USENIX'23 Artifact Appendix: Are Consumers Willing to Pay for Security and Privacy of IoT Devices?

Pardis Emami-Naeini[*], Janarth Dheenadhayalan[†], Yuvraj Agarwal[†], Lorrie Faith Cranor[†]

[*]*Duke University*   [†]*Carnegie Mellon University*

A   Artifact Appendix

A.1   Abstract

By conducting a two-phase online study on Prolific, we quantified the impact of various security and privacy improvements on Internet of Things (IoT) consumers' purchase behavior. Through designing an incentive-compatible experiment using the multiple price list (MPL) methodology, we captured participants' willingness to pay for transparency over security and privacy enhancements of smart devices. We constructed three regression models for each phase of our online study to quantify and explain participants' risk perception, willingness to purchase, and willingness to pay. In this artifact, we provide participants' de-identified survey data that we used to construct these models, the analysis code in R and STATA that we used to build the regression models, and the output files.

A.2   Description & Requirements

A.2.1   Security, privacy, and ethical concerns

Conducting the statistical models for this paper does not introduce any risks. In addition, we de-identified the raw survey data to preserve participants' data privacy.

A.2.2   How to Access

The artifact, including the raw, de-identified survey data, analysis files, model output files, and the README file, is hosted on GitHub and could be accessed via the following stable URL: https://github.com/pemamina/USENIX23_MonteryValueSP_Artifact/tree/e88e7eb5630996756f14335bf32abc4e9298e97a.

A.2.3   Hardware Dependencies

None.

A.2.4   Software Dependencies

We used an open source tool, RStudio, to run two of the regression models (risk_clmm and purchase_clmm). Since R currently does not allow constructing mixed effects interval regressions, we used STATA to build the model to explain participants' willingness to pay. We downloaded RStudio via the following link: https://www.RStudio.com/products/RStudio/download/. We downloaded R using the following link: https://cran.RStudio.com/. We obtained STATA by using the following link: https://www.STATA.com/.

A.2.5   Benchmarks

None.

A.3   Set-Up

A.3.1   Installation

After installing RStudio and R, we need to install (`install.packages("ordinal")`) and load (`library(ordinal)`) the ordinal library required to construct CLMM models. This process is shown in lines 2 and 3 of `phase_one_analysis.R`, `phase_two_analysis.R`. No package needs to be installed in STATA to conduct mixed-effects interval regression.

A.3.2   Basic Test

Here we use dataset `ologit.csv` from the OARC website (https://stats.oarc.ucla.edu/). This dataset includes four variables: 1) `apply`: nominal categorical variable with three levels (0, 1, 2) showing how likely it is that the student will apply for grad school, 2) `pared`: categorical binary variable, showing whether parents have attended college (1) or not (0), 3) `public`: categorical binary variable, showing whether the school the student has attended is public (1) or not (0), and 4) `gpa`: continuous numeric variable, showing the student's GPA score.

Analysis goal. We would like to understand the impact of parents' college education (`pared`) on students' likelihood of applying to college (`apply`). Similar to our risk perception model and the willingness to purchase model, the dependent variable in this test (`apply`) is ordinal categorical. Therefore, we will construct a cumulative link model (CLM) to explain the impact of `pared` on `apply`.

```
## Loading the required library for ordinal
↪  regression.
library(ordinal)
```

```r
## Loading the dataset ``ologit.csv''
dataset <-
↪  read.csv("https://stats.idre.ucla.edu/stat/data/ologit.csv")

## Changing the type of dependant variable
↪  \texttt{apply} from numerical (levels = 0, 1,
↪  2) to ordinal categorical (levels = "unlikely",
↪  "somewhat likely", "very likely")
dataset$apply <- factor(dataset$apply, labels =
↪  c("unlikely", "somewhat likely", "very likely"),
↪  ordered = TRUE)

## Changing the type of independent variable
↪  \texttt{pared} from numerical (levels = 0, 1)
↪  to nominal categorical (levels = "not attend",
↪  "attend")
dataset$pared <- factor(dataset$pared, labels = c("not
↪  attend", "attend"))

## Construction the CLM to explain the impact of
↪  \texttt{pared} on \texttt{apply}.
apply.clm <- clm(apply ~ pared, data = dataset)

# Showing the results
summary(apply.clm)
```

The output should look like:

```
formula: apply ~ pared
data:    dataset

 link  threshold nobs logLik  AIC    niter max.grad cond.H
 logit flexible  400  -361.40 728.79 5(0)  1.25e-10 9.3e+00

Coefficients:
           Estimate Std. Error z value Pr(>|z|)
paredattend  1.1275     0.2634    4.28 1.87e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 '
↪  ' 1

Threshold coefficients:
                        Estimate Std. Error z value
unlikely|somewhat likely   0.3768     0.1103   3.415
somewhat likely|very likely 2.4519     0.1826  13.430
```

By exponentiating the estimate in this model, we will calculate the odds ratio of 3.09. This shows that for students whose parents did attend the college, the odds of being more likely to apply to grad school is 3.09 times that of students whose parents did not attend the college.

A.4 Evaluation Workflow

The results of each phase of our study is based on three regression models. The regression results of phase one are included in `phase_one_CLMM_output.txt` and `phase_one_STATA_output.txt` and the regression results of phase two are included in `phase_two_CLMM_output.txt` and `phase_two_STATA_output.txt`. Here we provide the exact R and STATA code that we used to reach these results.

A.4.1 Major Claims

(C1): Our cumulative link mixed models describe participants' risk perception and willingness to purchase behavior in the first phase of our study. This is proven by the experiment (E1) described in Sections 4.2 and 4.3 of the paper, whose results are reported in Table 2 in the paper.

(C2): Our interval regression model describes participants' willingness to pay in the first phase of our study. This is proven by the experiment (E2) described in Sections 4.2 and 4.3 of the paper, whose results are reported in Table 2 in the paper.

(C3): Our cumulative link mixed models describe participants' risk perception and willingness to purchase behavior in the second phase of our study. This is proven by the experiment (E3) described in Sections 5.2 and 5.3 of the paper, whose results are reported in Table 4 in the paper.

(C4): Our interval regression model describes participants' willingness to pay in the second phase of our study. This is proven by the experiment (E4) described in Sections 5.2 and 5.3 of the paper, whose results are reported in Table 4 in the paper.

A.4.2 Experiments

(E1): Cumulative link mixed models in R: Risk perception and willingness to purchase models for phase-one study.

```r
## Loading the required library for regression
↪  analysis
library(ordinal)

## Loading the survey data
dataset <- read.csv("phase_one_survey_data.csv")

## Specifying the dependent variables as
↪  ordinal categorical
dataset$risk_perception_coded <-
↪  factor(dataset$risk_perception_coded, order =
↪  TRUE,
                         levels = c("1", "2", "3",
                                    "4", "5"))
dataset$willingness_to_purchase_coded <-
↪  factor(dataset$willingness_to_purchase_coded,
↪  order = TRUE,
                       levels = c("1", "2", "3",
                                  "4", "5"))

## Specifying the independent variables as
↪  categorical
dataset$order_scenario <-
↪  factor(dataset$order_scenario)
dataset$correct_definition_number <-
↪  factor(dataset$correct_definition_number)

## Setting the baseline for model independent
↪  variables
dataset$mostProtective_leastProtective_pair <-
↪  as.factor(dataset$mostProtective_leastProtective_pair)
dataset$smart_device <-
↪  as.factor(dataset$smart_device)

dataset$mostProtective_leastProtective_pair <-
↪  relevel(dataset$mostProtective_leastProtective_pair,
↪  "main_personal")
dataset$smart_device <- relevel(dataset$smart_device,
↪  "smoke")
```

```r
## Constructing the Risk Perception Model
risk_clmm <- clmm(risk_perception_coded ~
↪ mostProtective_leastProtective_pair + smart_device
↪ +
                 correct_definition_number +
                 ↪ order_scenario +
                     (1|participant), data =
                     ↪ dataset, link = "logit"
)
summary(risk_clmm)


## Constructing the Willingness to Purchase
↪ Model
purchase_clmm <-
↪ clmm(dataset$willingness_to_purchase_coded ~
↪ mostProtective_leastProtective_pair + smart_device
↪ +
                    correct_definition_number +
                    ↪ order_scenario +
                 (1|participant), data = dataset,
                 ↪ link = "logit"
)
summary(purchase_clmm)
```

(E2): Mixed interval regression model in STATA: Willingness to pay model for phase-one study. We first need to import our CSV file. Since this file has long participant quotes, we should ensure the values of the cells do not overflow. We will specify the following parameters when importing the datafile: delimiter(comma), bindquote(strict), and stripquote(yes).

```stata
* We create a label to show the order of
↪ independent variables.
. label define factor_lab 1 "main_personal"
. label define device_lab 1 "smoke" 2 "speaker"

* We recode the independent variables with the
↪ new baseline
. encode mostprotective_leastprotective_p,
↪ generate(attribute_value) label(factor_lab)
. encode smart_device, generate(device_value)
↪ label(device_lab)

* We construct the mixed interval regression
↪ and set participant as the random effect.
. meintreg minimum_willingness_to_pay
↪ maximum_willingness_to_pay i.order_scenario
↪ i.correct_definition_number i.device_value
↪ i.attribute_value || participant:
```

(E3): Cumulative link mixed models in R: Risk perception and willingness to purchase models for phase-two study.

```r
## Loading the required library for regression
↪ analysis
library(ordinal)

## Loading the survey data
dataset <- read.csv("phase_two_survey_data.csv")

## Specifying the dependent variables as
↪ ordinal categorical
dataset$risk_perception_coded <-
↪ factor(dataset$risk_perception_coded, order =
↪ TRUE,
```

```r
                              levels =
                              ↪ c("1",
                              ↪ "2", "3",
                              ↪ "4", "5"))
dataset$willingness_to_purchase_coded <-
↪ factor(dataset$willingness_to_purchase_coded,
↪ order = TRUE,
                                        levels
                                        ↪ =
                                        ↪ c("1",
                                        ↪ "2",
                                        ↪ "3",
                                        ↪ "4",
                                        ↪ "5"))

## Setting the baseline for model independent
↪ variables
dataset$label_type_comparison <-
↪ as.factor(dataset$label_type_comparison)
dataset$label_type_comparison <-
↪ relevel(dataset$label_type_comparison, "Z vs Y")

## Constructing the Risk Perception Model
risk_clmm <- clmm(risk_perception_coded ~
↪ label_type_comparison +
                 (1|participant), data = dataset,
                 ↪ link = "logit"
)
summary(risk_clmm)


## Constructing the Willingness to Purchase
↪ Model
purchase_clmm <-
↪ clmm(dataset$willingness_to_purchase_coded ~
↪ label_type_comparison +
                    (1|participant), data =
                    ↪ dataset, link = "logit"
)
```

```stata
* We create a label to show the order of
↪ independent variables.
. label define comparisonOrder 1 "Z vs Y" 2 "X vs Y"
↪ 3 "X vs Z"

* We recode the independent variables with the
↪ new baseline
. encode label_type_comparison,
↪ gen(typeComparisonCat) label(comparisonOrder)

* We construct the mixed interval regression
↪ and set participant as the random effect.
. meintreg min max i.typeComparisonCat || participant:
```

A.5 Notes on Reusability

None.

A.6 Version

Based on the LaTeX template for Artifact Evaluation V20220926. Submission, reviewing and badging methodology followed for the evaluation of this artifact can be found at https://secartifacts.github.io/usenixsec2023/.