Telemetry and Logging

David Zhai,
Computer Science and Software Engineering
University of Washington Bothell
Bothell, WA USA
Email: davidz5@uw.edu

I. Industry Trends and needs

Telemetry and logging are two concepts that are vital in monitoring and troubleshooting modern day software systems. At a basic level, telemetry aims to collect user data from a distance. This generally focuses on real-time performance metrics related to an event occurring [1]. This data can then be used to better understand issues in the software, the health of the system, and user behavior. This essentially helps product creators monitor their product once it has shipped, and offers insight into valuable metrics such as feature usage, product engagement frequency, preferences, and crash reports [2]. On the other hand, logging focuses on creating a historical record for post-incident analysis and debugging in order to help diagnose errors and code flows [2].

II. Current Solutions

Telemetry and logging are techniques used in a broad range of industries such as the automotive industry, healthcare, gaming industry, and far more. The reasoning behind this is simple - telemetry is the process of collecting vast amounts of data to be processed later, and is extremely useful when gauging the success criteria of a product. This concept is not mutually exclusive with any single industry - when data can be cheaply collected and analyzed to improve the success criteria of a product, the methods to do so are of interest to everybody.

A. Automotive Telemetry

The automotive industry uses telemetry to collect, transmit, and analyze data from vehicles. Data collection is automatically done through embedded sensors and then transmitted in real time to be processed within a centralized system for data analysis to occur [3]. Some basic metrics that are taken in the automotive industry relate to the performance of the vehicle; engine performance, fuel efficiency, speed, location, driver behavior, etc. Taking these metrics allow for automotive manufacturers to evaluate the performance of their vehicles. However, there are many concerns when it comes to telemetry in the automotive industry. Firstly, it is difficult to ensure data security and privacy, which comes down to employing reliable (and often expensive) ways of encryption and authentication. Furthermore, vehicles have become more diverse in recent years, meaning that forming an industry wide standard for telemetry data in different vehicles is a challenge [3].

B. Healthcare

Healthcare systems put effort into monitoring patient vitals. Telemetry in this case would focus on patients who have been diagnosed with a condition, or need to be monitored during recovery. Vital signs such as blood pressure, heart rate, respiratory rate, and blood oxygen levels can be measured remotely through wearable devices. For patients physically in the hospital, devices are commonly used that monitor vital signs, and alert nurses or doctors when safe thresholds are crossed [4]. Healthcare telemetry can even be seen in more commercial settings, such as in fit-bits and other wearable health devices.

C. Gaming Industry

Telemetry is vital in the gaming industry for monitoring player engagement, especially for games which have an interest in maintaining competitive integrity. For example, being able to see which game mechanics players gravitate towards, as well as how players engage with gaming content is vital for the long term health of the game. This information, if collected and analyzed correctly, should help game developers balance game play and retain their player base. Telemetry in the gaming industry is a great example of how analyzing telemetry data can be difficult and misleading. For example, say that you have a highly competitive multiplayer game where teams of five face off against another team of five individuals. Each individual has their own game character, which are usually based off predetermined skills or abilities. You might collect data on the win rates of each character, thinking that is a valid way to balance the game and maintain competitive integrity. However, many factors might influence the win rate of a character; what if a character has a lower win rate at lower skill levels, but is higher than average at elite levels of skill? What if a character has a lower win rate for the first hundred games that the average player plays it, but then suddenly skyrockets? In the gaming industry, the relevant data to collect might seem simple, but without deep analysis, it is easy to miss the point of the data, and employ patches or updates that go in the wrong direction.

D. IoT devices

IoT devices are nothing but a network of devices that are embedded with network connectivity, sensors, and software [5]. As IoT devices have the hardware (sensors) needed to collect data from their environment, as well as the network

connectivity required to transmit it, IoT devices have become a huge source of telemetry data [5].

III. Current Issues for telemetry

The issues related to telemetry are not related to the motivation in collecting data, but rather in the actual data analysis. For example, industries agree on the importance of data collection, but most of the work comes in identifying the actual data that should be collected, as well as interpreting that data. Both of these tasks are vital to telemetry, yet have huge potential for biases. For example, a team might have 100 metrics they could collect, but somehow choose the least impactful ones. In this case, they would be missing out on the most vital metrics, and may push useless features or updates. Furthermore, telemetry data needs to scale, meaning it takes a considerable amount of time to collect enough data to be meaningful. This introduces data storage issues, as storing vast amounts of telemetry data not only makes it more difficult to analyze, but also more costly. Data collection practices also need to be considered, as it takes device power to collect and send telemetry data, while also respecting privacy.

A. Current issues for logging

When it comes to logging, there are a few major challenges that the industry must address. The first is setting up a guideline on producing high quality logging code, as research suggests current logging code is written and revised in an ad-hoc fashion [6]. Another issue stems from the fact that logging code is often entangled with source code; this means that evolving systems need to figure out how to maintain and update logging code as their source code is changed [6].

IV. Critical Analysis

As it stands, telemetry and logging are widely used in industries such as healthcare, gaming, automotive, IoT, and far more. They provide a vital source of data which can help industries monitor their products for performance metrics as well as potential improvements. For industries focused more on software products, logging becomes a vital tool for incident analysis. While nobody doubts the usefullness of telemetry and logging, the costs and process associated with successful telemetry and logging are serious cons that need to be evaluated. The main issues with telemetry is that you only get the data you actively search for. This means useless data could be collected, and useful data could be missed. Furthermore, since telemetry data is more useful in higher quantities, storage considerations become a problem, which is a whole other problem. In essence, telemetry and logging in industry are two things everyone knows they should be doing, but the cost and process of doing so deter many.

V. New solutions

A. Telemetry solutions

Given that data collection and storage in an unbiased manner is the main challenge of telemetry, it could be useful to employ AI in order to create best practices for each industry. For example, a strong AI model may be able to analyze an entire codebase, understand the product, and draw upon best practices to determine which data points should be collected and stored for future analysis. While it is true that AI models are still biased due to being trained upon biased data, if the model was trained on a sufficient amount of industry code, it could still prove useful. If an AI model were somehow able to accomplish this while being unbiased, it would completely solve the data collection part of telemetry; only useful data would be collected, which also means that storage costs would go down.

B. Logging solutions

For logging, the main pain points are the fact that the industry lacks a standardized way to produce high quality logging code, as well as maintain existing logging code in evolving systems. One solution could be to train a strong AI model on source code with logging spanning many industries. The AI model could potentially analyze the patterns used by successful companies/high quality codebases. This would allow for a model to gain insight on best logging practices based upon what currently exists in the industry. It could then be trained to automatically generate logging code alongside source code.

References

[1] "Telemetry vs. logging: What is the difference?" 2015. [Online]. Available: https://sematext.com/glossary/telemetry-vs-logging/#:~:text=Telemetry%20focuses%20on%20real%2Dtime,post%2Dincident%20analysis%20and%20debugging.

[2] "What is telemetry? how telemetry works, benefits of telemetry, challenges, tutorial, and more," 2015. [Online]. Available: https://stackify.com/telemetry-tutorial/#:~:text=But%20logging%20is%20a%20tool,data%20from%20real%2Dworld%20use.

[3] "Investigating vehicle telemetry - what it is, why it matters, and what lies ahead," 2023. [Online]. Available: https://indeema.com/blog/intelligent-vehicle-telemetry-for-electric-bikes-case-study

[4] "Telemetry unit: A comprehensive guide," 2023. [Online]. Available: https://www.trustedhealth.com/blog/what-is-a-telemetry-unit#:~:text=a%20telemetry%20nurse!-,What%20is%20a%20Telemetry%20Unit%20in%20a%20hospital?,by%20a%20bedside%20cardiac%20monitor.

[5] "How telemetry powers the internet of things (iot)," 2019. [Online]. Available: https://www.hp.com/us-en/shop/tech-takes/how-telemetry-powers-the-internet-of-things-iot

[6] "Improving the software logging practices in devops," 2019. [Online]. Available: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8802761