Modern Passenger Vehicles as Cyber Threat Source: Analyses of Surveillance Options through Smart Vehicles

Henrietta Hegyi

Óbuda University, Doctoral School on Safety and Security Sciences; Népszínház u. 8, 1081 Budapest, Hungary; hegyi.henrietta@uni-obuda.hu

László Erdődi

University of Oslo, Department of Informatics, Gaustadalleen 23B, N-0373, Oslo, Norway; laszloe@ifi.uio.no

Abstract: In recent years, the automotive industry has witnessed a paradigm shift in the capabilities and functionalities of modern passenger vehicles. These vehicles have evolved beyond their conventional roles as modes of transportation to become complex data-generating and transmitting entities. With the integration of advanced sensors, processors, and communication technologies, contemporary automobiles continually exchange data with manufacturers and third-party entities. This intricate and ongoing communication opens up new avenues for both innovation and concern, as it involves the transmission of a diverse range of data types, such as battery status, autonomous driving functionalities, traffic information, faulty image recognitions, and more. Continuous network connectivity and data transfer provides new options for cyber threat actors. Transmitting hidden data using different covert channels can promote surveillance activities through smart vehicles. Because of the huge amount of data that has to be transferred and also due to network reliability issues, specific protocols and 5G technology are recommended using in connected vehicles such as MQTT that ensures flexible and fast data transfer. This paper aims to explore the multifaceted landscape of data exchange in modern passenger vehicles and delve into the potential security and privacy implications that arise from this practice.

Keywords: automotive industry; telecommunication; cybersecurity; covert channels; mqtt

1 Introduction

In the contemporary automotive landscape, vehicles have evolved into dynamic data hubs, akin to moving data centres. This transformation is propelled by the integration of onboard Internet-of-Things (IoT) sensors and sophisticated

computing units designed to collect and process vital information related to vehicle operations [1]. These technological advancements have ushered in an era where automobiles have become intricate amalgamations of sophisticated hardware, intricate software, and intricate IoT devices—all working in concert to deliver a wide array of services.

In the realm of smart vehicles, several studies has examined the vulnerabilities associated with modern automobiles and the utilisation of the MQTT protocol as a covert channel.[2], [3], [4], [5] It is designed with a publish-subscribe model for limited network bandwidth where remote devices can have network constraints. Using MQTT for smart vehicle data exchange is obvious, but still a notable gap exists, as there has been a lack of comprehensive studies that amalgamate these two aspects. Specifically, there has been a dearth of research shedding light on how vehicles, being mobile and potentially life-threatening devices, can become susceptible to security breaches due to the covert utilisation of the MQTT protocol. The convergence of these two critical domains—automotive security and MQTT-based covert channels—remains largely unexplored, presenting a compelling avenue for investigation.

Furthermore, it is imperative to consider the nascent state of regulatory frameworks pertaining to information security in the context of personal automobiles. Currently, there exists a limited and evolving landscape of regulatory standards and legal provisions governing the mandatory security assessments and audits of automotive products or their constituent information technology elements. This regulatory void extends to the absence of stringent guidelines obligating manufacturers and suppliers to subject their vehicular and informational components to rigorous scrutiny. Due to the highly intricate nature of an automobile as a complex product, conducting the foundational risk analysis required for establishing information security measures on it proves to be exceptionally challenging. [6] The absence of mandated audits and security assessments, both at the manufacturer and supplier levels, accentuates the potential vulnerabilities in the automotive ecosystem. [7]

The data streams emanating from modern vehicles have evolved into a torrent of information that flows into the cloud-based repositories of manufacturers and third-party service providers. Notably, these data encompass not only operational parameters but also insights into drivers' behaviour, habits, and preferences. The cases of companies like Tesla, which collect extensive real-time data from their vehicles, have demonstrated the capability of leveraging this information for purposes such as vehicle performance optimization, over-the-air updates, and even advanced driver-assistance system (ADAS) improvement. [8] However, the possible utilisation of these data for surveillance purposes cannot be overlooked. Recent events, exemplified by controversies surrounding the use of data by entities such as the Chinese government in relation to surveillance of government buildings, raise questions about the extent to which modern passenger vehicles can inadvertently become tools of surveillance. [9]

2 Automotive Industry and Telematics

Telematics data analysis associated with the automotive industry has garnered significant attention lately. Connected cars provide continuous data about internal behaviour such as speed, location or maintenance requirements. This information can be used in real-time analysis to improve overall safety, reduce cost or improve performance of commercial vehicles.

Telematics data [10] encompasses a wealth of information collected from vehicle information systems. These data capture various measurements, either at regular intervals (e.g., mileage and speed) or in response to system changes (e.g., error codes). Telematics data serve as a pivotal tool for gaining detailed insights into the mobility patterns of individual drivers.

Analysing telematics data can be useful to obtain the driver's behaviour. For example, Bergman et al. [10] focused on the examination of acceleration data in the context of a Swiss car retailer. The analysed dataset was derived from speed data collected between 2019 and 2021 with rich granularity. The objective of the analysis was to tailor its product offerings to individual customer needs more effectively. Specifically, the aim was to identify both typical and atypical usage patterns of specific car models and variants. This endeavour is intricately tied to discerning driver behaviour, a task made possible through the analysis of telematics data. Customers who grant consent for their personalised data to be processed enable the creation of enhanced customer profiles. Each driver's dataset comprised approximately 50,000 data points, enabling in-depth analysis and visualisation of their driving behaviour. These profiles, in turn, empower sales personnel to provide more informed recommendations during the future vehicle purchase decisions. It was concluded that telematics data can be harnessed to gain insights into driver behaviour and vehicle usage, ultimately benefiting the automotive industry.

Despite the growing availability of telematics solutions, the implementation of live production systems based on machine learning models remains relatively limited. Notably, the insurance industry has embraced telematics data, using it as a foundation for calculating usage-based premiums.

Due to the growing significance of telematics data, more and more research [28] [29] exemplifies the practical application of telematics data analysis in the automotive industry, showcasing how such data can be leveraged to enhance customer profiling and refine product offerings. This underscores the growing significance of data-driven insights in shaping the future of the automotive sector.

2.1 Commonly Used Protocols

Modern passenger vehicles communicate with internet-accessible servers through a variety of protocols, facilitating the exchange of data between the vehicle's onboard systems and external entities. These protocols often include HTTP (Hypertext

Transfer Protocol) and HTTPS (Hypertext Transfer Protocol Secure) for web-based communication, allowing vehicles to send and receive information securely over the internet. Additionally, MQTT (Message Queuing Telemetry Transport) is commonly used for efficient and lightweight data transmission, ideal for real-time monitoring and control. These communication protocols play a pivotal role in enabling the seamless flow of data between vehicles and remote servers, supporting features ranging from remote diagnostics and software updates to cloud-based analytics and beyond. However, as the scope and sensitivity of data transmitted continue to expand, ensuring the security and privacy of these communication channels becomes increasingly critical.

A. Ullah et al. [11] explore the expanding realm of connected vehicles and the inherent challenges associated with ensuring secure and reliable communication. The study investigates the protocols and architectures employed for vehicle-to-vehicle (V2V) and vehicle-to-infrastructure (V2I) communication, shedding light on the potential vulnerabilities and threats within these systems. By highlighting both the solutions and challenges, this research serves as a foundational work in understanding the communication protocols crucial for modern vehicle connectivity. S. Kumari et al. [12] presents a comprehensive review of the Internet of Things (IoT) connectivity paradigm in the context of vehicles. This research examines the integration of IoT principles in vehicular networks, emphasising the diverse data types transmitted, ranging from safety-critical information to infotainment data. The study critically assesses the communication protocols employed, including HTTP, MQTT (Message Queuing Telemetry Transport), and CoAP (Constrained Application Protocol), while also addressing the security and privacy implications inherent in the transmission of sensitive vehicular data.

The selected studies underscore the depth and breadth of research concerning data communication, security, and privacy in modern passenger vehicles. These investigations contribute valuable insights into the protocols, architectures, vulnerabilities, and potential solutions related to vehicular data transmission. As the automotive industry continues to evolve, the findings from these studies provide a solid foundation for developing strategies to enhance the security and privacy of connected vehicles. It is of considerable interest to ascertain the reasons behind MQTT's ascension to prominence within the automotive industry. Several original equipment manufacturers (OEMs) and suppliers have offered connectivity solutions within their vehicles for an extended period. The initial generation of these connectivity solutions frequently relied upon Short Message Service (SMS) and Hypertext Transfer Protocol (HTTP). Regrettably, these technologies were not purpose-built to cater to the demands of the connected car paradigm. HTTP, for instance, relies on stable network connectivity, whereas connected vehicles operate within mobile networks characterised by their inherent unreliability. SMS, on the other hand, fails to provide the real-time messaging experience essential for the seamless operation of connected car services. For instance, in the case of one company, the act of unlocking a car door using SMS and HTTP consumed an

excessive duration of over 30 seconds. Such delays are patently unacceptable to consumers who anticipate immediate responsiveness within their digital interactions. Furthermore, SMS and HTTP incurred exorbitant costs due to their substantial network bandwidth requirements. Therefore, one of the ideal solutions for connected cars is the MQTT protocol that was also analysed with details by, e.g. HiveMQ [13].

3 Covert Channels and MQTT

Covert Channels Analysis is a specialised field within computer security and network security. The concept of "covert channels" has its origins in the field of computer security and information assurance. It dates back to the mid-20th Century when the development of computers and digital systems began. The initial pioneers in investigating "covert channels" were American military and intelligence professionals during the Cold War era. In the late 1970s and early 1980s, research on "covert channels" gained increasing prominence in the domains of computer security and intelligence activities. During these years, the first methods and techniques for detecting covert communication and safeguarding against such channels in computer systems were developed. [14]

Its primary objective is to identify and analyse hidden or covert communication mechanisms that malicious actors may employ to transmit information illicitly or conceal secret content. Covert channels refer to communication channels that deviate from intended and authorised communication paths, exploiting various operational aspects or vulnerabilities within a system or network. These channels are often concealed and difficult to detect, potentially facilitating activities such as data leakage or unauthorised access to a system. In the realm of Covert Channels Analysis, researchers and cybersecurity professionals strive to uncover and document such clandestine communication mechanisms. They also develop defensive strategies and tools to mitigate their risks. The ultimate goal is to understand how these covert channels operate and to establish measures for prevention or detection within computer systems and networks. Because of this, Covert Channels Analysis plays a critical role in information security, aiming to identify and eliminate covert and illicit communication mechanisms to safeguard the integrity and confidentiality of computer systems and networks.

3.1 Covert Channels Analysis Implemented to MQTT

Given the increasing data exchange between passenger vehicles and telecommunications networks, where data traverses the internet and reaches various stakeholders, safeguarding the security of different network layers becomes paramount. With the advent of 5G, security-by-design has been a central focus for standard developers, incorporating various modern tools to fortify network layers

(however, the sheer variety of settings and flexible programmability can potentially be exploited). [15] Despite these efforts, popular protocols like MQTT have not seen substantial security advancements in their latest versions, even in MQTT v5. Consequently, assuming that 5G networks inherently guarantee data security for vehicle communication would be a misconception, as protocols themselves can still exhibit vulnerabilities. [16]

In our investigation, we build upon the research conducted by Velinov et al. [17], who extensively explored MQTT-based covert channels and successfully identified 16 such channels within the MQTT protocol. Their comprehensive study serves as the cornerstone of our analysis. The MQTT-based covert channel system model investigated in their research encompasses a covert sender (CS) and one or more covert receivers (CRs). It is noteworthy that the model allows for the possibility of multiple covert senders, although our present study primarily focuses on a single CS. Furthermore, MQTT's inherent versatility permits any client to assume the role of a publisher, fostering bidirectional communication and facilitating group-based communication among users. This adaptability enables CS and CRs to operate within the same network or across disparate networks. Crucially, Velinov et al. [17] introduce a pivotal distinction between two specific submodels: Direct Covert Channels (DCC) and Indirect Covert Channels (ICC). In the Direct Covert Channels submodel (DCC), CS directly communicates with CRs, presenting two distinct possibilities: either the broker serves as the CS (DCCa), or the broker assumes the exclusive role of the CR (DCCb). Conversely, in the Indirect Covert Channels submodel (ICC), CS engages in indirect communication with the CRs through the intermediary role of the broker. This configuration eliminates direct interaction between CS and CRs, allowing for asynchronous covert sending and covert receiving processes. In this scenario, the broker operates as a proxy node, forwarding any messages it receives, all while remaining oblivious to the concealed data exchange.

Covert channels can be established in MQTT by exploiting various protocol features and message characteristics. Some common areas to investigate include:

- Topic Names: MQTT relies on topic names to route messages. Covert channels can be established by encoding information within topic names, which may go unnoticed if not thoroughly examined.
- QoS Levels: Different QoS levels in MQTT (0, 1, and 2) provide varying degrees of message reliability. Attackers may use variations in QoS levels to create covert timing channels.
- Payload: MQTT messages contain payloads that can be manipulated to convey hidden data. Attackers can encode information within message payloads to establish covert channels.
- Retained Messages: MQTT allows the retention of messages by brokers, which can be misused to create channels for unauthorised data access.

According to the above-mentioned sources, to implement covert channel analysis in MQTT, security professionals should follow these steps:

- Traffic Monitoring: continuously monitor MQTT network traffic to detect anomalies, such as unusual topic patterns, message sizes, or QoS level variations.
- Payload Inspection: scrutinise message payloads for hidden information or unusual encoding techniques.
- Pattern Recognition: utilise pattern recognition algorithms to identify suspicious message patterns that may indicate the presence of covert channels.
- Behavioural Analysis: analyse the behaviour of MQTT clients and brokers for unusual communication patterns or deviations from expected norms.
- Security Policies: establish and enforce MQTT security policies that specify allowed topic structures, QoS levels, and payload formats.

In our study, we examine the exploitation of payload-based covert channels among those previously introduced. Specifically, we focus on altering the content of MQTT messages, aiming to maintain the appearance of normal message operation while enabling the assembly of a hidden message at the receiving (subscriber) end when specific patterns are known. This methodology draws upon the field of steganography, which involves concealing one type of data within another, such that the extracted message remains imperceptible to unauthorised observers within the given data type. Additionally, the choice of topic names plays a pivotal role in veiling the true nature of the data. MQTT employs topic names to route messages to specific subscribers. By opting for topic names that imply a connection to accelerometers or other related sensors, even when the payload indeed contains geolocation data, further obfuscation can be achieved. For instance, a topic like "acceleration/sensor01" could be employed, giving the impression that accelerometer data is being transmitted.

3.2 Commonly Used Data of Smart Vehicles

Due to the sensitive nature of the subject matter, the available data have not shed light on the precise data transmission mechanisms of passenger vehicles. However, various independent data privacy organisations do provide some valuable sources on the topic. Caltrider et al. [18] defines the scope of data collection in modern cars and the implications for their privacy and security. The importance of responsible data handling practices in the automotive sector was also highlighted. Data collection practices of modern automobiles and the destinations of this collected data was published on the Mozilla Foundation's website in 2023.

The researcher provides insights about the diverse range of data collected by today's cars, which includes information related to vehicle performance, location, driving

habits, and even user interactions with infotainment systems. They elucidate the various sources of this data, such as sensors, GPS systems, cameras, and onboard computers, highlighting that data collection is not limited to a single source but is a result of the synergy among multiple components within the vehicle.

The research also sheds light on where the collected data goes after being gathered by the vehicle. It points out that this data is often transmitted to the car manufacturer's servers or third-party service providers, leading to potential privacy concerns. The primary purposes of data collection include improving vehicle performance, enhancing safety features, and providing convenience features for drivers and passengers. However, it also underscores the importance of understanding the extent to which data is used for advertising and marketing purposes.

Based on the available information and previous research we identified several communications and data that are typical or optional in smart vehicles.

- GPS data is necessary to track the position of the car. Relevant traffic information can only be provided from remote if the car shares its actual position. This data is typical in many smart phone applications when the user enables the app to use GPS unconditionally.
- Car sensor data provides information from all the equipment inside the car. The temperature of the engine or the brakes as well as the tire pressure and many other internal sensors regularly report their status including possible error messages. All these sensor data can be stored locally, but the trend is to store these data remotely to avoid manipulation with the data by the owner (second hand car sales). We believe that a significant amount of this data will be stored remotely in the near future in well controlled databases, that also means that all this data has to be continuously sent over the air.
- Battery cell information is crucial data for electric vehicles. Battery technology develops at a rapid pace. The real value of an electric car is significantly influenced by the battery performance such as the charging speed or the consumption as well as battery degradation. To follow the battery performance the car should report the battery status regularly.
- Communication with the owner. Smart cars can report information to the owner such as if the window remains open or if the car alarm is activated. Furthermore, the owner can use the car smart phone app to remotely turn on windscreen heating, etc. This requires remote communication through a third party component.
- Software update is rather asynchronous communication. The owner initiate the update and the car receives and install the new plugins or software updates.
- Insurance data is related to driving habits. Although it practically does not exist and is only planned for Tesla vehicles now, we assume that the

benefits of such a solution will make it popular in the near future. This would require to continuously send relevant amounts of data about the driving details such as speeding, dangerous brakes, situations when the car had to intervene, etc.

- Self driving data is considered the richest data source. The self driving module continuously improves by learning more and more traffic situations and exceptions. A situation that was misunderstood by the self driving module has to be evaluated and studied to help the software performance improve. Such a function requires sending pictures taken by the camera of the car to understand picture recognition failures.

All this data transmission provides a theoretical option for the threat actors to use the continuous information flow to hide hidden data in it. We analysed different options that a threat actor might be interested in. Based on the covert channel bandwidth the threat actor might want to send:

- relatively small amount of information, e.g. regular data about the position of the vehicle
- bigger amount of data, e.g. voice samples recorded inside the car
- huge amount of data e.g camera pictures taken by the car

This research focuses on analysing the characteristics of the different options such as sending pictures in sensor data. For all options we consider that the car software is compromised, so the focus is only on the command and control covert channel and not other parts of the cyber kill chain such as the weaponization or exploitation.

4 Modelling of Covert Channel Usage Implemented to Smart Vehicles Using the MQTT Protocol

In this chapter, we embark on a modelling journey that leverages MQTT infrastructure to send numeric data (for example temperature, acceleration or coordinates) from a publisher to a broker. The data transmitted aligns with the data types typically sent by a standard smart car (or any IoT equipment sensors).

The methodological foundations of steganography we used were guided by the approaches Jessica Fridrich, [19] Mehdi Kharrazi, [20] and Neil F. Johnson [21]. However, to perform the specific modelling, we also leveraged results and practical techniques from the field of data science. These are referenced within the relevant sections.

It is crucial to acknowledge that data generated by vehicles can accumulate in substantial volumes, and the frequency and volume of data transmission can significantly impact network load and data capacity. Therefore, striking a balance between the frequency of data transmission and the utility of the information is

paramount to ensure that thes data remain meaningful and useful. Understanding and optimising data transmission in the context of automotive telematics is essential for both vehicle performance monitoring and ensuring the safety and efficiency of operations. Consequently, the development and deployment of telematics systems should consider the specific requirements of each use case to strike the right balance between real-time data updates and the practicality of data handling.

In our modelling experiment a publisher was configured to transmit temperature and related data to a designated broker. The data transmitted via MQTT closely emulated the acceleration data typically sent by a smart car. Such data is vital for monitoring vehicle performance and ensuring passenger safety. By mimicking this data stream, we aimed to create a realistic testing environment for our experiments.

4.1 Sending Visual Information Using Covert Channel of MQTT

We delve into the data science realm, emphasising the importance of transforming various images into numerical representations. Such transformations enable the application of various machine learning techniques to significantly enhance image quality. Notably, this process is instrumental in the restoration of old and blurred photos, as demonstrated in [22] that we used as a reference. During our modelling experiment, we made the assumption that the car's information system is equipped with an application capable of converting a photograph taken by the vehicle into numerical data as described below. Subsequently, this numerical data, in conjunction with sensor-generated information, is transmitted to a broker.

We set out to transform an image of the Death Star, a fictional space station from the Star Wars franchise, into a numeric sequence that could be perfectly reconstructed. This intriguing endeavour required precision and attention to detail.

In our experiment, each pixel of an image was represented by RGB (Red, Green, Blue) values. These values can be expressed either in hexadecimal format or as triples of decimal values ranging from 0 to 255. This representation allowed us to capture the intricate details of the image.

A high-quality photograph typically consists of numerous pixels. For instance, a photograph with dimensions of 900 by 1260 pixels contains a total of 1,134,000 pixels. In terms of decimal values, this amounts to a staggering 3,402,000 individual data points.

According to our experiments, the resolution required for clearly identifying a person 30 metres away from a CCTV camera depends on various factors, including the size of the person in the frame, the quality of the camera's lens, and the level of detail you want for identification. However, as a general guideline, a minimum resolution of 1080p (1920x1080 pixels) is often considered a standard for facial recognition in surveillance applications.

To put the data volume into perspective, let's consider an IoT (Internet of Things) device that generates one unit of data per second, such as acceleration readings. Transmitting the aforementioned data volume would require a significant amount of time. In essence, it would take approximately 3,402,000 seconds that is about 39,375 days to transmit this data continuously at a rate of one unit per second.

Storing pictures as pixel arrays is not as efficient as using picture compression. Higher compression can lead to quality loss. An average JPG images can compress the file to 10% of the original size with only a little loss of quality. Converting a JPG image to a series of numbers efficiently typically involves a process called image digitization or image quantization. [23], [24] This process essentially converts the image's pixels, which are represented by colour values, into a numerical format. Considering 10:1 compression rate, one JPG image would require around 4 days to send. JPG is able to provide even more efficient compression up to 50:1. However, this compression rate would damage the quality in a way that would make face recognition impossible. Therefore, in our analysis we ruled out such an option. On the other hand, JPEG2000 supports both lossless and lossy coding this format could work well for such task.

Other picture compression algorithms can provide lossless compression. A GIF image average ratio is about 4:1, PNG images can reach the same ratio as JPG without quality loss. Based on this, we considered 4 days for one image without quality loss.

One unit of data per seconds might be realistic for car telematics such as engine temperature or acceleration data but it can be too low if we consider more data exchange oriented services such as insurance car tracking, advanced car monitoring or advanced driving assistance functions. For advanced car tracking we considered 4 units of data per seconds. This assumption means that one picture can be sent in one day in a covert channel considering compromised control software of the car.

As it was previously discussed, self-driving functionality might require sending pictures regularly taken by the car to report misunderstood traffic situations. Hiding pictures in pictures can be much more effective than hiding picture pixels in sensor data. Considering 5% useful data in a picture, one hidden picture can be sent in 20 other pictures about misunderstood traffic situations. This significantly increases the covert channel transmission speed that can go up until several pictures per day in case of frequent traffic misunderstanding.

In the worst case the self-driving tool can send the dedicated picture as a misunderstood traffic situation – that is falsely labelled – as it is. In that case no cover channel is needed, only an intentional false picture recognition.

In conclusion, our experiment highlights the versatility of MQTT infrastructure in facilitating data transmission. With well controlled telematics data we concluded that the car is only able to send only one picture every 4 days in covert channel, but as functionality increases such as with self driving picture sending the cover channel

transmission speed can easily be increased very significantly. Table 1 shows our summary about sending pictures in covert channel

Table 1
Summary of visual data sent in covert channel

	Picture as bitmap	10% picture compression
Average telematics	one picture in 40 days	one picture in 4 days
Advanced car tracking	one picture in 10 days	one picture in every day
Self-driving data	several pics per day	several pics per day

4.2 Sending Audio Information Using MQTT Covert Channels

The widespread use of digital data in various real-life applications has necessitated the development of effective security measures. Steganography techniques have emerged as a means to achieve efficient data secrecy. Within this context, innovative and versatile methods for audio steganography have been proposed. The primary objective of steganographic systems is to establish secure and robust mechanisms for concealing a significant amount of confidential data.

Given the prevalence and popularity of digital audio signals, they have become an attractive choice for conveying sensitive information, especially in the context of voice communications. Because of that, digital audio steganography has gained prominence in concealing data within emerging telecommunications technologies such as voice-over-IP and audio conferencing. The diversity of steganographic criteria has led to a wide range of system design techniques.

Djebbar et al. [25] conducts a comprehensive review of contemporary digital audio steganographic methods and assesses their performance based on indicators of robustness, security, and data-hiding capacity. Additionally, a classification of steganographic models based on their role in the embedding process, with a particular emphasis on robustness was provided. The discussion distinguishes between temporal domain techniques, which prioritise data-hiding capacity, and transform domain methods, which leverage masking properties to make embedded data noise imperceptible. Encoded domain methods, meanwhile, prioritise data integrity in challenging environments like real-time applications. A classification of techniques based on their involvement in the voice encoding process was also analysed, providing a means to assess their robustness. Performance evaluations, encompassing imperceptibility and steganalysis, are conducted for the reviewed techniques. The study reveals that the frequency domain is often preferred over the temporal domain, and music signals are identified as superior covers for data hiding in terms of capacity, imperceptibility, and undetectability. Ultimately, the paper

underscores the diversity and abundance of existing audio steganography techniques, which expand the range of potential applications. The choice of one technique over another depends on specific application constraints, including requirements for data-hiding capacity, data security, and resistance to potential attacks.

As the quantization of sounds and audio steganography represent highly intricate fields, our model adopts a simplified theoretical approach, grounded in the study of real-world examples. In the development of this model, we considered existing models [26], [27] used during audio encoding and decoding, enabling us to quantize and transform audio data in a manner that maximises efficiency while preserving quality and clarity of sound during decoding.

To transmute a 3-minute audio file into a linear array of numerical values, amenable for clandestine transmission via the MQTT protocol under the guise of conventional numerical data, one must engage in the process of audio digitization. This process entails meticulous sampling of the audio signal at regular intervals and its subsequent quantization into discrete numerical representations. The process for converting an audio file into numerical data involves several crucial steps. First, the audio file is imported, with formats like WAV or PCM being preferred due to their compatibility with digital manipulation. The selection of a sampling rate, such as 44.1 kHz, would be a reasonable decision for normal voice as it governs the pace at which the audio signal is sampled. On the other hand, the quality of 44.1 kHz is not necessarily needed for hidden voice transmission. Sampling involves the systematic extraction of samples from the audio waveform at uniform intervals. Each sample corresponds to the instantaneous amplitude of the audio signal at a specific moment in time. Quantization is the next critical phase. Choosing an appropriate bit depth is paramount, as it significantly affects data resolution. For instance, opting for a 16-bit quantization scheme provides 65,536 distinct numerical values. In this step, a mapping procedure transforms the amplitude values of the sampled audio into discrete numerical equivalents aligned with the chosen bit depth. This intricate operation converts continuous amplitude data into quantized numerical representations. Subsequently, the discretized numerical data is organised into a one-dimensional array or enumerated list, setting the stage for MQTT payload transmission. This entails using an MQTT client library compatible with the chosen programming language to clandestinely disseminate the payload in an MQTT message.

Should the need arise to reconstruct the original audio signal at the recipient's end, a reverse operation is executed. This process involves undoing the quantization procedure, converting numerical values back into audio samples, thus enabling the faithful reconstruction of the audio file. In our model, we use only 5 kHz as sampling rate and analyse a 3-minute long speech sample. With these foundational parameters in mind, the computation of the total number of audio samples constituting the 3-minute audio file is effectuated as follows:

Total Samples = Sampling Rate x Duration

Employing a customary sampling rate of 5 kHz:

Total Samples = 5000 samples per second x (3 minutes x 60 seconds per minute)

Total Samples = 900,000 samples

Given that each individual sample is delineated via 8 bits (16 bit quantization scheme is not needed) (equivalent to 1 byte), the determination of the cumulative data volume in bytes emerges as:

Total Data (in bytes) = Total Samples x 1 bytes per sample

Total Data = 900,000 bytes

In more commonplace units of measurement: Total Data approximates 0.9 megabytes (MB).

We considered 1 byte as one number. To calculate the time it takes to send 900,000 numbers through MQTT protocol at a rate of 1 number per second, we can use the following formula:

Time (in seconds) = Total Numbers / Sending Rate

Time = 900,000 numbers / 1 number per second = 900,000 seconds

Now, we can convert seconds to other time units if needed. For example:

In minutes: 900,000 seconds / 60 seconds per minute ≈ 15,000 minutes

In hours: 15,000 minutes / 60 minutes per hour ≈ 250 hours

In days: 250 hours / 24 hours per day ≈ 10.41 days

For one data unit per second an estimated 0.9 megabytes of 8-bit data packets would be imperative for the encoding of a 3-minute audio file, based on a sampling rate of 5 kHz, and using an 8-bit quantization scheme. It is prudent to acknowledge that this computation remains a simplified rendition, devoid of considerations regarding any additional data overhead incurred by the MQTT protocol or potential encodings/compression methods deployed during the transmission process.

Similarly to our picture sending algorithm we considered more network intensive services such as advanced car tracking with 4 data units in every second. This approach also means 3 days to send a 3 minute audio sample.

Latest research on voice cloning [31] showed that only 3 seconds of voice recording is enough to clone someone's voice. Voice cloning can be one of the most popular phishing techniques in the future, having a 5 seconds sample from the owner's voice or from any other passengers can be relevant for the attackers. 5 seconds audio length significantly decreases the necessary time for sending the voice sample.

In addition, self-driving functions with the reporting of falsely recognized traffic situations opens up the options for voice sending more effectively. Embedding voice samples into pictures significantly decreases the time required for sending a voice sample. In Table 2 we summarised our result for sending audio.

Table 2
Summary of audio data sent in covert channel

	5 seconds audio	3 minutes audio
Average telematics	4 hours	10 days
Advanced car tracking	1 hour	2.5 days
Self driving	within a few minutes	within few hours

4.3 Sending Location Data Using MQTT Covert Channels

Sending geolocational data through the MQTT protocol in a clandestine manner, masking it as, for instance, accelerational or temperature data, is indeed a feasible endeavour. The MQTT protocol, known for its adaptability and versatility, imposes no stringent constraints on the nature of transmitted messages. However, such an endeavour necessitates several considerations. Latest research [30] shows that identifying two sensitive locations for one person can easily lead to identifying a person in special jobs. Knowing the physical location of a connected vehicle and its owner can have the same consequences as tracking GPS data of a smart phone through phone apps.

For sending location data in covert channels the challenge lies in representing geolocational data within the message payload in a manner that conceals its true nature, lending it the appearance of other data types. This might involve encoding latitude and longitude coordinates as numerical values or employing an encoding scheme that simulates the characteristics of accelerometer readings. The essence of covert transmission is in making the geolocation data inconspicuous within the message content.

A person's physical location on the Earth's surface can be expressed using coordinates, typically involving latitude and longitude values. These coordinates represent angular measurements that define a specific point on the Earth's surface. The precision of expressing someone's location through coordinates depends on several factors, including the number of decimal places used in the coordinates.

Latitude and longitude coordinates are typically expressed in degrees, with additional decimal places for increased precision. A coordinate may include decimal fractions of a degree, such as minutes and seconds, for finer accuracy.

The number of decimal places in the coordinates determines the level of precision. For example:

- A coordinate with one decimal place (e.g., 40.1° N, 75.2° W) provides a rough location, accurate to approximately 11 kilometres or 7 miles.
- Coordinates with two decimal places (e.g., 40.12° N, 75.24° W) narrow the location down to about 1.1 kilometres or 0.7 miles.
- Three decimal places (e.g., 40.123° N, 75.246° W) improve precision to roughly 110 metres or 360 feet.
- Four decimal places (e.g., 40.1234° N, 75.2468° W) offer even greater accuracy, around 11 metres or 36 feet.

To express someone's location with high precision, a coordinate may include more decimal places. However, it's essential to consider practical limitations, as excessive decimal places can introduce computational complexities and may not be necessary for most applications.

The payload of MQTT messages serves as ideal material to send covert geolocational data because of it being numerically easily writable information. Therefore, no need to apply any advanced steganographic methods. The key here is to ensure that, to an observer, the data appear as if it were, for instance, typical accelerometer data.

5 Results of the Modelling Experiment

In this research, we have conducted an investigation into the covert transmission of three distinct data types (images, audio recordings, and location data) using the MQTT protocol.

In our advisory model we assumed that the main software of the connected vehicle is compromised so any data coming from telematics or the car applications is available locally for the attacker. Our research focused on the covert channel using the connected car’s communication, more concretely the possible bandwidth of the covert channel by analysing different communication types of the connected vehicle. We assumed that the speed of the transmission is only limited by the functionality and not the theoretical network transmission capability (we considered fast 5G network connected cars).

For images and audio recordings, the conventional data exchange, embedding hidden pictures and audio samples directly into message payloads of average telematics data did not prove to be an efficient approach. This limitation primarily stems from the substantial volume of data associated with the conversion of these media types into simplified numerical representations. Transmitting such

voluminous data under real-world conditions would necessitate unreasonably extended timeframes.

On the other hand, more advanced functionality such as advanced car tracking or self driving methods can provide enough bandwidth for the threat actors to send hidden data in covert channels.

Geolocation data has exhibited a higher level of adaptability to steganographic techniques in all types of data transmission. It lends itself well to covert transmission, rendering it indistinguishable from alternative data categories, such as data originating from accelerometer sensors, thanks to subtle obfuscation mechanisms.

It is evident from our findings that vehicles are increasingly capable of transmitting larger quantities of data through telecommunication networks, driven in part by advancements such as 5G technology. It is imperative to note that data transmission techniques that may currently appear inefficient due to their perceived sluggishness could potentially become highly efficient in the future. This is attributable to the growing affordability and ease of data transmission, along with the ever-increasing speed and simplicity with which it can be accomplished. Consequently, the timeframes described in this study may significantly shorten, making the utilisation of these channels increasingly worthwhile.

Conclusions

In our study, we have made observations and drawn conclusions regarding the security of passenger vehicles in the context of their data transmission capabilities through telecommunication networks, particularly facilitated by technologies like 5G.

First and foremost, we have identified a significant gap in the existing IT security regulatory framework applicable to automotive manufacturers. It has become apparent that the current regulations do not adequately ensure the information security of passenger vehicles as products. Moreover, our research has highlighted a distinct lack of comprehensive studies concerning the types and quantities of data transmitted by passenger vehicles in contemporary contexts. Furthermore, our investigation has revealed the potential for protocols such as MQTT to serve as covert channels for the transmission of confidential information. While MQTT may not inherently lend itself to the covert transmission of large data volumes, it can be effectively utilised for concealing simpler data types through the use of appropriate steganographic tools. In light of these findings, we propose that further exploration of this domain is both necessary and worthwhile. Additionally, enhancing the transparency of the information technology systems employed by the automotive industry could play a crucial role in bolstering security by facilitating the early detection of vulnerabilities. In conclusion, based on our current knowledge, it is conceivable that data can be surreptitiously transmitted from vehicles without the knowledge of either the vehicle owner or the service provider. This underscores the

importance of continued research and vigilance in addressing emerging security challenges in the automotive sector.

References

[1] Dorri, A., Steger, M., Kanhere, S. S., & Jurdak, R. (2019) A blockchain-based solution to automotive security and privacy. In Blockchain for Distributed Systems Security (pp. 95-116)

[2] N. Hussein and A. Nhlabatsi, "Living in the Dark: MQTT-Based Exploitation of IoT Security Vulnerabilities in ZigBee Networks for Smart Lighting Control," IoT, Vol. 3, No. 4, pp. 450-472, Nov. 2022, doi: 10.3390/iot3040024

[3] M. Praveen, A. Raza, és M. Hasib, "Open-Source Security Testing Tools for IoT Protocols - MQTT and Zigbee," IEEE, PDF. doi: 10.1109/ASET56582.2023.10180709

[4] I. Vaccari, S. Narteni, M. Aiello, M. Mongelli and E. Cambiaso, "Exploiting Internet of Things Protocols for Malicious Data Exfiltration Activities," in IEEE Access, Vol. 9, pp. 104261-104280, 2021, doi: 10.1109/ACCESS.2021.3099642

[5] A. J. Hintaw, S. Manickam, M. F. Aboalmaaly, és S. Karuppayah, "MQTT Vulnerabilities, Attack Vectors and Solutions in the Internet of Things (IoT)," IETE Journal of Research, Vol. 69, No. 6, pp. 3368-3397, May 4, 2021. doi: 10.1080/03772063.2021.1912651

[6] A. Koubatis és J. Schonberger, "Risk management of complex critical systems," IJCIS, Vol. 1, pp. 195-215, 2005. doi: 10.1504/IJCIS.2005.006119

[7] H. Hegyi, "The Information Security of Personal Vehicles from the Perspective of Information Security Experts," JOURNAL OF SECURITY SCIENCE, Vol. 5, No. 2, pp. 47-58, 2023

[8] A. J. Hintaw, S. Manickam, M. F. Aboalmaaly, és S. Karuppayah, "MQTT Vulnerabilities, Attack Vectors and Solutions in the Internet of Things (IoT)," IETE Journal of Research, Vol. 69, No. 6, pp. 3368-3397, May 4, 2021

[9] J. Bourne, "China's Tesla restrictions expose growing concern about AVs' digital privacy," Insider Intelligence, Jun. 23, 2022 [Online] Available: https://www.insiderintelligence.com/content/china-s-tesla-restrictions-expose-growing-concern-about-avs-digital-privacy

[10] Bergmann, Svenja; Seeliger, Arne; and Cenedese, Alberto, "Visualizing Time Series Data for Early Stage Business Analytics - The Case of Vehicle Telematics Data" (2022)

[11] A. Ullah, A. H. Abdullah, I. S. Amiri, M. B. H. Said, "Connected Vehicles: Solutions and Challenges," in IEEE Transactions on Vehicular Technology, Vol. 67, No. 9, pp. 8346-8357, Sept. 2018

[12] S. Kumari, A. K. Singh, A. V. R. Reddy, S. Kumari, "IoT-Connected Vehicle for Future Networks: A Review," in IEEE Access, Vol. 7, pp. 124857-124874, 2019

[13] MQTT Standard for Connected Car." HiveMQ [Online] Available: https://www.hivemq.com/article/mqtt-standard-for-connected-car/

[14] D. E. Denning, "Secure personal computing in an insecure network," Communications of the ACM, Vol. 22, No. 8, pp. 476-482, Aug. 1979, doi: 10.1145/359138.359143

[15] A. Dutta and E. Hammad, "5G Security Challenges and Opportunities: A System Approach," 2020 IEEE 3rd 5G World Forum (5GWF), Bangalore, India, 2020, pp. 109-114, doi: 10.1109/5GWF49715.2020.9221122

[16] Mileva, A., Velinov, A., Hartmann, L., Wendzel, S., et al. (2021) Comprehensive analysis of MQTT 5.0 susceptibility to network covert channels. Computers & ..., IEEE [Online] Available: https://www.sciencedirect.com/science/article/pii/S0167404821000316

[17] Velinov, A., Mileva, A., Wendzel, S., & Mazurczyk, W. (2019) Covert channels in the MQTT-based Internet of Things. IEEE Access [Online] Available: https://ieeexplore.ieee.org/abstract/document/8890870

[18] J. Caltrider, M. Rykov, Z. MacDonald (2023) What Data Does My Car Collect About Me and Where Does It Go? Retrieved from: https://foundation.mozilla.org/en/privacynotincluded/articles/what-data-does-my-car-collect-about-me-and-where-does-it-go/

[19] Fridrich, J. (2009) Steganography in Digital Media Principles, Algorithms, and Applications. Cambridge University Press, Cambridge

[20] Chandramouli, R., Kharrazi, M., Memon, N. (2004) Image Steganography and Steganalysis: Concepts and Practice. In: Kalker, T., Cox, I., Ro, Y.M. (eds) Digital Watermarking. IWDW 2003. Lecture Notes in Computer Science, Vol. 2939, Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-540-24624-4_3

[21] N. F. Johnson and S. Katzenbeisser, "A survey of steganographic techniques," in Information hiding, May 2000, pp. 43-78 [Online] Available: https://www.academia.edu/download/11025045/22359536_lese_1.pdf

[22] "How to Convert a Picture into Numbers," KDnuggets, URL: https://www.kdnuggets.com/2020/01/convert-picture-numbers.html

[23] J. Fridrich, M. Goljan, és D. Soukal, "Perturbed quantization steganography with wet paper codes," in MM&Sec '04: Proceedings of the 2004 workshop on Multimedia and security, Sep. 2004, pp. 4-15, doi: 10.1145/1022431.1022435

[24] A. Almohammad, G. Ghinea and R. M. Hierons, "JPEG Steganography: A Performance Evaluation of Quantization Tables," 2009 International

Conference on Advanced Information Networking and Applications, Bradford, UK, 2009, pp. 471-478, doi: 10.1109/AINA.2009.67

[25] F. Djebbar, B. Ayad, K. Abed Meraim, and H. Hamam, "Comparative study of digital audio steganography techniques," EURASIP Journal on Audio, Speech, and Music Processing, Vol. 2012, No. 25, 2012

[26] N. Iwakami, T. Moriya and S. Miki, "High-quality audio-coding at less than 64 kbit/s by using transform-domain weighted interleave vector quantization (TwinVQ)," 1995 International Conference on Acoustics, Speech, and Signal Processing, Detroit, MI, USA, 1995, pp. 3095-3098 Vol. 5, doi: 10.1109/ICASSP.1995.479500

[27] Tsung-Han Tsai and Chuh-Chu Yen, "A high quality re-quantization/quantization method for MP3 and MPEG-4 AAC audio coding," 2002 IEEE International Symposium on Circuits and Systems (ISCAS), Phoenix-Scottsdale, AZ, USA, 2002, pp. III-III, doi: 10.1109/ISCAS.2002.1010358

[28] K. P. Kamble, "Smart vehicle tracking system" International Journal of Distributed and Parallel Systems (IJDPS) Vol. 3, No. 4, July 2012

[29] Nhien-An Le-Khac, Daniel Jacobs, John Nijhoff, Karsten Bertens c and Kim-Kwang Raymond Choo, "Smart vehicle forensics: Challenges and case study" Future Generation Computer Systems Volume 109, August 2020, pp. 500-510

[30] The Intercept, "American phone-tracking firm demoed surveillance powers by spying on CIA and NSA", URL: https://theintercept.com/2022/04/22/anomaly-six-phone-tracking-zignal-surveillance-cia-nsa/

[31] FreeThink, "Microsoft's new AI needs just 3 seconds of audio to clone a voice" URL: https://www.freethink.com/robots-ai/voice-cloning-vall-e