Privacy Concerns with Embracing Smart Cities

- Evaluating the impact of Chief Privacy Officer appointment on public privacy sentiment

Mandhani Kushal
Pranjal Sharma
mkushal@ncsu.edu
psharma9@ncsu.edu
North Carolina State University
Raleigh, North Carolina

ABSTRACT

The project focuses on the impact the presence of a Chief Privacy Officer (Chief Privacy Officer) has on people's adoption of smart cities. To deal with all privacy-related concerns for smart cities, we consider the notion of appointing a Chief Privacy Officer for the smart city by drawing parallels with the industry. This idea has had significant success in the industry, because it ensures that there is an entire office dedicated to designing and implementing policies to protect user data. On the other hand, most smart cities lack such dedicated offices and executives while they are most vulnerable to different types of privacy attacks. We conduct a comprehensive user study to understand if citizens of a smart city would be more comfortable sharing their personal data with the government if their city appointed a Chief Privacy Officer.

1 PROBLEM MOTIVATION

For governments, the need for smarter and more efficient ways of governance has led to increasing interest in smart cities. In order to achieve its goal of providing better services to its citizens, a smart city collects enormous amounts of data, processes the same to make optimal decisions, and then manipulates the city operations through active feedback [11] [5]. This leads to concerns over privacy, which is one of the major reasons why smart cities have not been adopted at the rate that they were expected to [6].

Having a dedicated office and executive for privacy could help avoid attacks on people's privacy and thereby gain more public confidence. However, the lack of relevant studies in the context of smart cities leads us to conduct a comprehensive user study. This study aims in understanding the impact the presence of a dedicated Chief Privacy Officer (and their office) has on people's willingness to share data with smart cities.

2 RELATED WORK

2.1 Privacy concerns in smart cities [7]

A framework to identify privacy concerns in the context of smart cities has been proposed. Two factors for categorization are: Data type (personal vs impersonal) and Purpose of data collection (service vs surveillance).

- `Quadrant I`: Personal Data collected for the purpose of providing a Service
- `Quadrant II`: Personal Data collected for the purpose of Surveillance
- `Quadrant III`: Impersonal Data collected for the purpose of Surveillance
- `Quadrant IV`: Impersonal Data collected for the purpose of providing a Service

This work provides us with a framework which would be useful in designing this user study.

Apart from this, this study also provides useful specific cases for each of the above quadrants and how they move to a different quadrant with a slight modification. Specifically, we are interested in these examples mentioned: Predictive Policing and Social Media Monitoring.

Predictive Policing[4] requires usage of impersonal data (location, time, etc.) to predict crimes (by surveillance). However, Police can start using personal identifiers (gender, race, ethnicity, etc.,) to make improvements in their predictions. This will lead to a transition from Quadrant III to Quadrant II.

Social Media Monitoring involves usage of Social Media Platforms for interacting with the citizens with the intention of providing services. This provides citizens with a platform to raise requests/concerns with the city officials directly with ease. However, the personal data collected and generated over these platforms can be used for analytics and sentiment analysis. This will lead to a transition from Quadrant I to Quadrant II. We have modified this case to come up with the scenario of Smart Complaint System (formally introduced in the Approach and Design section), where the threat to privacy is more pronounced.

2.2 Security and Privacy Challenges in Smart Cities [3]

This work first formally defines smart cities. The paper then highlights key privacy-related challenges in smart cities. The challenges include privacy threats due to data sharing and mining, threats in mashup data, insufficient cloud security, secondary use of data, and threats of Artificial Intelligence.

2.3 Do You Need A Chief Privacy Officer? [1]

This work highlights the situations which demand a Chief Privacy Officer role's necessity. It claims that when the risk of facing privacy issues is high, the need for a Chief Privacy Officer role is imminent. Since smart cities face a very high risk of privacy breaches, we aim to identify citizens' perspective on the appointment of Chief Privacy Officer.

2.4 The Evolving Role of the Privacy and Security Officer [2]

This work highlights the evolution of the privacy officer over the years and how this role transitioned from a few well-defined responsibilities to constantly requiring awareness. It is intriguing to evaluate if such a well-defined role can give confidence to the citizens of a smart city with their privacy.

3 APPROACH AND DESIGN

3.1 Data Collection Methodology

We designed a comprehensive user survey on Qualtrics to collect data. To respect the privacy of survey takers, all the responses were anonymized before analysis. To make the survey more effective, multiple response formats were used like 5-point likert scale (including matrix) and text box.

The survey was shared with the users using a link on emails and social media platforms. A total of 63 responses were received, which after filtering were reduced to 32 (filtering criterion will be discussed in detail in the section: Evaluation and Results).

3.2 User Study Design

The project uses the quadrant privacy framework introduced by Zoonen [7] to create appropriate question categories for the user study. In particular, the questions are based on data type (personal vs impersonal) and purpose of data collection (service vs surveillance).

Furthermore, instead of asking questions right away, we put users in different scenarios before eliciting response to pertaining questions. The scenarios have served two primary purposes: first, provide users with enough context for providing informed answers and second, keep users engaged during the survey.

The survey was designed as follows:

- The survey is divided into two sections, one taking input when there is no Chief Privacy Officer and the other after a Chief Privacy Officer has been appointed.
- Each of these two sections has three scenarios each (repeated) which have been identified by us. Each scenario has one primary question. However, in the scenario with Chief Privacy Officer, different follow-up questions are also asked based on the confidence level exhibited by user's response to the primary question.

Each section consists of three scenarios:

(1) **Predictive Policing**:
Predictive Policing is a practice aimed at preventing crimes by predicting them based on historical crime and surveillance data.

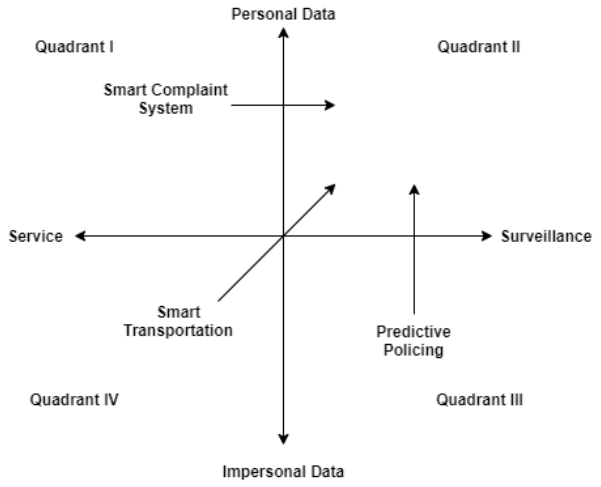

Figure 1: Transition across Quadrants for each scenario

This scenario focuses on understanding user's perspective when predictive policing operates primarily in Quadrant II (in both the cases: with and without the appointment of Chief Privacy Officer). The primary question for this scenario tries to gauge user response if they were told that police were using personal details like gender, religion and race for predictive policing (predominantly based in Quadrant II).

After explaining the pertinent responsibilities (enforce laws and policies, publish technical details, anonymize data) of a Chief Privacy Officer, the same question was asked to the users to understand if the appointment of Chief Privacy Officer leads to a change in the privacy confidence with the said practice.

If a user showed high confidence for this scenario in the presence of Chief Privacy Officer, they were asked to rate how helpful the different factors/responsibilities were for achieving high confidence. On the other hand, if a user showed low confidence, they were asked to list the practices (using textual input) that would've increased their confidence.

(2) **Smart Complaint System**:

A Smart Complaint System involves the use of complaint kiosks deployed in public places to conveniently raise complaints and requests with government departments.

This scenario focuses on understanding user's perspective when they are exposed to a Smart Complaint System, operating predominantly in Quadrant II (in both the cases). The primary question for this scenario tries to gauge user response if they were told that the department operating the complain system could run analytics on the request/complaint data to make inferences about their daily routine.

After explaining the pertinent responsibilities (notify about data collection, enforce laws and policies, publish technical details, anonymize data) of a Chief Privacy Officer, the same question was asked to the users to understand if the appointment of Chief Privacy Officer leads to a change in the privacy confidence with the said system.

If a user showed high confidence for this scenario in the presence of Chief Privacy Officer, they were asked to rate how helpful the different factors/responsibilities were for achieving high confidence. On the other hand, if a user showed low confidence, they were asked to list the practices (using textual input) that would've increased their confidence.

(3) **Smart Transportation**:

Smart Transportation involves the use of deployed cameras at bus stops to assess the demand, and adjust the number/frequency of buses accordingly.

This scenario focuses on gaining user's perspective when they are exposed to Smart Transportation, operating predominantly in Quadrant II (in both the cases). The primary question for this scenario tries to gauge user response if they were told that the system employs facial recognition which could also be used to make inferences about their whereabouts and travel patterns.

After explaining the pertinent responsibilities (notify about data collection, enforce laws and policies, publish technical details, anonymize data) of a Chief Privacy Officer, the same question was asked to the users to understand if the appointment of Chief Privacy Officer leads to a change in the privacy confidence with the said system.

If a user showed high confidence for this scenario in the presence of Chief Privacy Officer, they

were asked to rate how helpful the different factors/responsibilities were for achieving high confidence. On the other hand, if a user showed low confidence, they were asked to list the practices (using textual input) that would've increased their confidence.

The survey design was done such that it would help identify any comfort level improvements by first understanding the baseline for each user in Section 1 and then finding the delta with Section 2.

The survey can be accessed here: click to access survey.

4 EVALUATION AND RESULTS

4.1 Filtering of Data

After data collection, the first task was to filter incomplete and unreliable data. We filtered the responses as follows:

- **Incomplete responses**: As the Qualtrics software captures all responses even if they haven't been submitted, we first filtered out all the responses that were not submitted successfully.
- **Unreliable responses**: Some of the responses we received were filled with a total duration of under 60 seconds. We marked these responses as unreliable and filtered them out.

4.2 Evaluation Methodology

The main aim during evaluation was to determine and showcase the change, if any, Chief Privacy Officer brings to user's privacy sentiment in Smart Cities. We also focus on understanding the factors that lead to this change.

4.2.1 Change in user sentiment after appointment

The first evaluation was to understand if there was any change in user's privacy sentiment after appointment of the Chief Privacy Officer.

Pie chart demonstrate the change in user's privacy sentiment:

We generated a pie chart and categorized the responses into four categories based on the user's comfort level in sharing data:

(1) **Significant increase**: This category is to indicate that a User's comfort level in sharing data has increased drastically after appointment of the Chief Privacy Officer. In the survey, this category is used when a response changed from [1/2/3] to [4/5] after appointment of Chief Privacy Officer.
(2) **Marginal increase**: This category is to indicate marginal increase after appointment of the Chief Privacy Officer. In the survey, this category is used when a response changed from [1] to [2] or [1/2] to [3] or [4] to [5] after appointment of Chief Privacy Officer.
(3) **No change**: This category indicates no change after appointment of the Chief Privacy Officer.
(4) **Decrease**: This category indicates that a User's comfort level has reduced after the appointment of the Chief Privacy Officer.

CCDF graph to demonstrate the change in user's privacy sentiment:

The likert scale from 1 to 5 was used so as to enable us to plot meaningful statistical graphs like a complementary cumulative distribution function (CCDF). The CCDF plot helps us better visualize the distribution of data we received as our responses[9]. Some more information that helps us understand this plot:

- **Y - axis**: The probability that a User's rating is greater than or equal to x (the X - axis value) - Y = P ($X \geq x$).
- **X - axis**: All possible comfort levels for User's ratings.

Note that we have used a modified CCDF graph where we consider Y = P($X \geq x$) instead of the typical Y = P($X > x$) and chose not to use the log scale to provide more readability. While reading this graph, it can be noticed that it starts with probability 1 when X=1, meaning that the probability that a user's rating is greater than or equal to 1 is 1.

4.2.2 Factors leading to high user comfort levels

The second evaluation was carried out to identify what factors lead to a high comfort level with data sharing after appointment of the Chief Privacy Officer in Smart City. These factors comprise of Chief Privacy Officer's responsibilities as described in the previous sections.

To showcase the impact of each factor, we have used a percent stacked bar chart [8] where each bar represents a factor. Each bar has multiple stacks (which represents

the percentage of users with particular impact level). We have defined three such stacks:

- Helpful - With ratings 4-5
- Neutral - With rating 3
- Not Helpful - With ratings 1-2

This provides insightful comparison between each factor and how much it contributes to this positive impact.

We have also generated a CCDF graph to showcase the impact of factors for each of the three scenarios.

4.2.3 *Generation of graphs*

These graphs were generated using python scripts based on matplotlib module. The GitHub repository can be accessed from https://github.ncsu.edu/psharma9/priv-smart.

4.3 Results

The following sub-sections contain charts showcasing the results of this user study.

4.3.1 *Change in user sentiment after appointment*

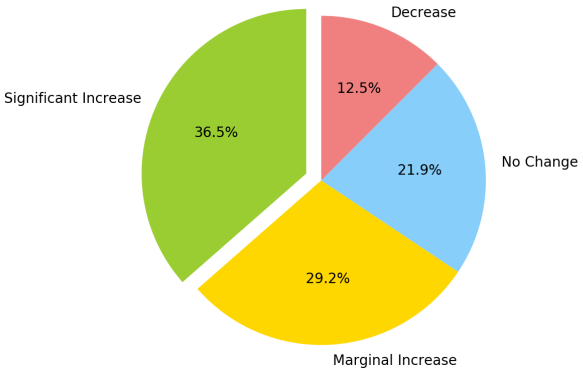

Figure 2: Pie chart showing change in user sentiment after appointment of Chief Privacy Officer across all scenarios

From Figure 2, we can notice that:

- Significant increase in 36.5% of respondent's comfort level for sharing their data after the appointment of Chief Privacy Officer in Smart Cities.
- An overall increase was observed in 65.7% of the respondents including marginal increase.
- 12.5% of the respondents indicated a decrease in their comfort level for sharing data.

Note: Each user's response for three scenarios have been considered for plotting the overall result.

For individual scenarios:

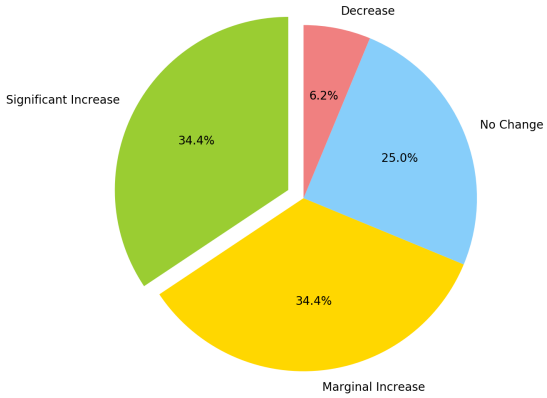

Figure 3: Pie chart showing change in user sentiment for Predictive Policing

From Figure 3, we can notice an overall increase in 68.8% of the respondents with a significant increase in 34.4% for Predictive Policing.

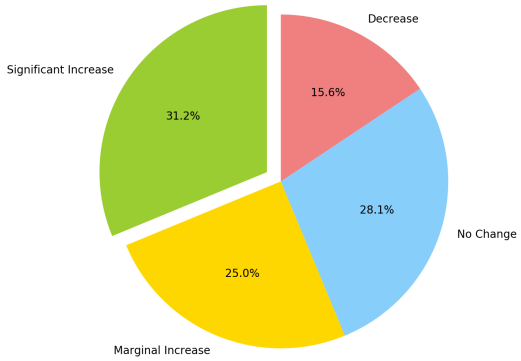

Figure 4: Pie chart showing change in user sentiment for Smart Complaint System

From Figure 4, we can notice an overall increase in 56.2% of the respondents with a significant increase in 31.2% for Smart Complaint System.

From Figure 5, we can notice an overall increase in 71.9% of the respondents with a significant increase in 43.8% for Smart Transportation.

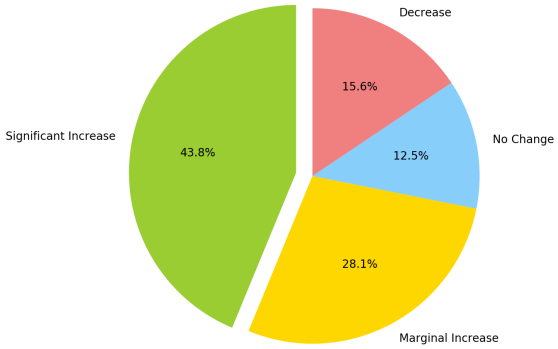

Figure 5: Pie chart showing change in user sentiment for Smart Transportation

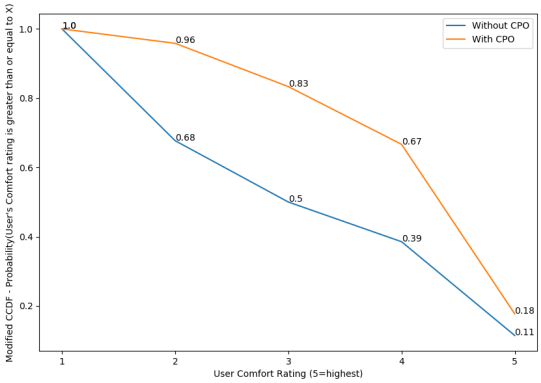

Figure 6: CCDF graph for user sentiment before and after appointment of Chief Privacy Officer for all scenarios

CCDF Graphs

From Figure 6, we can notice that for all scenarios:

- The curve for "With CPO" (orange) is much higher than the curve for "Without CPO" (blue). This means that the distribution of data for "With CPO" indicates a higher comfort level than "Without CPO" distribution.
- For instance, P(X ≥ 4) is 0.67 for "With CPO" and 0.39 for "Without CPO". This means that more responses in "With CPO" are greater than 4 (67%) in comparison with "Without CPO" (39%).

From Figure 7, we can notice that for Predictive Policing:

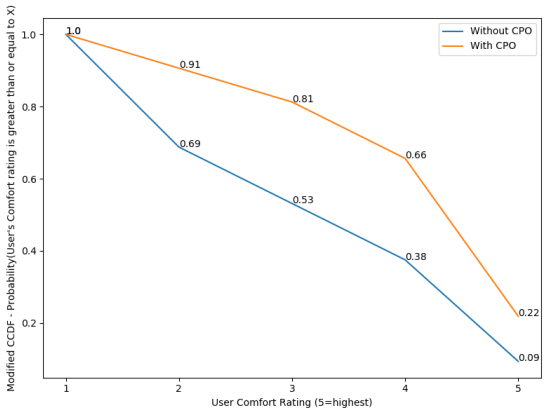

Figure 7: CCDF graph for change in user sentiment for Predictive Policing

- Again, we notice the curve for "With CPO" (orange) is always higher than the curve for "Without CPO" (blue) indicating an increase in user's privacy sentiment.

We notice the same trend in CCDF graphs in Figures 8 and 9 for Smart Complaint System and Smart Transportation respectively. However we can notice that among the scenarios, we see largest improvement in "Smart Transportation" (more difference between the curves).

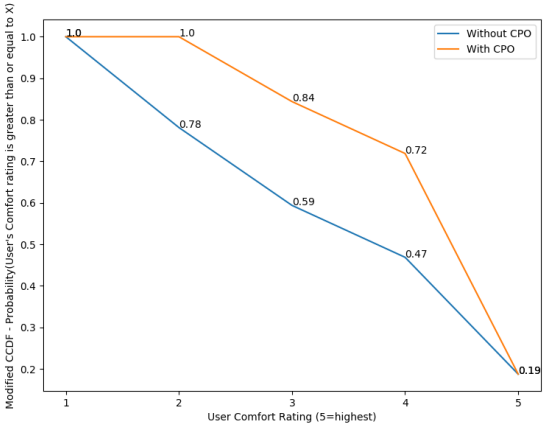

Figure 8: CCDF graph for change in user sentiment for Smart Complaint System

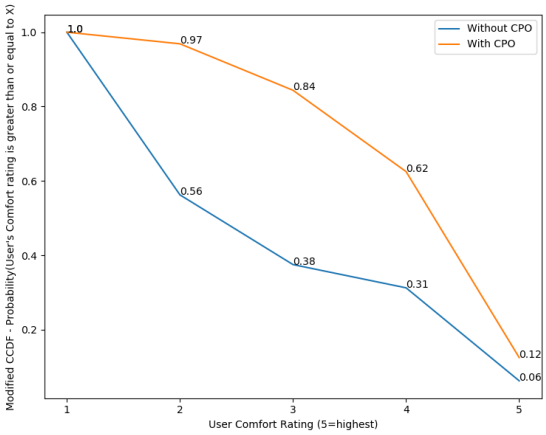

Figure 9: CCDF graph for change in user sentiment for Smart Transportation

4.3.2 Factors leading to high user comfort levels

The following results show how each of the factors (practices of the Officer) contributes towards achieving high confidence in sharing their data with Smart cities.

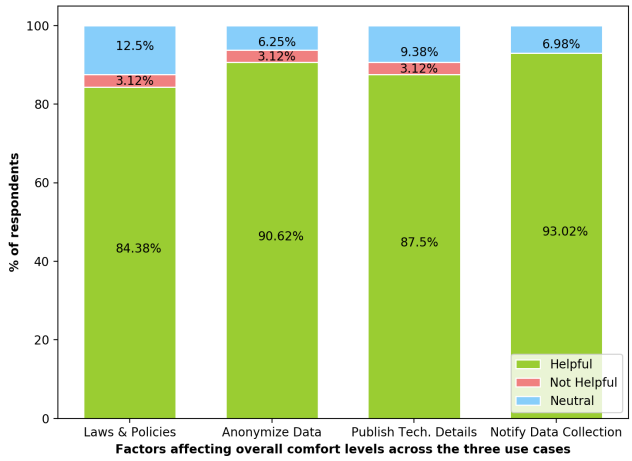

Figure 10: Stacked Bar Chart showing impact of each factor towards achieving high user confidence for all scenarios

From Figure 10, we can see that among users with high confidence:

- 93.02% of users found "Notify about Data Collection" an important factor, followed by "Anonymize data" (90.62% of users).
- 12.5% of users found "Enforce Laws and Policies" as not helpful followed by "Publish Technical Details" (9.38% of users).
- All factors have been found helpful by more than 84% of the users.

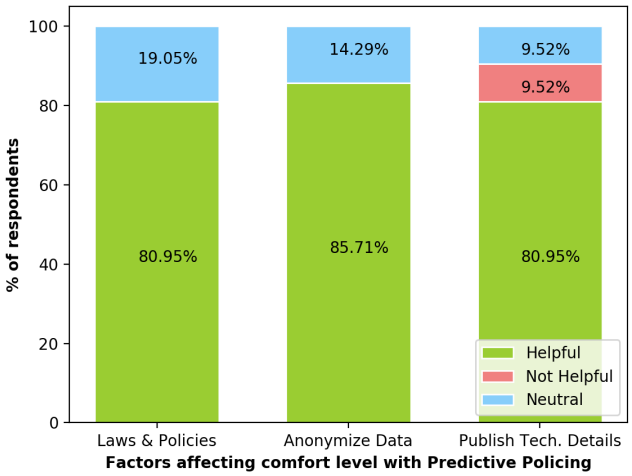

Figure 11: Stacked Bar Chart showing impact of each factor towards achieving high user confidence for Predictive Policing

From Figure 11, we can see that among users with high confidence for Predictive Policing:

- All factors have been found helpful by more than 80% of the users with "Anonymize data" as the most popular with 85.71% of users.
- 19.05% of users found "Enforce Laws and Policies" unhelpful.

From Figure 12, we can see that among users with high confidence for Smart Complaint System:

- 91.3% of user's found "Anonymize Data", "Publish technical Details" and "Notify about Data Collection" as helpful.

From Figure 13, we can see that among users with high confidence for Smart Transportation:

- 95% of user's found "Anonymize Data" and "Notify about Data Collection" as helpful while 90% of users found "Publish technical Details" and "Enforce Laws and Policies" helpful.
- All the factors are found useful by more than 90% of users for this scenario.

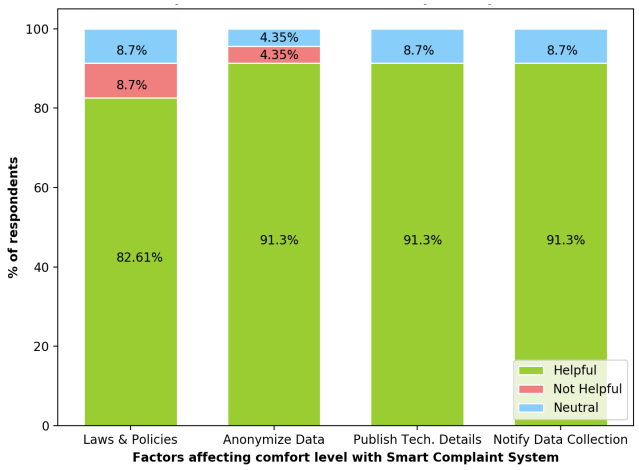

Figure 12: Stacked Bar Chart showing impact of each factor towards achieving high user confidence for Smart Complaint System

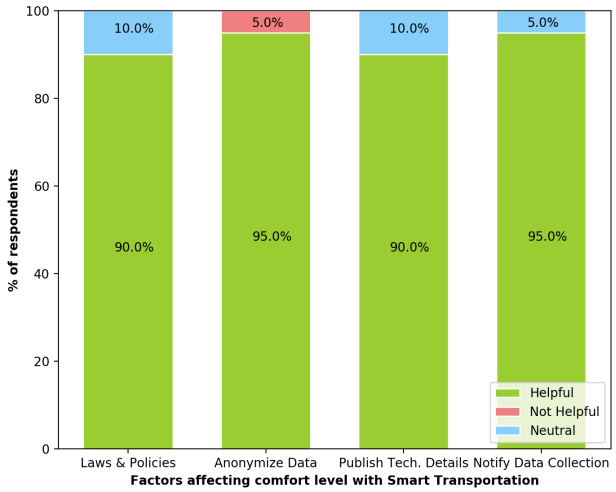

Figure 13: Stacked Bar Chart showing impact of each factor towards achieving high user confidence for Smart Transportation

CCDF Graphs for impact of factors on achieving high confidence:

From Figure 14, we can notice that for Predictive Policing:

- The curves for all the factors are usually high (as expected) with "Anonymize Data" leading (staying higher more often than other factors).

From Figure 15, we can notice that for Smart Complaint System:

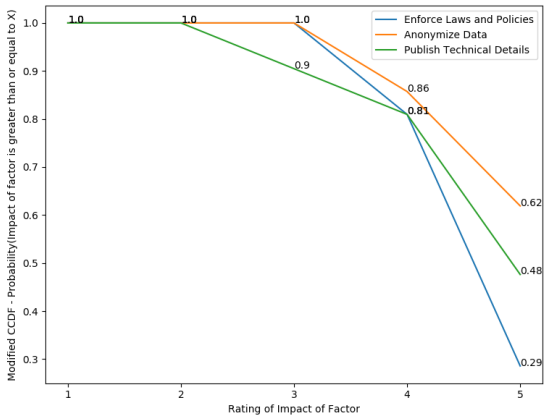

Figure 14: CCDF graph for impact of each factor towards achieving high user confidence for Predictive Policing

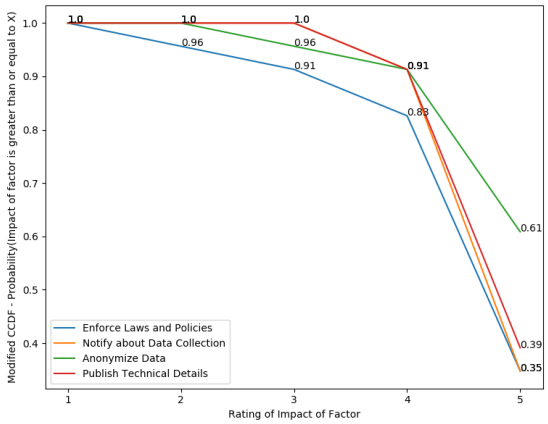

Figure 15: CCDF graph for impact of each factor towards achieving high user confidence for Smart Complaint System

- The curves for all the factors are high (as expected) with "Enforce Laws and Policies" having least impact (staying lower more often than other factors).

From Figure 16, we can notice that for Smart Transportation:

- The curves for all the factors are high (as expected, with no significant variations).

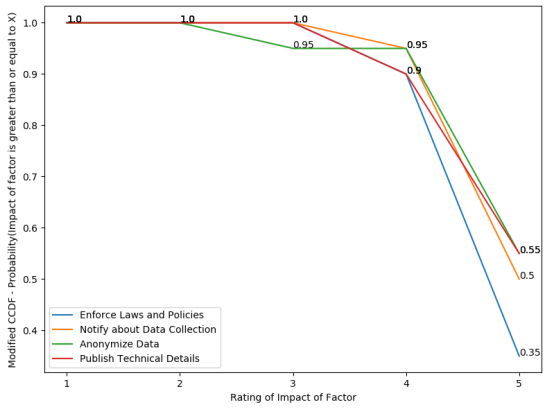

Figure 16: CCDF graph for impact of each factor towards achieving high user confidence for Smart Transportation

4.3.3 *Suggested factors by users with low comfort levels*

For the users with low comfort levels in sharing their data in the presence of CPO, they suggested some factors apart from the Chief Privacy Officer's appointment which will boost their comfort level. They have been identified across the three scenarios:

(1) **Predictive Policing:**
- Predictive Policing should only use criminal records and valid security breaches and not do complete surveillance.
- The smart city police should undergo training to have no biases.

(2) **Smart Complaint System:**
- The smart city should have stricter rules and fines on violation of privacy laws and policies.
- Third-party auditing (on a yearly basis) to review privacy practices.

(3) **Smart Transportation:**
- Use of other simpler methods instead of cameras. Ex: Proximity sensors (even if it means lower accuracy)
- On-demand available information similar to "The Freedom of Information Act"[10].

Overall, we can see that the appointment of a Chief Privacy Officer in Smart Cities significantly improves the privacy sentiment of 36.5% of the users. All the factors discussed contribute significantly to the high comfort level of users to share data, with "Notify about Data Collection" (except for Predictive Policing) and "Anonymize Data" leading with a small margin.

5 LIMITATIONS AND FUTURE WORK

Some of the limitations are:

- **Sample Size**: For a more thorough analysis, more responses would've been needed.
- **Respondent Diversity**: Current respondents were mostly students (some were working professionals). More diversity in terms of age and exposure to smart cities would've been ideal.

The future work will be along the following lines:

- **More factors**: Since the survey follows a factorial design, more factors can be added thereby making it more fine-grained and comprehensive.
- **Clustering responses**: Clustering techniques can be used to automatically group user responses mentioning techniques that would increase their confidence levels

6 INDIVIDUAL CONTRIBUTIONS

Equal contributions were made during survey design and distribution.

Evaluation was divided into five parts.

- Part 1: Response Filtering - Mandhani
- Part 2: Plotting Pie Charts - Pranjal
- Part 3: Plotting Stacked Bar Charts - Pranjal
- Part 4: Plotting CCDF Graphs - Mandhani
- Part 5: Confidence Decrease Analysis and Tabulation - Pranjal and Mandhani

ACKNOWLEDGMENTS

To Professor Anupam Das, for his unconditional support and guidance throughout this project.

REFERENCES

[1] Steven C. Bennett. 2007. *Do You Need A Chief Privacy Officer?* Retrieved December 08, 2019 from https://www.jonesday.com/files/Publication/65a036d5-fccc-4489-83a6-20eecc6969ff/Presentation/PublicationAttachment/0e519319-ab0f-44c8-b193-224473166495/Privacy_Bennett%20.pdf

[2] MA RHIA C.H.P.S. S.S.G.B. Bowen, Rita K. 2015. The Evolving Role of the Privacy and Security Officer. *Journal of AHIMA* 86, 6 (06 2015), 46–47. https://proxying.lib.ncsu.edu/index.php/login?url=https://

search.proquest.com/docview/1702085963?accountid=12725 Copyright - Copyright American Health Information Management Association Jun 2015; Last updated - 2015-08-07.

[3] Trevor Braun, Benjamin C.M. Fung, Farkhund Iqbal, and Babar Shah. 2018. Security and privacy challenges in smart cities. *Sustainable Cities and Society* 39 (2018), 499 – 507. https://doi.org/10.1016/j.scs.2018.02.039

[4] Andrew Guthrie Ferguson. 2016. Policing predictive policing. *Wash. UL Rev.* 94 (2016), 1109.

[5] Albert Meijer and Manuel Pedro Rodríguez Bolívar. 2016. Governing the smart city: a review of the literature on smart urban governance. *International Review of Administrative Sciences* 82 (2016), 392 – 408. https://doi.org/10.1177/0020852314564308

[6] Daniel Newman. 2019. *Are Privacy Concerns Halting Smart Cities Indefinitely?* Retrieved Dec 8, 2019 from https://www.forbes.com/sites/danielnewman/2019/01/08/are-privacy-concerns-halting-smart-cities-indefinitely/

[7] Liesbet van Zoonen. 2016. Privacy concerns in smart cities. *Government Information Quarterly* 33, 3 (2016), 472 – 480. https://doi.org/10.1016/j.giq.2016.06.004 Open and Smart Governments: Strategies, Tools, and Experiences.

[8] Wikipedia contributors. 2019. Bar chart — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Bar_chart&oldid=925835611 [Online; accessed 9-December-2019].

[9] Wikipedia contributors. 2019. Cumulative distribution function — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Cumulative_distribution_function&oldid=929058561 [Online; accessed 9-December-2019].

[10] Wikipedia contributors. 2019. Freedom of Information Act (United States) — Wikipedia, The Free Encyclopedia. https://en.wikipedia.org/w/index.php?title=Freedom_of_Information_Act_(United_States)&oldid=926894704 [Online; accessed 9-December-2019].

[11] K. Zhang, J. Ni, K. Yang, X. Liang, J. Ren, and X. S. Shen. 2017. Security and Privacy in Smart City Applications: Challenges and Solutions. *IEEE Communications Magazine* 55, 1 (January 2017), 122–129. https://doi.org/10.1109/MCOM.2017.1600267CM