POLITECNICO DI TORINO
Repository ISTITUZIONALE

Can social media shape the security of next-generation connected vehicles?

Original
Can social media shape the security of next-generation connected vehicles? / Scarano, N., Mannella, L., Savino, A., DI CARLO, S.. - (2024), pp. 1-4. (2024 IEEE 30th International Symposium on On-Line Testing and Robust System Design (IOLTS) Rennes (FRA) 3-5 July 2024) [10.1109/IOLTS60994.2024.10616053].

Availability:
This version is available at: 11583/2993131 since: 2024-10-07T15:46:29Z

Publisher:
IEEE

Published
DOI:10.1109/IOLTS60994.2024.10616053

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright
IEEE postprint/Author's Accepted Manuscript

©2024 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collecting works, for resale or lists, or reuse of any copyrighted component of this work in other works.

(Article begins on next page)

26 November 2026

Can social media shape the security of next-generation connected vehicles?

Nicola Scarano, Luca Mannella, Alessandro Savino, Stefano Di Carlo∗,
Politecnico di Torino, Department of Control and Computer Engineering, Turin, Italy
e-mail: {name.surname}@polito.it

***Abstract*—The increasing adoption of connectivity and electronic components in vehicles makes these systems valuable targets for attackers. While automotive vendors prioritize safety, there remains a critical need for comprehensive assessment and analysis of cyber risks. In this context, this paper proposes a Social Media Automotive Threat Intelligence (SOCMATI) framework, specifically designed for the emerging field of automotive cybersecurity. The framework leverages advanced intelligence techniques and machine learning models to extract valuable insights from social media. Four use cases illustrate The framework's potential by demonstrating how it can significantly enhance threat assessment procedures within the automotive industry.**

***Index Terms*—Online Social Networks, Cyber Threat Intelligence, Automotive Security, Threat assessment**

I. Introduction

The Automotive Industry is experiencing a substantial change in its innovation drivers, moving from traditional domains to new areas such as hardware or software. Thanks to this shift, we have experienced significant updates in various car components, including advanced driver assistance, brake systems, ignition, and engine. On the other hand, the increasing complexity of vehicular systems is opening the industry to new harms that carmakers have never experienced [1]. Cyberattacks increasingly target vehicle systems, threatening the system's security and drivers' safety [2]–[4]. This phenomenon is alarming carmakers and pushing innovations in automotive-oriented cybersecurity [5].

Online Social Networks (OSNs) serve as platforms where many users share opinions and information across diverse topics via the Internet [6], as they can aggregate insights and attention in response to real-world events, transcending geographical boundaries. Automatic information gathering from openly accessible sources like blogs, publications, or social networks is called Open Source Intelligence (OSINT). This term includes all those frameworks that involve collecting and processing large amounts of data to extract knowledge from it.

This study was carried out within the SERICS - Security and Rights in the CyberSpace and received funding from the European Union NextGenerationEU (PIANO NAZIONALE DI RIPRESA E RESILIENZA (PNRR) – MISSIONE 4 COMPONENTE 2, INVESTIMENTO 1.3 – D.D. 1556 11/10/2022, PE00000014). This manuscript reflects only the authors' views and opinions. Neither the European Union, nor the European Commission, nor the Ministry can be considered responsible for them.
(*) Correspondence: stefano.dicarlo@polito.it

979-8-3503-7055-3/24/$31.00 ©2024 IEEE

Specifically, OSNs over the years have become valuable data sources for researchers who have gleaned insights into various domains like political campaigns and sociological studies [7]–[9].

Even though the automotive domain is experiencing an increase in cyber threats, there remains a conspicuous absence of Open Source Cyber Threat Intelligence (OSCTI) research within the automotive domain. This paper proposes a framework for Social Media Intelligence (SOCMINT) in automotive security, or Social Media Automotive Threat Intelligence (SOCMATI), and presents four use cases that could significantly benefit from its exploitation. The system proposed will enhance the industry's capacity to anticipate, identify, and mitigate emerging threats in an increasingly interconnected and digitalized automotive landscape.

The remainder of this paper is organized as follows: Section II reviews the current state of the art and provides the main concepts necessary to discuss the proposed approach. Section III outlines the methodologies proposed for performing SOCMATI. Finally, Section V concludes the paper and discusses possible future research directions.

II. Background and related work

OSINT leveraging openly accessible data facilitates the retrieval of significant insights. Hwang et al. [10] conceptualize OSINT within related terms: (i) *Intelligence* refers to information of espionage, collection of data for a specific purpose, (ii) *Open Source Data (OSD)* stands for unprocessed general data, (iii) *Open Source Information (OSINF)* general data partially filtered on requirements, and (iv) *Validated OSINT (OSINT-V)* representing OSINT with a high degree of certainty. In contrast, Cyber Threat Intelligence (CTI) instead refers to collecting and analyzing data related to threat data. It is a strategy widely adopted in IT systems, mainly focusing on Indicator of Compromise (IoC) [6]. While social media-based data collection is referred to as SOCMINT [11], etc. [12]. However, the existing platform (threat sharing platform or general repositories with public OSCTI reports) seldom documents and registers automotive attacks, where information remains notably scarce.

Social media platforms offer a viable resolution to the data scarcity issue, serving as prolific reservoirs of threat-related data. Platforms such as YouTube, Reddit, and Telegram are recognized as robust ecosystems harboring extensive cybercrime communities. Numerous studies [13]–[15] have

effectively harnessed CTI derived from social network data, facilitating the extraction of advanced threat information by analyzing user interactions, communications, and preferences.

However, a few studies have investigated the potential of leveraging social media data to enhance cybersecurity in the automotive sector. Oberti et al. [16] focuses solely on Twitter as a data source, presenting a comparable workflow but with limited detail on data processing techniques and methodologies. Additionally, despite providing a comprehensive overview of standards and context in the automotive domain, there appears to be a lack of understanding of the broader OSCTI field, encompassing cybersecurity and data mining. Instead, the work of Bertoglio et al. [17] suffers from inaccuracies in CTI terminology, a lack of detailed methodological discussion spanning data sources to word embedding and Machine Learning (ML) techniques, and vague definitions of solutions, with inadequate perspectives on potential applications and future work.

In this context, supporting SOCMATI must complement competencies across all the three macro areas interested: automotive, data science, and cybersecurity 1.

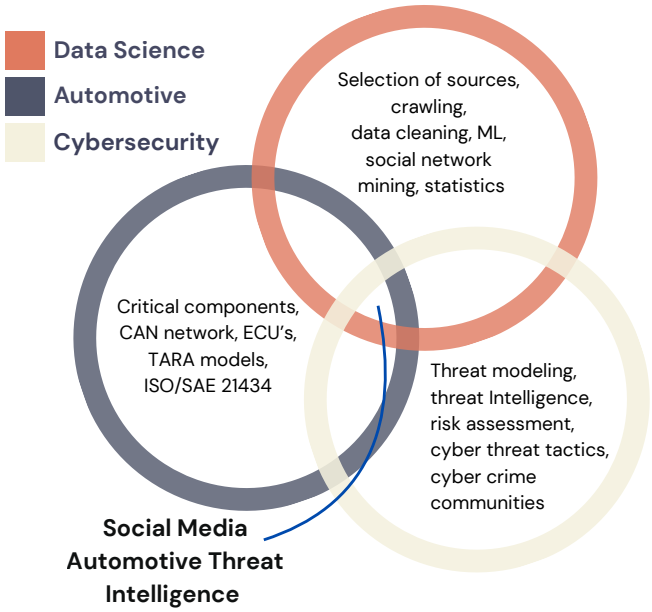

Fig. 1. This figure highlights three major competencies to design and implement the proposed framework.

For this reason, we propose a new framework that exploits the work done in the existing OSCTI field and adapts it to the needs and requirements of the growing automotive security domain.

III. Proposed Framework

The framework integrates SOCMINT into a cyber threat mitigation strategy. In particular, the data-driven approach proposed using a corpus of open data allows the automation of part of the cyber security assessment and improves its accuracy, depth, and speed. The approach proposed at this stage requires the cybersecurity expert to supervise all the phases of the data-driven approach. The framework can be modeled as a pipeline of seven phases as illustrated in Figure 2 following a V-model. The initial phase involves collecting information on vulnerabilities and potential threats, thus building a preliminary threat model. The subsequent phase focuses on defining and extracting atomic data like keywords, which is the starting point for data extraction and analysis methods. We identify and select open-source data in the third phase, concentrating primarily on social media platforms. This involves thoroughly evaluating various social media channels, groups, and chats to identify those pertinent to our research objectives. The next phase entails the application of a data science pipeline to the extracted data, leveraging ML techniques. The final phases, on the right side of the V-model, focus on exploiting the extracted knowledge. These include analyzing and visualizing the results, updating the preliminary synthesis with new findings, and exploiting the processed understanding on a specific use case.

This systematic approach ensures that the framework captures critical threat information and repeats the process multiple times, continually enhancing its threat model through iterative updates and refinements. In the following sections, the paper briefly describes the key components of our framework.

A. Data acquisition and cleaning

The proposed framework relies extensively on data, its most valuable and sensitive component. Therefore, identifying reliable data sources and implementing rigorous data cleaning, transformation, and processing pipelines are fundamental to the framework's effectiveness and reliability. Communication platforms such as Telegram, Reddit, and YouTube have emerged as significant data reservoirs for cybersecurity analysis due to their substantial engagement from cybercrime communities. Although X (formerly known as Twitter) has restricted open access to its Application Programming Interface (API) by imposing a fee, other publicly accessible social networks continue to offer invaluable data that can be leveraged for comprehensive cybersecurity research. Conversely, Meta's platforms require specific ethical safeguards (ethical committee agreement) for automatically accessing third-party data since they might present user's personal information. These platforms provide official APIs that allow researchers to gather text-based data and pertinent metadata methodically. In particular, we can categorize the data scraped into three key types:

- Textual Data: posts, comments, discussions, video transcripts, chat logs.
- Interaction Information: metrics include replies, likes/dislikes, shares/retweets.
- Metadata: timestamps, geographical locations (when available), user profiles.

Once the data is collected, the pre-processing phase is crucial for ensuring its cleanliness and standardization, which

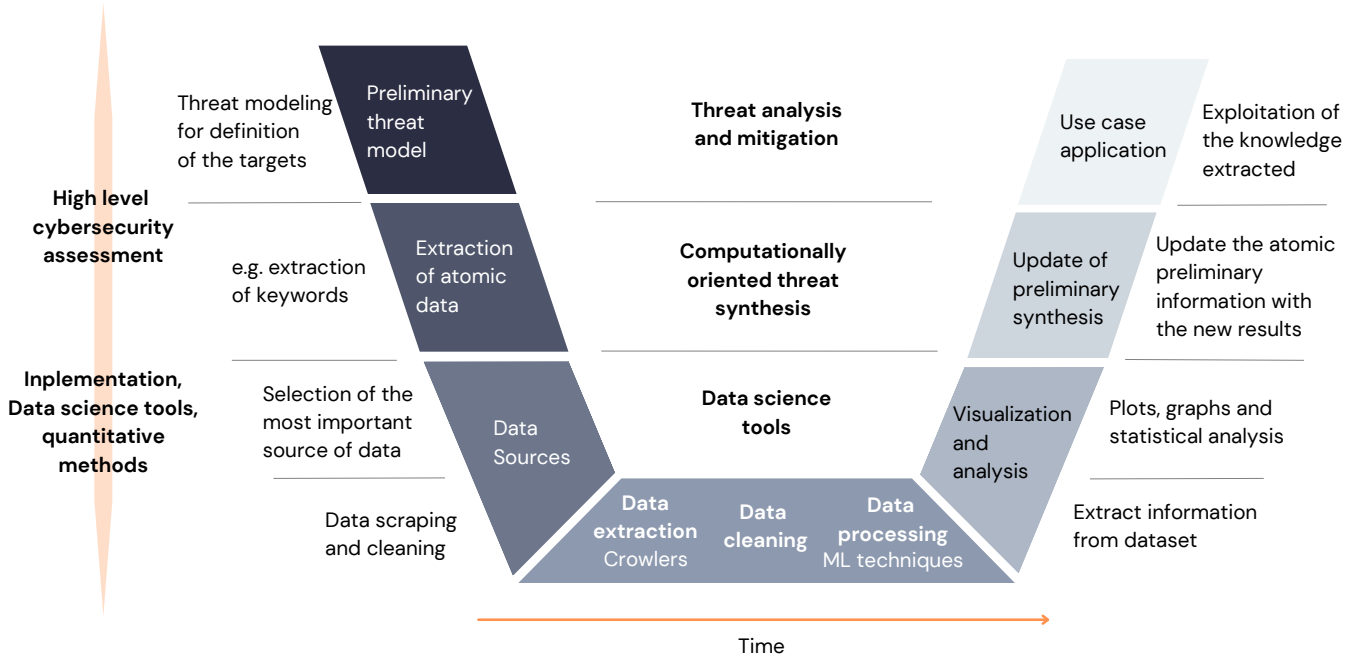

Fig. 2. The picture illustrates the V-model of the framework flow presented in the paper. The upper part of the model represents the initial and final stages of the workflow, focusing on structuring and validating the extracted threat information, such as the threat model. In contrast, the lower part of the model details the implementation steps. It begins with extracting preliminary data from the qualitative threat analysis, followed by a data science pipeline involving data extraction, cleaning, and processing. It concludes with visualizations and updates to the preliminary synthesis.

is essential for effective analysis. The processed data is then meticulously organized into structured datasets.

B. Processing techniques

The framework relies on a ML pipeline to extract detailed information from the corpus of data crawled from network sources. Based on these data types presented in Section III-A, we have selected a list of processing techniques for social media analysis and threat intelligence that have been successfully deployed.

Time Series Analysis is a specialized method for examining a sequence of data points collected over regular intervals. In this type of analysis, data points are consistently recorded at set intervals throughout a specified duration instead of being captured intermittently or randomly. Promising are the Autoregressive Integrated Moving Average (ARIMA) models used for short-term forecasting of trends in threat events and attack patterns activity and the Long Short-Term Memory (LSTM) networks for long-term trend analysis that allow capturing seasonal and temporal effects over extensive datasets instead.

Natural Language Processing (NLP) methods combine the techniques of linguistics and computer science to create algorithms that can interpret human language. They can be helpful in processing and extracting information from comprehensive corpora of text data. In particular, the applications of SOCMINT and CTI could benefit from exploiting methods to extract keywords related to a specific topic. In this context, word embedding methods (Term Frequency–Inverse Document Frequency (TF-IDF), Word2Vec, etc) and topic modeling techniques (e.g., Named Entity Recognition (NER), Latent Dirichlet Allocation (LDA), Density-Based Spatial Clustering of Applications with Noise (DBSCAN), etc) offer great potential.

Interaction Analysis exploits network analysis for text mining. Analyzing large corpora of data from social media can unveil connections between concepts, entities, or documents by representing them as nodes in a network and their relationships as edges. It also allows us to identify clusters or communities within a network, revealing cyber-criminal groups and threat topics and enhancing our understanding of the underlying structure of the data.

IV. Use cases

In this section, we present four use cases that may deeply benefit from exploiting the social media threat intelligence approach described in this paper.

1) Learn attack tactics information: Attackers use social network crime communities to exchange information about attack strategies. The ability to analyze chats, groups, blogs, or videos in an automated way can allow us to learn new tactics or update old ones. The automatic update allows carmakers and security experts to adopt or update new defense mechanisms. In this case, NLP techniques are key. In particular, it is required to detect topics inside the whole corpus of data and, after selecting the topic, extract hot words inside each single cluster. Then, using anomaly detection techniques could allow us to

detect new words related to new techniques or technologies used for a specific attack.

2) Threats trend detection and prediction: The collected data can be used to detect and predict cyber threat trends. Trends can be analyzed based on geographical data, making these predictions granular and specific to different regions. The performance of this type of analysis requires NLP capabilities alongside time series analysis methods. In this case, extracting quantitative metrics is critical to perform trend prediction. Examples of valuable data in this context are engagement and interaction metrics extracted from content on a specific topic.

3) Define threat intelligence metrics: Defining metrics to assess and extract relevant data is crucial in evaluating the threat intelligence process. Inspired by [6], which presents a set of metrics for comparing threat intelligence sources and assessing their suitability for specific purposes, this approach aims to ensure a thorough assessment of threat intelligence effectiveness. Metrics can be extracted to assess a channel's quality or the relevance of a specific search query in a YouTube search. The formal definition of metrics allows researchers to define better the extent to which the data collected can meaningfully support their intended uses.

4) Data Driven Risk Assessment: Risk assessment is a well-known mitigation strategy, already described in automotive ISO/SAE 21434 [18]. However, this is usually done by directly analyzing the platform to protect it. Introducing SOCMATI can help extract risk metrics and compute risk models, contributing to cybersecurity and safety. This use case needs the construction of merging protocols that exploit the information extracted using SOCMINT with already existing risk models. A Proof of Concept (PoC) of the use of extracted Twitter data for automatic update of the risk model is presented in [16].

V. Conclusion and future work

This paper proposed a framework that exploits CTI techniques for building a social media cybersecurity framework tailored to the automotive domain: SOCMATI. The proposed framework leverages intelligence methods and data analysis techniques to extract valuable information from social media. Several use cases demonstrate the framework's potential to enhance threat assessment procedures in the automotive industry significantly.

Future research will focus on implementing and evaluating this framework in real-world scenarios. This initial development represents only the beginning; numerous challenges remain in elevating automotive cybersecurity to match the expertise found in traditional IT sectors. A particularly urgent task is the development of ontologies to improve the understanding of attack patterns and response strategies.

In conclusion, the diffusion of OSCTI in the automotive industry is essential. It will enable the sector to anticipate better, identify, and mitigate emerging threats in an increasingly interconnected and digitalized automotive landscape.

References

[1] Mandiant, "Global perspectives on threat intelligence report," https://www.mandiant.com, 2024, https://www.mandiant.com/global-perspectives-on-threat-intelligence (accessed May 25, 2024).

[2] F. Oberti *et al.*, "Mitigation of automotive control modules hardware replacement-based attacks through hardware signature," in *2021 51st Annual IEEE/IFIP International Conference on Dependable Systems and Networks - Supplemental Volume (DSN-S)*, 2021, pp. 13–14.

[3] S. Jeong *et al.*, "Aero: Automotive ethernet real-time observer for anomaly detection in in-vehicle networks," *IEEE Transactions on Industrial Informatics*, vol. 20, no. 3, pp. 4651–4662, 2024.

[4] S. Longari *et al.*, "Cannolo: An anomaly detection system based on lstm autoencoders for controller area network," *IEEE Transactions on Network and Service Management*, vol. 18, no. 2, pp. 1913–1924, 2021.

[5] S. Misto Kirdi *et al.*, "CARACAS: vehiCular ArchitectuRe for detAiled Can Attacks Simulation," 2024.

[6] V. G. Li *et al.*, "Reading the tea leaves: A comparative analysis of threat intelligence," in *28th USENIX Security Symposium (USENIX Security 19)*. Santa Clara, CA: USENIX Association, Aug. 2019, pp. 851–867. [Online]. Available: https://www.usenix.org/conference/usenixsecurity19/presentation/li

[7] N. Aggrawal and A. Arora, "Behaviour of viewers: Youtube videos viewership analysis," *International Journal of Business Innovation and Research*, 2019. [Online]. Available: https://api.semanticscholar.org/CorpusID:202331567

[8] T. Khaund *et al.*, "Telegram: Data collection, opportunities and challenges," in *Information Management and Big Data*, J. A. Lossio-Ventura *et al.*, Eds. Cham: Springer International Publishing, 2021, pp. 513–526.

[9] J. Pastor-Galindo *et al.*, "The not yet exploited goldmine of osint: Opportunities, open challenges and future trends," *IEEE Access*, vol. 8, pp. 10 282–10 304, 2020.

[10] Y.-W. Hwang *et al.*, "Current status and security trend of osint," *Wireless Communications and Mobile Computing*, vol. 2022, pp. 1–14, 02 2022.

[11] T. Riebe, *Privacy Concerns and Acceptance Factors of OSINT for Cybersecurity: A Representative Survey*. Wiesbaden: Springer Fachmedien Wiesbaden, 2023, pp. 221–248. [Online]. Available: https://doi.org/10.1007/978-3-658-41667-6_14

[12] P. Gao *et al.*, "Threatkg: A threat knowledge graph for automated open-source cyber threat intelligence gathering and management," 2022. [Online]. Available: https://arxiv.org/abs/2212.10388

[13] B.-D. Le *et al.*, "Gathering cyber threat intelligence from twitter using novelty classification," in *2019 International Conference on Cyberworlds (CW)*, 2019, pp. 316–323.

[14] A. Bose *et al.*, "A novel approach for detection and ranking of trendy and emerging cyber threat events in twitter streams," in *2019 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, 2019, pp. 871–878.

[15] V. Behzadan *et al.*, "Corpus and deep learning classifier for collection of cyber threat indicators in twitter stream," in *2018 IEEE International Conference on Big Data (Big Data)*, 2018, pp. 5002–5007.

[16] F. Oberti *et al.*, "Psp framework: A novel risk assessment method in compliance with iso/sae-21434," in *2023 53rd Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*, 2023, pp. 60–67.

[17] L. Bertoglio *et al.*, "Machine learning and artificial intelligence boosting automotive threat intelligence," in *20th escar Europe - The World's Leading Automotive Cyber Security Conference (15. - 16.11.2022)*, 2022.

[18] ISO, "Iso/sae 21434:2021," https://www.iso.org/, https://www.iso.org/standard/70918.html).