How System Failures and Ransomwares Affect Drivers' Trust and Attitudes in an Automated Car? A Simulator Study

Payre, W., Perello March, J., Sabaliauskaite, G., Jadidbonab, H., Shaikh, S., Nguyen, H. N. & Birrell, S

DOI 10.54941/ahfe1002764

Publisher: AHFE International

Human Interaction and Emerging Technologies (IHIET 2022), Vol. 68, 2022, 453–460
https://doi.org/10.54941/ahfe1002764

How System Failures and Ransomwares Affect Drivers' Trust and Attitudes in an Automated Car? A Simulator Study

William Payre[1], Jaume Perelló-March[1], Giedre Sabaliauskaite[2], Hesamaldin Jadidbonab[2], Siraj Shaikh[2], Hoang Nguyen[2], and Stewart Birrell[1]

[1]National Transport Design Centre, Coventry University, Coventry, CV1 2TT, UK
[2]Systems Security Group, Centre for Future Transport and Cities, Coventry University, Coventry, CV1 5FB, UK

ABSTRACT

As cars are becoming more automated and connected, their vulnerability to cyber-attack is also increasing. This research aims to understand how drivers react to a cyber-attack affecting personal data in a connected and automated car. Thirty-seven participants participated in a driving simulator study where a ransomware popped-up on the centre console while driving in automated mode. An inductive content analysis was conducted to examine drivers' responses to open-ended questions. The analysis showed that participants identified a range of themes including 1) their interpretation of the ransomware displayed on the in-vehicle screen, 2) the expected effects of this message on the vehicle and its components (automated driving system and in-vehicle display) and 3) the attitudes and feelings drivers experienced. Drivers were primarily concerned about the detrimental effects of the ransomware on the automated driving system and road safety, but were less concerned with respect to personal data encryption. When not ignored, the cyberattack negatively affected trust in the automated system and drivers' emotions.

Keywords: Ransomware, Automation, Automotive, Cyber-attack, Behaviour, Attitudes, Qualitative analysis

INTRODUCTION

With the constant implementation of software solutions and digital interfaces in modern vehicles, cars have become increasingly connected and automated, making road safety and cybersecurity intertwined (Trope et al., 2018). Cybersecurity applied to automotive seeks to protect cars from malevolent electronic attacks. Connected and automated vehicles (CAVs) are expected to provide drivers with more comfort and safety, provided that automated driving systems perform better than human drivers and are not vulnerable to security breaches.

While some qualitative research has investigated cybersecurity knowledge and associated risks in the automotive industry (Morris et al., 2020), little attention has been devoted to drivers' responses to cyber-attacks while using

an automated car. The main benefits of a qualitative approach are (a) providing insight into drivers' experience when exposed to a cyber-attack in a CAV; (b) generating an understanding of the effect of a security breach on drivers' perception of CAV, and (c) the capacity to focus on subjective aspects of trust in automated driving related to cybersecurity. The present study sought to bridge a gap in the human factors-cybersecurity qualitative knowledge base. To fulfil this objective, participants were administered closed and open-ended questions after experiencing a cyber-attack, i.e. ransomware, in an automated driving simulator. The textual data collected were analysed using inductive content analysis, a method deemed suitable when the topic investigated has not been yet examined (Elo & Kyngäs, 2008).

The epistemological stance adopted for this study is pragmatism, which means researchers are knowledgeable on the subject investigated. Hence, they are not completely objective when analysing and interpreting the data, which will influence the findings (Bryant, 2009). In this instance, all the authors have a background in psychology or cybersecurity, including some experience in automotive experience.

This study aimed to perform an inductive content analysis on responses to open-ended questions that followed a cyber-attack during a driving simulator study. No *a priori* hypotheses were formulated because this study was inductive.

METHOD

Material, Experimental Procedure and Measures

Volunteer participants comprised of a convenience sample of 38 adults with 15 females and 22 males. On average, they were 36.2 years old (SD = 12.5) with 15.7 years of driving experience (SD = 13.1), and reported a yearly mileage of 7737 (SD = 5891). One participant dropped out from the study due to simulator sickness (N = 37). The high fidelity driving simulator consisted of a full-body vehicle with three degrees of freedom. A 7" touchscreen display attached to the centre console allowed drivers to toggle the automated driving mode on or off, see the status of the vehicle (i.e., manual vs. automated mode) and create a user profile including forename, surname, email and password (Fig. 1). These details would later be used to simulate a ransomware attack but were neither stored nor recorded. Participants provided informed consent and the study complied with the Coventry University Research Ethics Committee code and the General Data Protection Regulation.

Participants started with a 5 min familiarisation practice involving driving manually on suburban roads and activating automated driving after merging onto a motorway. Then, they completed three counter-balanced 12 min trials on a motorway for 15 miles, where the automated system would operate either seamlessly (i.e., no failure) or unreliably (i.e., turn signals not activating during an overtaking manoeuvre for the silent failure vs. ransomware attack for the explicit failure). After activating automated driving, participants undertook a word search task that lasted for the whole trial. The ransomware popped on the in-vehicle display while the car drove in automated mode (inspired by the WannaCry ransomware and Wolf et al., Lambert,

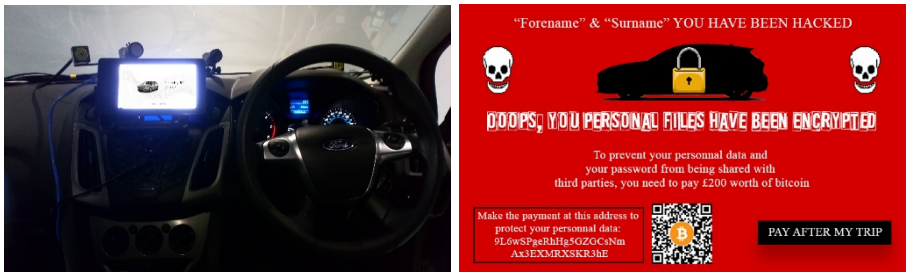

Figure 1: (left) Driver's view of the instrument cluster and HMI used for activating the automation (right) Ransomware displayed on the in-vehicle screen.

2017; see Fig. 1). Drivers were free to resume manual control and reactivate the automation at any time. The study lasted 2 hours and participants received a £20 voucher as a compensation for their time.

Drivers answered the following open-ended question after the *silent* and the *explicit failure* scenarios to understand their perception of the incident: *Did you notice anything abnormal with the conditionally automated vehicle?* For the sake of conciseness, this paper will only cover answers from the *explicit failure* scenario.

Analytical Strategy

The inductive content analysis serves to produce knowledge and understand the meaning of qualitative data (Hsieh & Shannon, 2005; Schreier, 2012). Kyngäs (2020) defined it as an open method. The first step is to set a unit of analysis that will allow identifying open codes in the text corpus. Afterwards, the similarities and differences of these open codes are assessed and then classified in sub-themes accordingly. The sub-themes are then categorised into first-order themes, which are further organised into broader, second-order themes. These second-order themes can be classified into wider categories i.e. general dimensions. This last step of the data classification depends on the notions elicited by the respondents and the nature of the data. Every dimensions and themes identified are labelled with content-characteristic words.

The inductive content analysis started with two researchers getting familiarised with the data. This step is crucial to capture a broader understanding of the textual data (Gale et al., 2013). Then, the two researchers who coded the drivers' responses returned to the original transcripts a few times during the analysis. The aim was to make sure the data they analysed and the results were connected. Following from that iterative process, the researchers debriefed each other to consolidate the trustworthiness of the results. The driving simulation context of the study was taken into consideration to avoid misinterpretation of the data. Trial coding helped testing the consistency (i.e., categories do not overlap) and validity (i.e., the extent to which the categories depict the material) of the coding frame created. After amending the coding frame a few times, its final version was utilised for interpretation and analysis.

RESULTS

Overall, 78 transcripts were extracted. Similar transcripts were merged to form 62 raw data themes. Table 1 presents the raw data, first-order and second-order themes with the general dimensions. For the sake of brevity, only one example illustrates each first-order theme. Four transcripts were excluded from the analysis as they were unrelated to the ransomware attack. Data saturation was completed after processing about 80% of the transcripts, yet all the transcripts were analysed.

First-Order Themes

The raw data themes were analysed by two researchers who then extracted 19 first-order themes related to the in-vehicle display (e.g., *spam*, *ransomware* and *perception*), vehicular safety (e.g., *attack mitigation*, *control* and *performance*) and drivers' personal experience of the situation (e.g., *responsibility*, *realism* and *alertness*). Results are presented in Table 1.

Second-Order Themes

The first-order themes were merged into eight second-order themes according to their conceptual resemblance. These second-order themes stemmed from the following logic: drivers saw an unusual message on the in-vehicle display that they evaluated in terms of the actions to be undertaken, and then reflected on their cognitive and affective state. The first set of second-order themes elicited by drivers was 1) *malfunction* 2) *security* and 3) *comprehension*. *Malfunction* encompassed errors (e.g. 'glitch') and defects (e.g. failure to block a spam: 'A spam popped up') from the Human-Machine Interface (HMI). *Security*-related to the cyber protection of the HMI and its exposure to external electronic threats, including the consequences on drivers' personal data (e.g. 'I had been hacked') and finances (e.g. 'bitcoin warning payment'). *Comprehension* described how long it took drivers to perceive the ransomware (e.g. 'it took me some time') and whether they understood which of the HMI or the automated driving system was affected.

Drivers also evoked the actions they performed and strategies they applied to mitigate the ransomware attack. These second-order themes were labelled 1) *behaviour* and 2) *safety*. *Behaviour* included the decisions made by drivers to lower the impact of the cyber-attack (e.g. '[I didn't] tap on the link') on driving (e.g. 'I needed to take over control'). With respect to *safety*, participants commented that they were concerned about whether resuming control was possible (e.g. 'I wasn't sure') and the driving performance (e.g. 'ransom does not affect driving performance').

The third set of second-order themes emphasised drivers' perception of their affective and cognitive state, along with the experimental environment. These three themes were labelled 1) *attitudes and feelings*, 2) *experimental environment* and 3) *situation awareness*. Comments participants made about their *attitudes and feelings* concerned how responsible they felt (e.g. 'It's the carmaker's problem'), their level of anxiety (e.g. 'it made me freak out'), trust in the vehicle-computer system (e.g. 'far less confident with the technology')

Table 1. Results of the inductive content analysis.

Raw data themes examples ($k = 62$)	1st order themes ($k = 19$)	2nd order themes ($k = 8$)	General dimensions ($k = 3$)
In-vehicle tablet interface glitch	Display failure	Malfunction	Human-Machine Interface (HMI)
A spam popped up	Spam		
I expected an audio warning	Warning		
The screen showed that I had been hacked	Malware	Security	
Bitcoin warning payment popped up	Ransomware		
Took me some time to notice the ransomware because of the word search puzzle	Perception	Comprehension	
Is it the tablet or the automated driving system that is hacked?	System status		
You only get hacked when you tap on the link so I didn't	Attack mitigation	Behaviour	Driving task
I needed to take over control of the car	Decision making		
I wasn't sure if the system could conflict with me taking back control	Control	Safety	
Ransom does not affect driving performance	Performance		Driving task
It is the carmaker's problem, not mine	Responsibility	Attitudes and feelings	Driver's insights
It made me freak out	Stress		
Far less confident with the technology	Trust		
Privacy is less important than vehicle control	Exposure		
I was more alert	Alertness	Situation awareness	
Paid more attention to the surroundings	Attentional focus		
Hacking didn't feel real	Realism	Experimental environment	
It was a simulation	Experimental setting		

and how driving safety was more important than personal data in this situation. Regarding *situation awareness*, they elicited that they were more alert and vigilant after the ransomware attack. Finally, some drivers commented on the *experimental environment* mentioning that the ransomware was bogus and that they were conscious of taking part in a study.

General Dimensions

The eight second-order themes were associated according to their similarities within the same topic. As a result, the three following general dimensions emerged from the verbatim collected during the study: 1) *HMI*, 2) *Driving task* and 3) *Driver's insights*.

With respect to *HMI*, drivers mentioned what they saw on the in-vehicle screen and how they interpreted it: a cyber-attack, a failure or a pop-up spam. Participants mentioned that they were unsure as to which of the infotainment or driving system was affected by the ransomware. It seemed to bring confusion to the situation and being engaged in a non-driving related task did not help drivers understand quickly what was going on. This relates to system transparency where users fail to understand the status of a system and its predictability (Alonso et al., 2018). *Malfunction*, *security* and *comprehension* are the components of the *HMI* dimension.

In terms of *driving task*, the ransomware led participants to ensure the automated driving system was operating satisfactorily. They also raised concerns on whether they could resume control, when they did not already. Some drivers also explained how they mitigated the propagation and malevolent effect of the ransomware by not tapping any link or button presented on the HMI (Fig. 1). The *driving task* dimension is composed of the second-order themes *behaviour* and *safety*. It focuses on the direct operations of the car's controllers and interfaces to ensure driving safety.

With reference to *driver's insights*, participants commented as to how the ransomware and the experimental settings affected their perception of the cyber-attack. This was illustrated in the *attitudes and feelings* theme. For instance, some individuals did not feel accountable for the ransomware and ignored it, despite the cyber-attack potentially affecting automated driving performance operation or personal data. This result is surprising as individuals are responsible for the driving but seem to be complacent when exposed to a cyber-attack. On the contrary, it was reported that this experience was stressful, resulting in a decrease in trust in the technology. In addition, some participants made real-time decisions to evaluate if the ransomware or the vehicle control was the most critical and urgent to deal with. The priority some participants gave to the ransomware over the control of the vehicle was insightful because it stressed the momentary shift of attention away from the driving task to concentrate on the HMI. This short period of distraction, where drivers sought seek to understand and assess the situation, may hinder road safety. However, after that assessment, some drivers declared they were more attentive to the road environment, which was reflected in the *situation awareness* theme. Finally, several participants indicated that the ransomware did not affect them a lot because of the *experimental environment*, the third

and last component of the driver's insights dimensions. This is a limitation of the present research as using a driving simulator may have decreased the perception of risk coming from the cyber-attack. As a result, some participants may not have felt threatened or concerned by the ransomware, despite sharing personal information at the beginning of the study.

CONCLUSION

The findings from the inductive content analysis suggest that individuals' principal concerns after experiencing a cyber-attack while in automated driving mode are:

1. The perceived type of the message displayed on the HMI: spam, cyber-attack or display failure,
2. Estimating the effect of the ransomware on the automated vehicle's performance and their capability to resume manual driving control,
3. Drivers' mostly negative attitudes and feelings towards the cyber-attack, even though some of them were not much concerned.

Results also shed light on the concepts affecting drivers after a ransomware appears on their in-vehicle display: trust in automation, situation awareness and driver behaviour. Responses proved diverse and went beyond the expected theme of trust that had been explored in previous quantitative research (Payre et al., 2022). Manufacturers and designers would benefit from these findings to help drivers focus on what is crucial when a vehicle is hacked: vehicle control and road safety. The aim is to avoid driver distraction and support road safety. For instance, an acoustic and visual warning system independent from the hacked HMI could be used to confirm the status of the vehicle and the automation to the drivers, call for assistance after the ongoing trip is over and avoid using the automated driving system until further notice.

ACKNOWLEDGMENT

This research was supported by the UKRI Trustworthy Autonomous Systems Hub (EP/V00784X/1).

REFERENCES

Alonso, V. and De La Puente, P., 2018. System transparency in shared autonomy: A mini review. Frontiers in Neurorobotics, 12, p. 83.

Bryant, A. (2009). Grounded Theory and Pragmatism: The Curious Case of Anselm Strauss. In Forum Qualitative Sozialforschung/Forum: Qualitative Social Research 10 (3), pp. 1–38.

Elo, S. and Kyngäs, H. (2008). The qualitative content analysis process. Journal of advanced nursing, 62(1), pp. 107–115.

Gale, N. K., Heath, G., Cameron, E., Rashid, S., and Redwood, S. (2013). Using the framework method for the analysis of qualitative data in multi-disciplinary health research. BMC Medical Research Methodology, 13(1), pp. 1–8.

Hsieh, H. F. and S. E. Shannon. (2005). Three Approaches to Qualitative Content Analysis.Qualitative Health Research 15 (9), pp. 1277–1288

Kyngäs, H. (2020). Inductive content analysis. In The application of content analysis in nursing science research, pp. 13–21. Springer, Cham.
Morris, D., Madzudzo, G. and Garcia-Perez, A. (2020). Cybersecurity threats in the auto industry: Tensions in the knowledge environment. Technological Forecasting and Social Change, 157, p. 120102.
Payre, W., Perelló-March, J., Sabaliauskaite, G., Jadidbonab, H., Shaikh, S., Nguyen, H. and Birrell, S. (2022). Understanding Drivers' Trust After Software Malfunctions and Cyber Intrusions of Digital Displays in Automated Cars. International Conference on Applied Human Factors and Ergonomics. Springer, Cham.
Schreier, M. (2012). Qualitative Content Analysis in Practice. Thousand Oaks, CA: Sage.
Trope, R.L. and Smedinghoff, T.J. (2018). Why smart car safety depends on cybersecurity. Scitech Lawyer, 14(4), pp. 8–13.
Wolf, M. and Lambert, R. (2017). Hacking trucks-cybersecurity risks and effective cybersecurity protection for heavy duty vehicles. Automotive-Safety & Security 2017-Sicherheit und Zuverlässigkeit für automobile Informationstechnik.