of devices that perform monitoring and control functions (e.g., process control, manufacturing, etc.) are connected via wired or wireless networks. Modern agriculture is a leader in experiencing this transformation, with ubiquitous data collection associated with planting, fertilizing, and harvesting of crops as well as with feeding, environmental control, and monitoring of livestock. BECS Technology, Inc., (BECS) is a small business that manufactures monitoring and control equipment for a number of markets (agriculture, aquatics, refrigeration, etc.) and seeks to maximize the societal benefits of the data made available by this connectivity. However, the current data sharing models we employ (like most companies) limits each individual organization (farmer) to seeing only the data logs of equipment that it directly owns. This data model significantly impedes the potential that could be realized.

The benefits that would accrue from aggregating data across farms are substantial. For instance, this can lead to better feed conversion and diminished usage of feed, water, and electricity. We can use machine learning techniques to improve yields as well as catch health-related issues earlier. While these opportunities can and do provide real benefit to farmers, there are challenges related to security, privacy, and data communication that must be overcome. Both Kumar and Patel [62] and Vasilomanolakis et al. [93] describe these challenges as being pervasive across all the IoT.

We desire that the farmer/owners of IoT-based data share those data with the broader community (including their peers) for the benefit of all, but this will only happen if we can assure owners that their data will be kept private. This proposal seeks to explore the feasibility of cross-organizational data sharing by addressing the challenge of data privacy. Additionally, we explore incentives for farmers to participate as well as techniques for communicating aggregated data to farmers to encourage better decision-making. These latter two items are of particular importance to small and mid-size farms, which are extremely unlikely to have a data analyst on staff.

Some high-profile examples [4, 5] have illustrated the difficulty inherent in maintaining the privacy of individuals whose data are shared, even when anonymized. Differential privacy [19, 20] provides a formal framework for ensuring bounds on data leaks about individuals; however, a number of questions remain about its use in practical settings.

What we propose to investigate in this SBIR project is how to transition the theory of differential privacy into useful practice on agricultural data, how to incentivize farmers to participate, and how to communicate such aggregated data back to the farmers. Differential privacy provides strong guarantees that no incremental harm comes to an individual (in our case, a farmer) if they choose to share their private data with a common database maintained by a trusted curator. Statistical queries to that database have added noise included in query responses to preserve privacy. This added noise introduces uncertainty to the data, making effective communication a primary challenge.

The privacy theory is robust and quite strong. What remains is transition into practice, and there are a number of interesting questions that we intend to address as part of this investigation.

- Differential privacy's theory depends upon appropriate tuning of a number of parameters (e.g., ϵ, δ). How the privacy budget impacts utility is not well understood. We will investigate this relationship empirically in the context of agricultural data sets.
- How do we provide appropriate incentives to encourage farmers to provide their data for the common good?

- How do we communicate to users both the query responses and the inherent associated uncertainty due to differential privacy?
- How do we combine the use of differential privacy with alternative approaches [16], such as k-anonymity [84, 91] and l-diversity [69]?

BECS will serve as the trusted curator of the database (it is already serving that role, providing segregated access to individual farmers of their own data). We will investigate how previous privacy-preserving machine learning experience [1, 89] translates into the space of agricultural data, what incentives are appropriate and effective [44], and how visualization can be used by laypersons [74] in the context of anonymized data.

This proposal is responsive to topic area 8.3 Animal Production and Protection as articulated in the FY 2018 Request for Applications from NIFA. Specifically, success in the proposed research will address two of the research priorities of the USDA within this topic area:

- Priority 1. Improve production efficiency
- Priority 3. Improve animal health and well-being

The availability of shared data can be a strong driver to benefit both of the above priorities.

In addition, while we are submitting to topic area 8.3, we see this proposal as also responsive to topic area 8.12 Small and Mid-Size Farms. While larger commercial organizations often have access to extensive data sets of their own, it is the smaller farms that stand to benefit the most from aggregating data across farms.

The commercial opportunities are significant. Currently, farmers pay a monthly fee for a data service provided by BECS, called Feed-Link, in which each farm's equipment makes periodic contact with BECS's servers (in the cloud). All the data collected during that period by the equipment is uploaded to a persistent database (also maintained in the cloud). Farmers then have access to data collected from their equipment in a number of forms, an illustration of the Feed-Link dashboard is described in Section 2.1.

At the simplest level, success in this research will enable farmers to not only have access to the data they already own, but it will also provide access to data from a much larger collection of farms (all the while, preserving the privacy of the farmers that have voluntarily provided those data). This can be an important commercial opportunity for BECS in two ways: (1) data that are more valuable to the farmers will likely support a higher price point for the monthly data service, and (2) increased data value will also increase the number of farmers likely to subscribe.

The successful completion of this research is important commercially on a much broader scale as well. While the initial implementation of the system will target equipment manufactured by BECS, the technology is potentially useful in a much broader setting. One can envision a stand-alone data marketplace in which the aggregation of mutually contributed data that are then shared (in a privacy preserving way) with the participants to be very attractive to anyone who can benefit materially from the collected data set.

2. BACKGROUND AND RELATED WORK

This section will first describe background information on BECS equipment and then the general state-of-the-art in differential privacy.

2.1. Agricultural IoT Data

BECS Technology's equipment are fairly typical devices in the Internet of Things (IoT). The devices monitor various aspects of animal husbandry: barn temperature, feed stocks, feed consumption, water consumption, etc. In addition, our equipment interfaces with equipment from other manufacturers for ventilation control, etc. Based on this information, the various controllers take actions (starting/stopping feed delivery augers, starting/stopping ventilation fans, etc.) to maintain the barn environment at the proper levels and ensure the animals are properly fed. Alarm conditions trigger notifications to service personnel. Sensor values and actions are logged, and these logs are frequently used when diagnosing the causes of alarms or other anomalous events. Remote access to all of the above information is clearly to the benefit of the animal owners/farmers.

While the notion of IoT might be new, the fundamental capability to access controller information remotely is not. BECS Technology's controllers have supported remote communications for more than 2 decades. Early controllers used modems attached to the telephone network (an option still available for those that need it), today controllers support TCP/IP connectivity via the Internet.

Remote capabilities include viewing of current status, downloading of data logs, and configuration of the equipment. Figure 1 shows a screenshot of the Feed-Link dashboard for a specific farm with 3 barns (organized into 2 distinct sides, North and South) in which the 12 grain bins have been instrumented. Note that, as described in Section 4, BECS Technology's equipment is sold, private-label, under the AP and Cumberland brands of GSI, a global agriculture company. The banner near the top indicates we are viewing data from the "Smith and Jones" farm from mid-July to mid-August.

There are links on the middle portion of the screen that drill down into more detailed status reports, notifications of anomolies, and options to download the logs in a spreadsheet-compatible form. This is also where the user indicates the range of data he/she wishes to view.

Each plot along the bottom shows bin weight over a one month timeframe for each bin and also totals for each side of each barn. Vertical jumps in the graphs represent feed deliveries, and the linear downward slope shows feed consumption. Nominally, bins are paired for each side of the barn, with one bin operational at a time, so there should also be portions of the plots that are horizontal, indicating that bin is not providing feed.

Figure 2 illustrates data logs collected over a month, showing the clear correlation between low feed consumption and high temperatures on two occasions. This is an indication of the kinds of things that can be learned from data collected for a single barn.

While the two figures show images from a desktop PC screen, modern remote communications capability is also supported via apps that run on smartphones and tablets.

In addition to diagnosing the root causes of issues in the barn the historical logs also enable the tracking of parameter changes by operators as well as support the demonstration and documentation of regulatory compliance. Using Feed-Link, these data logs are collected automatically and the information retained in the cloud for easy access by the owners/operators of the equipment (the farmer, in most instances).

It should be clear that preserving privacy without concurrently ensuring security would be a completely unacceptable state of affairs. The security of our systems is state-of-the-art [12, 13], with special attention given to ease-of-use considerations, as there is ample evidence that security measures that are difficult to implement are frequently circumvented by users [26, 42, 88].

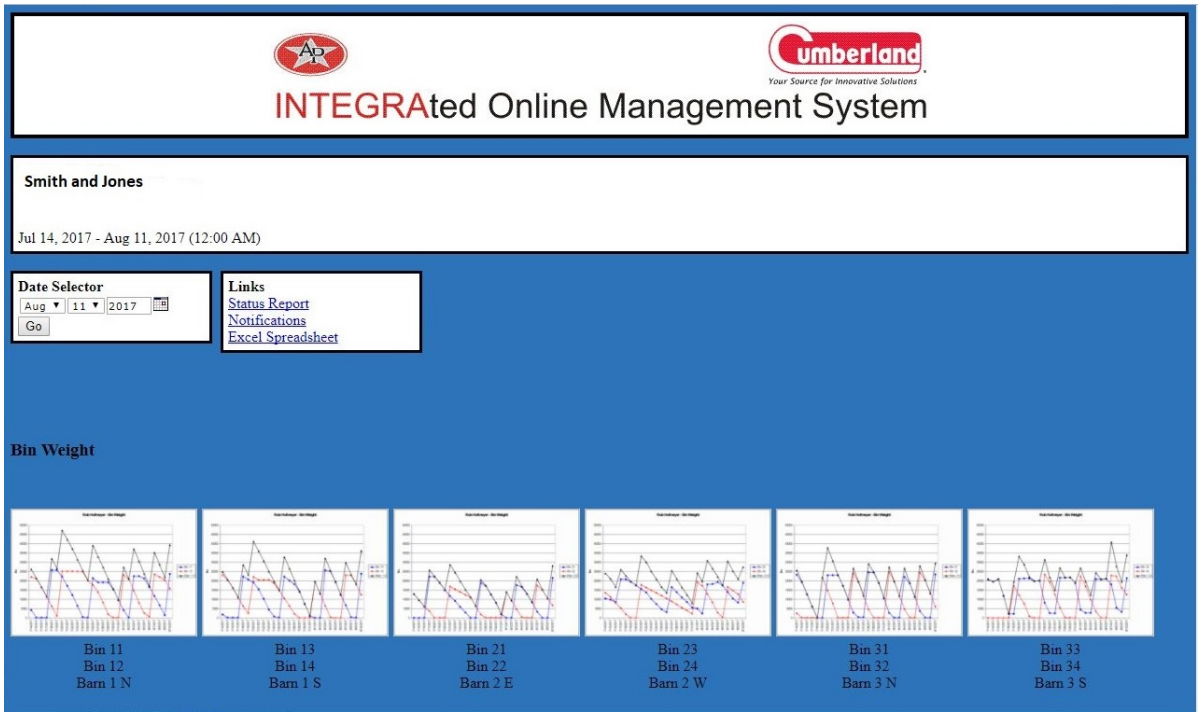


Figure 1: Dashboard display of Feed-Link system.

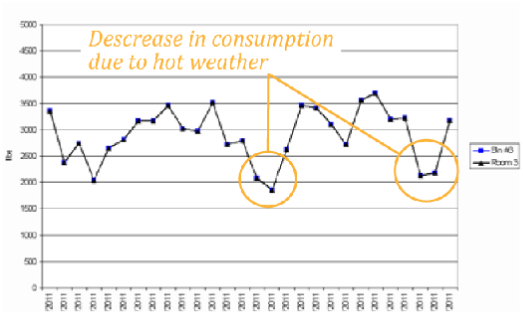
Currently, the data access model is such that farmers have access to data from their own equipment, but there is no data sharing. Because individual farms each typically have a limited number of barns (between 1 and 4 would be typical at an individual facility), there is limited data available for learning to take place. This is true whether the learning is happening in an automated way (e.g., using modern machine learning techniques) or manually (using visualization tools). Each of these approaches will benefit from broader experiential coverage. Changing this current state of affairs is the specific purpose of this proposed research project.

2.2. Privacy Theory

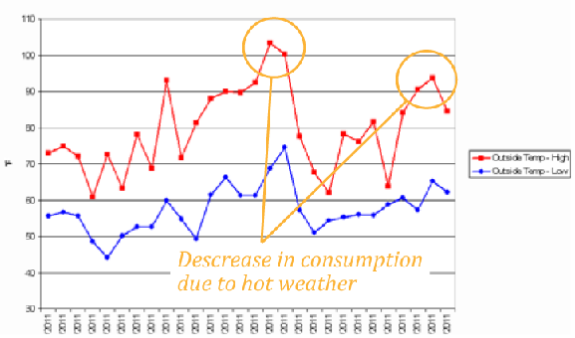
Our goal is to make aggregated data available while preserving the privacy of individual data owners (farmers). Differential privacy provides strong theoretical guarantees in this area. Dwork and Roth [19, 20] provided the seminal initial work in this area, and more recently Nguyen et al. [73] and Wang et al. [94] have reviewed the differential privacy literature as it has been applied to various practical circumstances. A number of groups have contributed to differentially private hierarchical data (referred to as histograms) [40, 96]. Most relevant to our interest is the work of Rastogi and Nath [82] and Fan and Xiong [22, 23], who describe approaches to time series data.

Differential privacy is far from the only way to tackle the problem. Others have described approaches such as k-anonymity [84, 91], l-diversity [69], other techniques aimed at hierarchical data sets [65], and some have asserted that multiple, combined approaches are superior to any individual technique [16]. TIPPERS [71] is an experimental infrastructure, built on a Honeywell building management system, that supports privacy research in the IoT space.

The core notion of differential privacy, informally stated, is that whether or not a single individual chooses to share his/her data to be a part of the collected data set does not impact the conclusions one



(a) Daily feed consumption.



(b) Daily high and low barn temperature.

Figure 2: Plot of data logs.

draws from the data set. This is typically accomplished by perturbing the results of queries against the data set by some random amount. Moving towards more formality [20], a randomized algorithm $\mathcal{M}$ is (ϵ, δ)-*differentially private* if for all $\mathcal{S} \subseteq \text{Range}(\mathcal{M})$ and datasets x and y differing in at most one record:

$$\Pr[\mathcal{M}(x) \in \mathcal{S}] \leq e^{\epsilon} \cdot \Pr[\mathcal{M}(y) \in \mathcal{S}] + \delta \tag{1}$$

where the probability is over the randomness of the algorithm $\mathcal{M}$. The parameter ϵ is often called the *privacy budget* [70]; (ϵ, δ)-differential privacy ensures that for all adjacent x and y, the privacy loss (as defined in [20]) will be bounded by ϵ with probability at least $1 - \delta$. Operationally, it is possible to achieve differential privacy by adding i.i.d. noise to query results.

Our circumstance is one in which time series data is to be queried, and we wish to preserve the privacy of the data owners (rather than that of an individual record). Fan and Xiong [22, 23] describe a mechanism in which the time series are first sampled, perturbed (by adding i.i.d. noise), and then the released series is composed using prediction (for the non-sampled points) and prediction-correction (for the sampled points). It is this approach that we will initially explore.

3. RESEARCH – PHASE I

The goal of the overall system is to support exploration of aggregated data. To achieve this goal, we will leverage our current IoT infrastructure to implement privacy-preserving data aggregation techniques from the differential privacy literature. The data processing, computation, and aggregation occur in the back-end where the data reside. The processed data sent to the front-end visualization system will contain aggregated data for clients and stakeholders to interact with.

During Phase I, our focus will be on addressing the following three research questions.

R1: Can we effectively reduce differential privacy theory to practice, balancing the conflicting concerns of privacy budget and data utility, in the context of the agricultural marketplace?

R2: What incentives are required for farmers to be willing to share their data?

R3: Can we effectively communicate aggregated time series data, with the inherent uncertainty imposed by differential privacy, to end users?

We anticipate addressing the first question through empirical evaluation, exercising one (or more) privacy-preservation algorithms across agricultural IoT data and assessing the resulting output in the privacy versus utility tradeoff space. We will address the second and third questions by exploring several incentive mechanisms and visualization approaches and assessing their effectiveness through user studies.[1]

3.1. R1: Data Aggregation and Privacy Preservation

We are interested in assessing our ability to reduce differential privacy theory into effective practice, and our approach to achieving that goal is empirical evaluation. That implies collecting data, performing experiments, and evaluating the experimental outcomes. We will describe each of these in turn.

Data Sources The Feed-Link system currently provides cloud-based access to agricultural sensor data owned by the individual farmer. BECS is the trusted curator of that data, responsible for its security, integrity, and availability. To provide experimental data for the proposed privacy preservation research, in consultation with GSI (our private-label partners), we will query current customers and ask permission to use historical data from their farms. There are a number of customers that regularly assist us in evaluating new products, etc., at their farm, such that we anticipate no difficulty in achieving data access from real-world installations.

Experiments Our initial empirical investigation will focus on the approaches described by Fan and Xiong [22, 23], utilizing their FAST (Filtering and Adaptive Sampling for differentially private Time-series monitoring) framework. Individual experiments will have as inputs:

- an example aggregated time series data set,
- settings for the privacy budget (i.e., values of ϵ and δ),
- parameter values for FAST (i.e., see Table 1 of [23]), and
- prediction/correction filter for FAST (Kalman vs. Monte Carlo)

and will provide an output time series that is differentially private. Using any number of techniques (e.g., relative error, correlation analysis), we can then compare the input (non-private) time series to the output time series. We will use 2^k factorial experimental design [56] to reduce the overall empirical search space to manageable levels, adding experiments as needed to fill in portions of the space that appear interesting.

Of course, the real world is never as simple as implied by the previous paragraph. E.g., $\delta = 0$ in the published version of FAST. Next, we describe our approach to several of the complications that we must plan to address. In the time series formulation of Fan and Xiong, they only consider a single series, while in our case we have multiple series measurements from a single farm. A typical system will monitor barn temperature, feed consumption, feed stocks, and water usage. Optional additional measurements might include air pressure, wind speed, wind direction, carbon dioxide levels, and many others. As such, the sensitivity of each aggregate query needs to be analyzed for each measure in order to hide the contribution of each farm.

It would be completely unreasonable to consider these separate measurements to be independent. In fact, there is ample reason to believe they are strongly correlated. For example, both feed consumption

[1] For all user studies, we will seek the appropriate human subjects approvals through the IRB at Washington Univ. in St. Louis.

and water consumption are strongly corrrelated with the age of the animals. As such, we will need to extend the approach of Fan and Xiong to support multiple time series. We will start by using the same reasoning as Fan and Xiong. For n series, we will allocate ϵ/n of the privacy budget to each series.

If this initial approach is too constraining, we will explore alternative partitionings of the privacy budget, in particular, approaches that dynamically form the partitioning guided by the tradeoffs between privacy budget and utility (which are likely to be different for different signals).

Another issue we must deal with is the fact that from any individual data series, it will be relatively straightforward to discern when animals are delivered to the farm, and when they are shipped out. If one connects this knowledge to alternate sources of information, individual farms might be readily identified. One approach we will investigate to deal with this issue is potentially to not publicly release absolute dates, but rather communicate the time series indexed by the age of the animals in the barn.

While our initial focus will be on the techniques described by Fan and Xiong, an alternative approach to time series data is presented by Rastogi and Nath [82]. In their approach, the time domain data are transformed into the frequency domain, appropriate additive noise is inserted to ensure differential privacy, and the inverse transform returns the data to the time domain. This technique requires that the entire data set be available prior to release; however, that is not an insurmountable obstacle in our circumstance.

Another practical consideration that we much address is the fact that, in addition to traditional time series data, our data sets also include event data (e.g., alarm conditions, feed deliveries, control parameter changes). For those that can be effectively encoded as binary variables (e.g., alarm status), we will start with that encoding. One possibility for discrete events is to encode their inter-event time and ensure that it is differentially private. Another is to use the w-event privacy notions introduced by Kellaris et al. [58]. In the most pessimistic case, we might need to suppress some raw data, if we cannot ensure its disclosure maintains privacy. In such a situation, we clearly need to quantitatively assess the utility implications of this choice. In any event, the Phase I effort will not consider event data, which will be investigated during Phase II.

To this point in the discussion, we have maintained strict compliance with differential privacy, looking to discover whether or not we can achieve sufficient utility at an acceptable privacy budget. If this is the case, we have succeeded in our goal. If this is not the case, all is not lost.

Clifton and Tassa [16] argue that other privacy preserving mechanisms, while not as strong as differential privacy from a theoretical perspective, are still quite valuable in practice. Just as we cannot prove perfect security, and therefore rely on multiple tiers of security apparatus, we can also exploit a similar approach to data privacy. We will investigate a multi-tiered approach to data privacy that has differential privacy at its core, but leverages the additional concepts of *suppression* and *generalization* which are commonly used means to transform data to comply with k-anonymity and/or l-diversity criteria [69]. Specifically, in our setting suppression would entail removing time series contributions from a (small) subset of organizations, whereas generalization would determine the level of discretization of time in the time series. We will use these techniques to preprocess the dataset before applying the algorithms for making the resulting data differentially private. The key intuition for this approach is that suppression would serve to reduce global sensitivity of the queries by removing organizations that are particularly identifiable in the dataset (e.g., those which are highly unusual). Similarly, using a coarser time series data would reduce the amount of noise necessary to make it differentially private, at the cost of utility loss of removing fine-grained information. We conjecture that the combined approaches provide us with sufficient leverage to allow for optimal balancing between utility and privacy.

Evaluating Outcomes For most of our experimental results, the outcomes of an experiment will be in the form of a multi-dimensional ROC curve (more precisely, a multi-dimensional regression ROC curve [24, 41, 72]), illustrating the tradeoffs between privacy budget (shown on one axis) and uncertainty (shown on the other axis). If we can effectively fix the parameter δ, as has been suggested [20], that leaves ϵ as the sole parameter describing privacy budget (at least in the case where we are only using differential privacy as the privacy mechanism), so we are down to one dimension there.

Similarly, if we can distill the uncertainty to a summary statistic (e.g., rms error or some other norm), uncertainty can also be reduced to a single dimension, and now we can actually plot a traditional ROC curve, showing the tradeoff between privacy budget and uncertainty. Clearly, this distillation down to a traditional ROC curve won't happen for every case, but we will exploit it whenever we can.

Given the existence of an ROC curve that realistically communicates the tradeoffs implied, what still remains is the judgment as to whether or not any achievable points in the tradeoff space are acceptable (i.e., effectively meet the needs of end users). Understanding this is key to commercial feasibility. We will evaluate this by collaborating with GSI to identify a set of end users to give us feedback on the tradeoffs.

3.2. R2: Formulating Appropriate Incentives

The goal of this proposal is to transition the theory of differential privacy into useful practice on agricultural data. To achieve this goal, we first need to obtain data from farmers. In this portion of the project, we explore how to design appropriate incentives to encourage users[2] with privacy concerns to provide their data. In particular, we propose to first empirically measure users' costs for releasing their data under different *privacy budgets* (i.e., the values of ϵ in the theory of differential privacy) and to design incentive mechanisms to obtain data from users.

Measuring Privacy Costs Differential privacy provides a formal framework in exploring the tradeoff between aggregation accuracy and privacy guarantees (in terms of privacy budgets). Intuitively, a mechanism with a stronger privacy guarantee (a smaller privacy budget) should provide stronger incentives for users to contribute their data. However, in practice, how the privacy budget influences users' motivation to contribute their data is not well understood.

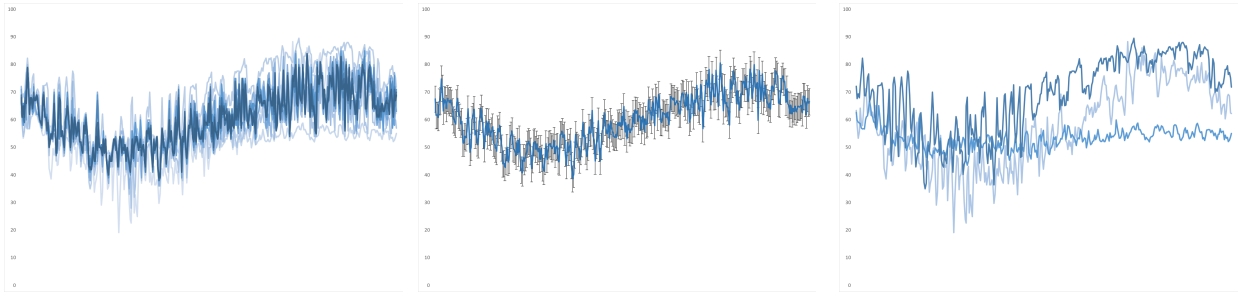
In this phase of the project, we will design and conduct experiments to measure users' privacy costs (i.e., how much they value their privacy) as a function of privacy budgets. As a first step, we will run a pilot study and collect data from online users via Amazon's Mechanical Turk [2]. In this pilot study, we plan to design simulated scenarios and measure how much money we need to offer to users in order for them to release their data under different private budgets. The collected information will reveal how sensitive users are about the privacy budget and provide us insights on the design of incentive mechanisms. We will conduct similar experiments with farmers about releasing their agricultural data in Phase II.

Designing Incentives With the information on users' private costs, we next will explore the design of incentives to motivate users to release their data.

In Phase I, we will explore the design of monetary incentives.[3] There have been several prior works on the study of purchasing data from users with privacy concerns [15, 27, 66, 67, 95]. However, most of

[2]Our proposed research can be generalized to scenarios of purchasing data from users with privacy concerns. Therefore, we use *users* instead of *farmers* when the discussion is generalizable.

[3]We will explore non-monetary incentives in Phase II. As an example, since farmers could learn how to optimize their yields from the aggregated data, the access to the aggregated data itself could serve as an incentive. We plan to explore how



(a) Line ensembles (b) Line with error bars (c) Hypothetical outcome plots

Figure 3: We will evaluate visualization techniques for representing uncertainty with time series data. The existed methods include: (a) Line ensembles, (b) error bars, and (c) hypothetical outcome plots [47] (an animation that randomly draws from the distribution of possible outcomes.)

these works assume little knowledge of users' privacy costs. In this research, we plan to explore whether we can design incentives based on our empirical measurements of users' privacy costs. For example, consider the posted-price mechanism, in which we post a fixed price to all farmers. Farmers will provide their data to us if and only if our posted price is higher than their costs for releasing the data. Even in this simple mechanism, we need to address several tradeoffs: when the privacy guarantee is strong (i.e., smaller privacy budget), we can obtain more accurate aggregated data, however, we need to pay more to farmers to obtain their data since their private costs are higher.

We plan to explore these tradeoffs in the posted-price mechanism and in other incentive mechanisms (such as auctions) in the literature. We will combine theoretical analysis and simulations to evaluate our results. In particular, we will theoretically prove the incentive properties of the proposed mechanisms (i.e., whether farmers will truthfully report their data and/or their privacy costs) and use simulated data to demonstrate the tradeoffs between privacy guarantee, data accuracy, and the total amount of money spent to acquire the data.

3.3. R3: Client Visualization

Communicating uncertainty is one of the greatest challenges we face today. It impacts many data-driven communities such as climate science [6, 79, 80, 90], business [34, 83] and medicine [36, 37, 59, 68, 78, 81, 90]. Visualization has been widely successful in helping people explore, reason, and make judgments with data, and prior work suggests that it may be key for improving comprehension of probability and uncertainty [17, 6, 79, 85, 86, 90]. Researchers have proposed several designs including line ensembles [80, 86], error bars [17], gradient plots [17, 37, 48, 92] and violin plots [17, 43, 57]. More recently, Hullman et al. introduced the hypothetical outcome plot which is an animated chart that randomly draws from the distribution of possible outcomes [47].

One advantage of visualization is that it can make complex or abstract concepts easier to grasp by making them visible. In theory, an effective visual representation can make traditionally difficult problems more concrete and easier to understand. In practice, however, nuanced contextual factors heavily impact the effectiveness of visualizations, making the right representation difficult to establish. While many designs exist, it is not clear whether these methods effective communicate uncertainty and support

to control the amount of information on the aggregated data farmers can access to encourage them to contribute more (and better) data.

decision-making in real-world tasks. Many open questions remain. For instance: *Which visualizations best communicates data with uncertainty*? and *How do visualizations impact decision-making?*

In this phase of the project, we will explore and evaluate methods for communicating uncertainty to non-experts. To identify the best techniques, we will rigorously examine existing methods. These include using line ensembles (Figure 3a), using standard error bars (Figure 3b), using animation (Figure 3c), as well as other techniques that leverage visual variables (e.g., scaling the size of visual elements, using color to encode the degree of uncertainty, and using transparency or blurring). We plan first to test the designs with a diverse population of study participants via Amazon's Mechanical Turk. Once we have some candidate visualizations, we will expand these results by testing real aggregated data with clients in their natural work environments. Dr. Ottley and her team have extensive experience in evaluating systems. Here we outline the steps we will take to evaluate the different designs.

Comparing Designs Our goal here is to determine the most effective visualization design for conveying meaningful information and for facilitating decision-making. We will conduct experiments via Amazon's Mechanical Turk using three different evaluation approaches.

(1) We will measure speed and accuracy as participants complete standard search and retrieval tasks.

(2) Participants will perform decision-making tasks, and we will investigate how each visualization impacts the decisions made.

(3) During our experiment, we will conduct additional post-tests using adaptations of standard questionnaires. For instance, we will use the NASA-TLX for subjective measures of workload [39], and the System Usability Scale for measuring perceived usability [3].

Application to Visualization Theory Our proposed work of exploring techniques for communicating uncertainty has applications beyond our problem statement. Visualizing uncertainty is a central challenge in the Information Visualization community with many applications including weather, health, business, and government. Evaluating and potentially developing novel visualization techniques for representing uncertainty will be a signification contribution to these application areas.

3.4. Judging Success

Ultimately, this project will be successful if we can effectively protect the privacy of individual data contributors (i.e., a small privacy budget) concurrently with releasing aggregate data that has high utility (i.e., low uncertainty). While theory tells us we cannot be perfect at one without sacrificing the other, the intention is to determine how close to that ultimate scenario it is to possible to be.

At the completion of Phase I, a judgment is required to determine if further effort towards the above ultimate goal is warranted. Because that judgment is both a commercial and a technical judgment, we plan to collaborate with GSI to identify individual potential customers for the assessment.

One source of information that we will pursue is the International Poultry Expo, a trade show sponsored by the U.S. Poultry & Egg Association, which is held annually in late January. This meeting is extremely well attended (tens of thousands participate) and, as such, will serve as fertile ground for helping us assess the degree to which we have a technical solution that: (1) provides reasonable privacy/utility tradeoffs, (2) effectively communicates those tradeoff choices to farmers, and (3) can be offered in a way that farmers are likely to participate.

The meeting is timely, in that it happens towards the end of the Phase I schedule period. During the meeting, we will solicit participants to assist us in evaluating our technical solutions. This is the primary method by which we will judge whether or not we have been successful in the Phase I effort.

4. RESEARCH TEAM AND PLAN

Here, we describe the research team that will perform the investigations we propose, and then follow that with the our collaboration plan and timeline for Phase I along with our current thoughts concerning the Phase II effort.

4.1. Team

Organized in 1991, BECS is privately held and employs approximately 100 people. It has outgrown its current manufacturing facility of 24,000 sq. ft. so is in the process of renovating and moving into a new 42,000 sq. ft. facility. Its intellectual property is protected with 12 issued patents and 2 patents pending.

The design engineering team at BECS is comprised of 10 individuals, of which Todd Steinbrueck is the software team lead (who will serve as PI on the SBIR grant) and Roger Chamberlain is the VP Engineering (who will serve as senior personnel on the SBIR grant).[4] Additional members of the design engineering team will contribute to the effort as needed.

This team has extensive experience in control equipment design and implementation, including custom implementation of both the hardware and the software. Presently, BECS designs and manufactures equipment for animal husbandry (the focus of this proposal), e.g., grain bin content monitoring, feed auger controls, environmental monitoring and control of animal houses; maintains proper water chemistry in aquatics systems; monitors and controls refrigeration in commercial settings, e.g., valve controllers that adapt automatically to different refrigerants; and delivers CATV signals, e.g., RF amplifiers, line extenders.

BECS provides equipment to the agriculture industry under a private-label arrangement with Grain Systems, Inc. (GSI), a subsidiary of AGCO Corporation, which is a publicly traded company (NYSE:AGCO) with net sales of $7.4 billion in 2016. GSI provides BECS equipment (and equipment manufactured by itself and others) to the market through both its Cumberland (poultry) and AP (swine) divisions.

As the vision of IoT becomes a reality, equipment providers generally need to evolve so that they partner with end users (farmers in our case) to become a service company as well as a manufacturing company. BECS and GSI both aspire to this goal. Quoting from AGCO's website,[5] "Grain, swine, and poultry producers need more than equipment, they need full-scale solutions that boost overall performance and productivity."

As described in Section 2, BECS has supported remote connectivity to its equipment for years. What is new is that rather than simply providing access to data (and expecting end users to make use of it on their own), BECS is increasingly finding ways to help end users by being actively involved in data analysis. While initial connectivity solutions merely supported pull semantics (i.e., users had to manually connect to the equipment to see what was going on), current systems actively contact users

[4] Roger Chamberlain also is Professor of Computer Science and Engineering at Washington University in St. Louis. While bringing over 25 years of research experience to bear on the problem at hand, Dr. Chamberlain will be serving in his capacity as VP Engineering at BECS and will maintain strict compliance with the conflict-of-commitment rules at the the university.

[5] `www.agcocorp.com/brands/gsi.html`

when conditions warrant (e.g., alarm conditions, feed stocks depleted), pushing information to farmers on demand. Through the combination of functionality within the control equipment and data services provided via the cloud, the vision is one in which animal husbandry is simpler to execute, more efficient, less resource intensive (both in terms of human resources and feed and water resources), and results in healthier animals.

BECS has received government funding on two separate occasions (see Section 5 for details), both of which were SBIR/STTR grants through the National Institutes of Health. We currently have a pending application with the National Science Foundation that focuses on similar issues as this proposal but is targeted at the aquatics industry.

We will collaborate on this SBIR project with Dr. Alvitta Ottley, Asst. Professor of Computer Science and Engineering at Washington University in St. Louis, Dr. Chien-Ju Ho, Asst. Professor of Computer Science and Engineering at Washington University in St. Louis, and Dr. Yevgeniy Vorobeychik, Assoc. Professor of Computer Science and Engineering at Washington University in St. Louis. In this proposed agenda, Dr. Alvitta Ottley will lead the design of the visualization front-end of the system that will allow clients to interact with the data. Dr. Ottley has built and evaluated a range of visualization tools and designs [7, 35, 76, 77]. Relevant to the proposed work, her prior research has focused on designing visualizations to decision support [35, 74, 75] for non-experts. Her work has also made significant advancement toward evaluating the effectiveness of visualization designs [77, 97], and understanding how users interactions with visualization tools [7, 76].

Dr. Ho will investigate the design of incentives to motivate farmers to contribute their data. Dr. Ho's research has focused on the learning and incentive problems in human-in-the-loop systems. In particular, he has worked on designing reputation systems and monetary contracts to encourage online users to contribute high-quality data [45, 46] and measuring how users respond to incentives in real-world applications [44].

Dr. Vorobeychik will investigate complementary notions of privacy, such as a combination of differential privacy and k-anonymity/l-diversity criteria. He has extensive prior research on privacy-preserving data sharing, focusing largely on medical data [**?**, **?**, **?**, **?**].

We will consult on the project with Dr. Liyue Fan, Asst. Professor of Information Security and Digital Forensics, University at Albany (SUNY). Dr. Fan's research interests are in data privacy, spatiotemporal data analysis, and database applications. Her research is one of two currently viable approaches to differentially private time series data, and is the one we plan to focus on in this project. Dr. Fan will commit 2 weeks of effort towards the project.

Together the proposal team has the necessary knowledge and experience to undertake the proposed work. More specific information regarding coordination and collaboration plans is provided below.

4.2. Research Plan

The primary participants in the project all reside in St. Louis, Missouri, so face-to-face meetings have a relatively low overhead. We will coordinate the project through periodic (e.g., biweekly) meetings that alternate locations (one meeting at BECS Technology's facility, and two weeks later another meeting on campus at Washington Univ. in St. Louis).

Dr. Fan is physically remote (she is in Albany, NY), so her primary method of interaction will be via electronic communication. We will use video-conferencing fairly extensively when face-to-face discussions are warranted. In addition, the Dept. of Computer Science and Engineering at Washington

University in St. Louis will invite Dr. Fan to present a colloquium on campus at or very near the beginning of the project. This will enable all of the parties to have a true face-to-face interaction without any travel expense charged to the grant.

A formal agreement will need to be put in place between the company and Washington U.; however, the two organizations have collaborated in the past (on a previous STTR grant through the National Institutes of Health). The terms of that previous agreement will form the basis of the required new agreement, and we anticipate no difficulties in finalizing the agreement itself.

The proposed research is suited for an eight-month research agenda. The schedule is shown in Figure 4. The various research components can initally be independently investigated, primarily in parallel by the two organizations, coming together in the assessments to be performed at the end of Phase I.

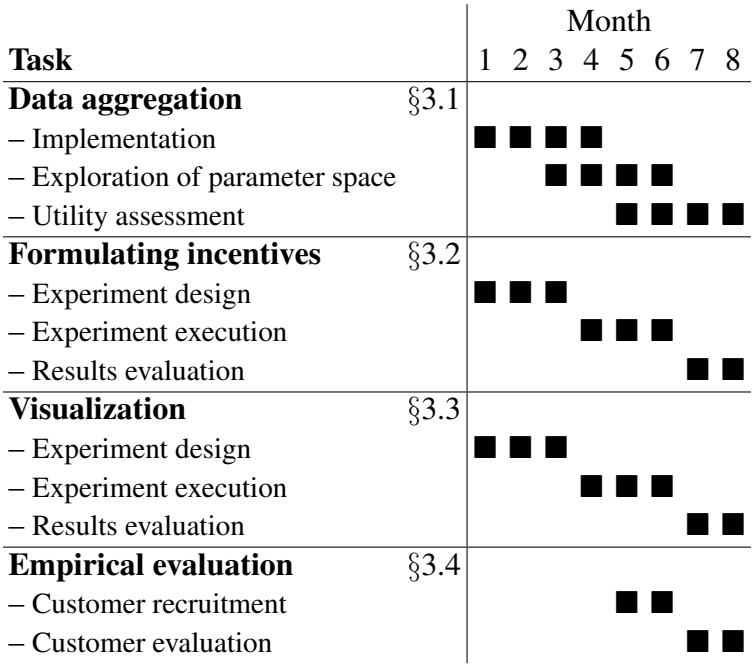


Figure 4: Project schedule, eight month duration.

The results of the assessments will inform the activities to be pursued during Phase II. If the result of Phase I is data that are appropriately anonymized, one can consider the focus of Phase II to be the determination of what it is we can effectively learn from the anonymized data.

The learning from data will come in two forms:

- **Machine learning** – we will investigate utilizing techniques described in the literature [1, 25, 89] for performing machine learning on differentially private data sets. We will apply techniques such as these to the problem of barn control, with the goal of diminishing feed consumption, providing tighter control, and improving feed conversion rates.
- **Human learning** – we will explore ways in which we can effectively communicate to laypersons an intuitive notion of "privacy budget," and how to visualize risk, which has been studied in the management and financial space [21, 87], but we are unaware of any work concerning privacy risk.

In addition to exploring what can be learned from data in a privacy preserving manner, there are a number of followup investigations that will be initiated during Phase I but will require additional effort during Phase II. First, our initial efforts will be focused on time-series data; however, event data are

also present in our data sets and we will need to investigate techniques for including them in properly anonymized form.

Second, both our incentive and visualization investigations during Phase I will be restricted to participants we recruit via Amazon's Mechanical Turk. This gives us flexible and inexpensive access to willing participants, with a broad set of backgrounds. However, we are particularly interested in the impact of these investigations on our target audience, farmers. The Phase II effort will include targeted studies to address this issue.

Third, the incentive mechanisms during Phase I will be limited to monetary incentives. As we discover what can be learned from the aggregated data, it is entirely possible that simply access to this information is a sufficient incentive for farmers to be willing to participate. This will also be a part of our Phase II investigation.

5. RESULTS OF PRIOR SBIR/STTR SUPPORTED RESEARCH

All information in this section about other companies (Hearing Emulations, LLC, and Exegy, Inc.), including quantities of both external and internal investments, is proprietary to those firms.

Hearing Aids Based on Models of Cochlear Compression – SBIR Phase I (1999), $100,000, and Phase II (2001), $546,284. PI: Julius Goldstein, Senior Personnel: Roger D. Chamberlain.

This research demonstrated improvements in the understandability of speech in noisy environments for hearing-impaired individuals. Commercialization of this work resulted in the formulation of a new company, Hearing Emulations, LLC, which subsequently received private investment of approximately $2 million. BECS's interest was sold to the investors, so we no longer have visibility into the company. Publications resulting from this work include [14, 31, 32, 33]. Patents resulting from this work include [28, 29, 30].

Fast Biosequence Annotation via Reconfigurable Hardware – STTR Phase I (2004), $307,508, and Phase II (2007), $510,614. PI: Jeremy Buhler, Senior Personnel: Roger D. Chamberlain.

This research resulted in FPGA-based accelerated implementations of several biosequence applications, including both DNA and protein alignment as well as RNA folding. Commercialization of this work resulted in a Memorandum of Understanding (MOU) between Exegy, Inc., and BECS. Exegy has subsequently invested approximately $500,000 in the technology. Under the terms of the MOU, Exegy is not required to disclose sales figures to BECS. Publications resulting from this work include [11, 18, 38, 49, 50, 51, 52, 53, 54, 55, 60, 61, 63, 64], which includes a best paper award [54]. Patents resulting from this work include [8, 9, 10] and one pending application.

6. CONCLUSION

It is clear that privacy of IoT-derived data is currently a challenge, and differential privacy theory has the potential to put privacy practice on a sound theoretical footing. What still remains to be seen is whether or not differential privacy theory is "ready for prime time."

Specifically, can the current state of the theory handle all of the practical considerations that much be resolved for it to be effective in the real world.

- Can we choose appropriate parameter values so as to achieve sufficient privacy and maintain data utility?
- Can we motivate farmers to contribute their data for the common good?
- Can we effectively communicate the anonymized query results to farmers in visual forms that include the inherent uncertainty due to privacy?
- Can we combine differential privacy theory with other privacy mechanisms for a multi-tiered privacy infrastructure?
- While ensuring that cross-organizational data are private, can we exploit those data to improve agricultural practice?

These are the questions we intend to address as part of this SBIR project. Success will enable benefits across many more markets than just agriculture, as virtually all of the IoT space has issues that, if not identical to those we face, are certainly comparable.

REFERENCES

[1] M. Abadi, A. Chu, I. Goodfellow, H. B. McMahan, I. Mironov, K. Talwar, and L. Zhang. Deep learning with differential privacy. In *Proc. of ACM SIGSAC Conf. on Computer and Communications Security*, pages 308–318. ACM, 2016.

[2] Amazon Mechanical Turk, 2107. `http://www.mturk.com`, accessed Oct. 2017.

[3] A. Bangor, P. T. Kortum, and J. T. Miller. An empirical evaluation of the system usability scale. *Int'l Journal of Human-Computer Interaction*, 24(6):574–594, 2008.

[4] M. Barbaro and T. Zeller. A face is exposed for AOL searcher No. 4417749. *The New York Times*, Aug. 9, 2006.

[5] R. Bell and Y. Koren. Lessons from the Netflix prize challenge. *SIGKDD Explor. Newsl.*, 9(2), 2007.

[6] K. Brodlie, R. A. Osorio, and A. Lopes. A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*, pages 81–109. Springer, 2012.

[7] E. T. Brown, A. Ottley, H. Zhao, Q. Lin, R. Souvenir, A. Endert, and R. Chang. Finding Waldo: Learning about users from their interactions. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):1663–1672, 2014.

[8] J. Buhler, R. Chamberlain, M. Franklin, K. Gyang, A. Jacob, P. Krishnamurthy, and J. Lancaster. Method and apparatus for performing similarity searching on a data stream with respect to a query string. U.S. Patent #7,917,299, issued Mar. 29, 2011.

[9] J. Buhler, R. Chamberlain, M. Franklin, K. Gyang, A. Jacob, P. Krishnamurthy, and J. Lancaster. Method and apparatus for performing similarity searching. U.S. Patent #8,515,682, issued Aug. 20, 2013.

[10] J. Buhler, R. Chamberlain, M. Franklin, K. Gyang, A. Jacob, P. Krishnamurthy, and J. Lancaster. Method and apparatus for performing similarity searching. U.S. Patent #9,547,680, issued Jan. 17, 2017.

[11] J. D. Buhler, J. M. Lancaster, A. C. Jacob, and R. D. Chamberlain. Mercury BLASTN: Faster DNA sequence comparison using a streaming hardware architecture. In *Proc. of Reconfigurable Systems Summer Institute*, July 2007.

[12] R. D. Chamberlain, M. Chambers, D. Greenwalt, B. Steinbrueck, and T. Steinbrueck. Layered security and ease of installation for devices on the Internet of Things. In *Proc. of IEEE Int'l Conf. on Internet-of-Things Design and Implementation*, pages 297–300. IEEE, Apr. 2016.

[13] R. D. Chamberlain, M. Chambers, D. Greenwalt, B. Steinbrueck, and T. Steinbrueck. Devices can be Secure and Easy to Install on the Internet of Things. In R. Gravina, C. Palau, M. Manso, A. Liotta, and G. Fortino, editors, *Interconnection, Integration, and Interoperability of IoT Systems*. Springer, 2018.

[14] R. D. Chamberlain, J. L. Goldstein, and D. Ivanovich. Implementation of hearing aid signal processing algorithms on the TI DHP-100 platform. In *Proc. of 37th Asilomar Conf. on Signals, Systems and Computers*, volume 1, pages 404–409, Nov. 2003.

[15] Y. Chen, S. Chong, I. A. Kash, T. Moran, and S. Vadhan. Truthful mechanisms for agents that value privacy. *ACM Trans. Econ. Comput.*, 4(3):13:1–13:30, Mar. 2016.

[16] C. Clifton and T. Tassa. On syntactic anonymity and differential privacy. *Trans. in Data Privacy*, 6:161–183, 2013.

[17] M. Correll and M. Gleicher. Error bars considered harmful: Exploring alternate encodings for mean and error. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2142–2151, 2014.

[18] R. Dor, J. M. Lancaster, M. A. Franklin, J. Buhler, and R. D. Chamberlain. Using queuing the-

ory to model streaming applications. In *Proc. of Symposium on Application Accelerators in High Performance Computing*, July 2010.

[19] C. Dwork. A firm foundation for private data analysis. *Communications of the ACM*, 54(1):86–95, 2011.

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

[21] M. J. Eppler and M. Aeschimann. A systematic framework for risk visualization in risk management and communication. *Risk Management*, 11(2):67–89, 2009.

[22] L. Fan and L. Xiong. Real-time aggregate monitoring with differential privacy. In *Proc. of 21st ACM International Conference on Information and Knowledge Management*, pages 2169–2173, New York, NY, USA, 2012. ACM.

[23] L. Fan and L. Xiong. An adaptive approach to real-time aggregate monitoring with differential privacy. *IEEE Transactions on Knowledge and Data Engineering*, 26(9):2094–2106, Sept. 2014.

[24] T. Fawcett. An introduction to ROC analysis. *Pattern Recognition Letters*, 27(8):861–874, June 2006.

[25] A. Friedman and A. Schuster. Data mining with differential privacy. In *Proc. of 16th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 493–502, New York, NY, USA, 2010. ACM.

[26] D. Gefen and D. W. Straub. The relative importance of perceived ease of use in IS adoption: a study of e-commerce adoption. *Journal of the Association for Information Systems*, 1(1):8, 2000.

[27] A. Ghosh and A. Roth. Selling privacy at auction. In *Proc. of 12th ACM Conference on Electronic Commerce*, pages 199–208, 2011.

[28] J. L. Goldstein. Hearing aids based on models of cochlear compression. U.S. Patent #6,868,163, issued Mar. 15, 2005. Assigned to Hearing Emulations, LLC.

[29] J. L. Goldstein. Hearing aids based on models of cochlear compression using adaptive compression thresholds. U.S. Patent #6,970,570, issued Nov. 29, 2005. Assigned to Hearing Emulations, LLC.

[30] J. L. Goldstein and R. D. Chamberlain. Hearing aids based on models of cochlear compression. European Letters Patent #EP 1 121 834, issued Apr. 2, 2003. Assigned to Hearing Emulations, LLC.

[31] J. L. Goldstein, M. Oz, P. Gilchrist, and M. Valente. Signal processing strategies and clinical outcomes for gain and waveform compression in hearing aids. In *Proc. of 37th Asilomar Conference on Signals, Systems and Computers*, volume 1, pages 391–398, Nov. 2003.

[32] J. L. Goldstein, M. Valente, and R. D. Chamberlain. Acoustic and psychoacoustic benefits of adaptive compression thresholds in hearing aid amplifiers that mimic cochlear function. *J. Acoust. Soc. Am.*, 109:2355(A), 2001.

[33] J. L. Goldstein, M. Valente, R. D. Chamberlain, P. Gilchrist, and D. Ivanovich. Pilot experiments with a simulated hearing aid based on models of cochlear compression. In *Proc. of International Hearing Aid Research Conference*, Aug. 2000.

[34] H. Griethe and H. Schumann. Visualizing uncertainty for improved decision making. In *Proc. of 4th International Conference on Business Informatics Research*, BIR 2005, pages 1–11, 2005.

[35] A. Hakone, L. Harrison, A. Ottley, N. Winters, C. Gutheil, P. K. Han, and R. Chang. PROACT: Iterative design of a patient-centered visualization for effective prostate cancer health risk communication. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):601–610, 2017.

[36] P. K. Han, W. M. Klein, and N. K. Arora. Varieties of uncertainty in health care: a conceptual taxonomy. *Medical Decision Making*, 31(6):828–838, 2011.

[37] P. K. Han, W. M. Klein, T. Lehman, B. Killam, H. Massett, and A. N. Freedman. Communication of

uncertainty regarding individualized cancer risk estimates: effects and influential factors. *Medical Decision Making*, 31(2):354–366, 2011.

[38] B. Harris, A. C. Jacob, J. M. Lancaster, J. Buhler, and R. D. Chamberlain. A banded Smith-Waterman FPGA accelerator for Mercury BLASTP. In *Proc. of 17th International Conference on Field Programmable Logic and Applications*, Aug. 2007.

[39] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology*, 52:139–183, 1988.

[40] M. Hay, V. Rastogi, G. Miklau, and D. Suciu. Boosting the accuracy of differentially private histograms through consistency. *Proc. VLDB Endow.*, 3(1-2):1021–1032, Sept. 2010.

[41] J. Hernandez-Orallo. ROC curves for regression. *Pattern Recognition*, 46(12):3395–3411, Dec. 2013.

[42] M. Hertzum, N. Jørgensen, and M. Nørgaard. Usable security and e-banking: Ease of use vis-a-vis security. *Australasian Journal of Information Systems*, 11(2), 2004.

[43] J. L. Hintze and R. D. Nelson. Violin plots: a box plot-density trace synergism. *The American Statistician*, 52(2):181–184, 1998.

[44] C.-J. Ho, A. Slivkins, S. Suri, and J. W. Vaughan. Incentivizing high quality crowdwork. In *Proc. of 24th International Conference on World Wide Web*, pages 419–429, 2015.

[45] C.-J. Ho, A. Slivkins, and J. W. Vaughan. Adaptive contract design for crowdsourcing markets: Bandit algorithms for repeated principal-agent problems. In *Proc. of Fifteenth ACM Conference on Economics and Computation*, pages 359–376, 2014.

[46] C.-J. Ho, Y. Zhang, J. Vaughan, and M. van der Schaar. Towards social norm design for crowdsourcing markets. In *The 4th Human Computation Workshop (HCOMP)*, 2012.

[47] J. Hullman, P. Resnick, and E. Adar. Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PloS one*, 10(11):e0142444, 2015.

[48] C. H. Jackson. Displaying uncertainty with shading. *The American Statistician*, 62(4):340–347, 2008.

[49] A. Jacob, J. D. Buhler, and R. D. Chamberlain. Accelerating Nussinov RNA secondary structure prediction with systolic arrays on FPGAs. In *Proc. of IEEE International Conference on Application-specific Systems, Architectures and Processors*, pages 191–196, July 2008.

[50] A. Jacob, J. Lancaster, J. Buhler, and R. D. Chamberlain. FPGA-accelerated seed generation in Mercury BLASTP. In *Proc. of 15th IEEE Symposium on Field-Programmable Custom Computing Machines*, pages 97–106, Apr. 2007.

[51] A. Jacob, J. Lancaster, J. Buhler, and R. D. Chamberlain. Preliminary results in accelerating profile HMM search on FPGAs. In *Proc. of 6th IEEE International Workshop on High Performance Computational Biology*, Mar. 2007.

[52] A. Jacob, J. Lancaster, J. Buhler, B. Harris, and R. D. Chamberlain. Mercury BLASTP: Accelerating protein sequence alignment. *ACM Trans. on Reconfigurable Technology and Systems*, 1(2):1–44, June 2008.

[53] A. C. Jacob, J. D. Buhler, and R. D. Chamberlain. Optimal runtime reconfiguration strategies for systolic arrays. In *Proc. of 19th International Conference on Field Programmable Logic and Applications*, pages 162–167, Aug. 2009.

[54] A. C. Jacob, J. D. Buhler, and R. D. Chamberlain. Design of throughput-optimized arrays from recurrence abstractions. In *Proc. of 21st IEEE International Conference on Application-specific Systems, Architectures and Processors*, pages 133–140, July 2010.

[55] A. C. Jacob, J. D. Buhler, and R. D. Chamberlain. Rapid RNA folding: Analysis and acceleration of the Zuker recurrence. In *Proc. of 18th IEEE International Symposium on Field-Programmable*

Custom Computing Machines, pages 87–94, May 2010.

[56] R. Jain. *The Art of Computer System Performance Analysis: Techniques for Experimental Design, Measurement, Simulation and Modeling*. John Wiley & Sons, Inc., New York, 1991.

[57] P. Kampstra. Beanplot: A boxplot alternative for visual comparison of distributions. *Journal of Statistical Software*, 28(1):1–9, 2008.

[58] G. Kellaris, S. Papadopoulos, X. Xiao, and D. Papadias. Differentially private event sequences over infinite streams. *Proc. VLDB Endow.*, 7(12):1155–1166, Aug. 2014.

[59] J. M. Kniss, R. Van Uitert, A. Stephens, G.-S. Li, T. Tasdizen, and C. Hansen. Statistically quantitative volume visualization. In *Visualization, 2005. VIS 05. IEEE*, pages 287–294. IEEE, 2005.

[60] P. Krishnamurthy, J. Buhler, R. Chamberlain, M. Franklin, K. Gyang, and J. Lancaster. Biosequence similarity search on the Mercury system. In *Proc. of IEEE 15th International Conference on Application-specific Systems, Architectures and Processors*, pages 365–375, Sept. 2004.

[61] P. Krishnamurthy, J. Buhler, R. D. Chamberlain, M. A. Franklin, K. Gyang, A. Jacob, and J. Lancaster. Biosequence similarity search on the Mercury system. *Journal of VLSI Signal Processing*, 49(1):101–121, Oct. 2007.

[62] J. S. Kuman and D. R. Patel. A survey on Internet of Things: Security and privacy issues. *International Journal of Computer Applications*, 90(11), Mar. 2014.

[63] J. Lancaster, J. Buhler, and R. D. Chamberlain. Acceleration of ungapped extension in Mercury BLAST. In *Proc. of 7th Workshop on Media and Streaming Processors*, Nov. 2005.

[64] J. Lancaster, J. Buhler, and R. D. Chamberlain. Acceleration of ungapped extension in Mercury BLAST. *Microprocessors and Microsystems*, 33(4):281–289, June 2009.

[65] A. H. Landberg, K. Nguyen, E. Pardede, and J. W. Rahayu. δ-dependency for privacy-preserving XML data publishing. *Journal of Biomedical Informatics*, 50:77–94, 2014.

[66] C. Li, D. Y. Li, G. Miklau, and D. Suciu. A theory of pricing private data. *ACM Trans. Database Syst.*, 39(4):34:1–34:28, Dec. 2014.

[67] K. Ligett and A. Roth. Take it or leave it: Running a survey when privacy comes at a cost. In *Proc. of 8th International Conference on Internet and Network Economics*, pages 378–391, 2012.

[68] C. Lundström, P. Ljung, A. Persson, and A. Ynnerman. Uncertainty visualization in medical volume rendering using probabilistic animation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1648–1655, 2007.

[69] A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowledge Discovery from Data*, 1(1):3:1–3:52, Mar. 2007.

[70] F. D. McSherry. Privacy integrated queries: An extensible platform for privacy-preserving data analysis. In *Proc. of ACM SIGMOD International Conference on Management of Data*, pages 19–30, New York, NY, USA, 2009. ACM.

[71] S. Mehrotra, A. Kobsa, N. Venkatasubramanian, and S. R. Rajagopalan. TIPPERS: A privacy cognizant IoT environment. In *Proc. of IEEE International Conference on Pervasive Computing and Communication Workshops (PerCom Workshops)*, pages 1–6, Mar. 2016.

[72] D. Mossman. Three-way ROCs. *Medical Decision Making*, 19(1):78–89, 1999.

[73] H. H. Nguyen, J. Kim, and Y. Kim. Differential privacy in practice. *Journal of Computing Science and Engineering*, 7(3):177–186, Sept. 2013.

[74] A. Ottley, B. Metevier, P. K. Han, and R. Chang. Visually communicating Bayesian statistics to laypersons. Technical Report TR-2012-02, Tufts University, Medford, MA, USA, 2012.

[75] A. Ottley, E. M. Peck, L. T. Harrison, D. Afergan, C. Ziemkiewicz, H. A. Taylor, P. K. Han, and R. Chang. Improving Bayesian reasoning: The effects of phrasing, visualization, and spatial ability. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):529–538, 2016.

[76] A. Ottley, H. Yang, and R. Chang. Personality as a predictor of user strategy: How locus of control affects search strategies on tree visualizations. In *Proc. of 33rd ACM Conference on Human Factors in Computing Systems*, pages 3251–3254. ACM, 2015.

[77] E. M. M. Peck, B. F. Yuksel, A. Ottley, R. J. Jacob, and R. Chang. Using fNIRS brain sensing to evaluate information visualization interfaces. In *Proc. of SIGCHI Conference on Human Factors in Computing Systems*, pages 473–482. ACM, 2013.

[78] M. C. Politi, P. K. Han, and N. F. Col. Communicating the uncertainty of harms and benefits of medical interventions. *Medical Decision Making*, 27(5):681–695, 2007.

[79] K. Pöthkow, B. Weber, and H.-C. Hege. Probabilistic marching cubes. In *Computer Graphics Forum*, volume 30, pages 931–940. Wiley Online Library, 2011.

[80] K. Potter, A. Wilson, P.-T. Bremer, D. Williams, C. Doutriaux, V. Pascucci, and C. R. Johnson. Ensemble-vis: A framework for the statistical visualization of ensemble data. In *Data Mining Workshops, 2009. ICDMW'09. IEEE International Conference on*, pages 233–240. IEEE, 2009.

[81] J.-S. Praßni, T. Ropinski, and K. Hinrichs. Uncertainty-aware guided volume segmentation. *IEEE Transactions on Visualization & Computer Graphics*, (6):1358–1365, 2010.

[82] V. Rastogi and S. Nath. Differentially private aggregation of distributed time-series with transformation and encryption. In *Proc. of ACM SIGMOD International Conference on Management of Data*, pages 735–746, New York, NY, USA, 2010. ACM.

[83] C. Rodriguez, F. Daniel, F. Casati, and C. Cappiello. Toward uncertain business intelligence: the case of key indicators. *IEEE Internet Computing*, 14(4):32–40, 2010.

[84] P. Samarati. Protecting respondents' identities in microdata release. *IEEE Trans. on Knowl. and Data Eng.*, 13(6):1010–1027, 2001.

[85] J. Sanyal, S. Zhang, G. Bhattacharya, P. Amburn, and R. Moorhead. A user study to compare four uncertainty visualization methods for 1D and 2D datasets. *IEEE Transactions on Visualization and Computer Graphics*, 15(6), 2009.

[86] J. Sanyal, S. Zhang, J. Dyer, A. Mercer, P. Amburn, and R. Moorhead. Noodles: A tool for visualization of numerical weather model ensemble uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 16(6):1421–1430, 2010.

[87] P. Sarlin. Macroprudential oversight, risk communication and visualization. *Journal of Financial Stability*, 27:160–179, Dec. 2016.

[88] B. Schneier. Stop trying to fix the user. *IEEE Security Privacy*, 14(5):96–96, Sept. 2016.

[89] R. Shokri and V. Shmatikov. Privacy-preserving deep learning. In *Proc. of ACM SIGSAC Conf. on Computer and Communications Security*, pages 1310–1321. ACM, 2015.

[90] D. Spiegelhalter, M. Pearson, and I. Short. Visualizing uncertainty about the future. *Science*, 333(6048):1393–1400, 2011.

[91] L. Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[92] S. Tak, A. Toet, and J. van Erp. The perception of visual uncertaintyrepresentation by non-experts. *IEEE Transactions on Visualization and Computer Graphics*, 20(6):935–943, 2014.

[93] E. Vasilomanolakis, J. Daubert, M. Luthra, V. Gazis, A. Wiesmaier, and P. Kikiras. On the security and privacy of Internet of Things architectures and systems. In *Proc. of International Workshop on Secure Internet of Things (SIoT)*, pages 49–57, Sept. 2015.

[94] J. Wang, S. Liu, and Y. Li. A review of differential privacy in individual data release. *International Journal of Distributed Sensor Networks*, Jan. 2015.

[95] D. Xiao. Is privacy compatible with truthfulness? In *Proc. of 4th Conference on Innovations in Theoretical Computer Science*, pages 67–86, 2013.

[96] Y. Xiao, L. Xiong, and C. Yuan. Differentially private data release through multidimensional partitioning. In W. Jonker and M. Petković, editors, *Proc. of 7th VLDB Workshop on Secure Data Management*, pages 150–168. Springer Berlin Heidelberg, Berlin, Heidelberg, 2010.

[97] C. Ziemkiewicz, A. Ottley, R. J. Crouser, A. R. Yauilla, S. L. Su, W. Ribarsky, and R. Chang. How visualization layout relates to locus of control and other personality factors. *IEEE Transactions on Visualization and Computer Graphics*, 19(7):1109–1121, 2013.