White paper

Internet of Things:

How to improve data transmission efficiency

orange™ Business Services

Summary

01 8 tips for enhancing data bandwidth efficiency

1.1 Choose the right protocol

Data transmission protocols matter. Different protocols are better suited to different applications and uses. If you need to transmit small data packets with a high frequency (e.g. for real time purpose) MQTT or AMQP protocols will deliver better performances than other protocols like HTTP.

1.2 Get relevant data

Always evaluate which data really needs to be retrieved. When an application is designed, it is very tempting to transmit all available data, to be prepared for future functionalities. But data transmission has a cost, and the choice of the data to be transmitted can later be modified over the air using appropriate data collection and management solution. That means you can safely focus on only data that is currently relevant. For example, an eco-driving solution requires retrieving the engine's RPM or gearbox ratio but not seatbelt status.

1.3 Accuracy is not always worth it

Accurate data is essential in some cases, but only nice-to-have in others. The definition of "accuracy" itself varies depending on the industry and application concerned.

But transmission has a cost. The more accurate a value is, the larger the number of bytes required, which increases data consumption. For example, when managing an air conditioning system, it is enough to know temperature variations of 0.5° and up, since greater accuracy does not increase value. On the other hand, managing a chemical process could require taking into account the slightest temperature variations.

1.4 Trigger alarms instead of pushing data

In many situations, you can trigger alarms based on set thresholds instead of retrieving data. For example, why retrieve the fuel tank level if the compelling event is the tank being empty? Or why retrieve geo-location when you only need geo-fencing? This leads us to our next tip: local processing.

1.5 Process locally whenever you can

Another way to reduce data consumption is to embed local processes. In the example of an eco-driving service that we gave previously, you should process onboard vehicle data locally to generate an eco-driving score and then send it. Then you no longer need to send frequent RPM or gearbox ratios, just a score that can be sent once.

1.6 Choose the right format

Now that you have determined what data you really need, how can you go even further? You should focus on data serialization. Is it binary, human-readable, standardized? You should adapt your serialization to match the product phase you are currently in. For example, you can use a human-readable format such as JSON, XML or any verbose serialization format for development phases and shift to less byte-hungry formats like TLV or even binary for production.

1.7 Raw vs differential retrieval

To get a measurement, you do not always need to retrieve it in full. You often can reconstruct it, meaning that you only retrieve the difference between the previous and new values of the measurement. For example, if you need to send a set of measurements, you should index them and only send those which have changed instead of sending the whole set.

1.8 Do you need real time?

Frequency is another essential point to consider. You need to set up a retrieval timing strategy to fit your needs. Do you need real time, pseudo real time, or only best-effort reporting? The frequency issue is crucial because it affects the bandwidth your application will consume. It will also determine the transfer protocol you can use, as explained in the first part of this article.

02 Applying data collection strategies for remote monitoring of industrial assets delivers a more cost effective solution

Let's apply some of the tips we just provided to remote monitoring of industrial assets such as mobile power stations.

These power stations are sold and used worldwide, and the user wants to make sure that they operate 24 hours a day and 7 days a week. In case of an incident or outage, remote diagnostics make it possible to better prepare for an intervention by a field technician.

In this case, there are different options to collect the data. Let's review some of these data collection strategies:

- Retrieve all of the data provided by the machine in a raw format (over 200 parameters collected on a bus) and in real time (no tips applied),
- Retrieve 20 critical parameters in a raw format in real time (tips 2 and 3 applied),
- Retrieve 20 critical parameters in a processed format in real time (tips 2, 3, 6 and 7 applied),
- Retrieve 20 critical parameters in a processed format in a daily batch time (tips 2, 3, 6, 7 and 8 applied),
- Retrieve only certain alerts (5 alerts) (tips 2, 3, 4, 6, 7 and 8 applied),

In each case, we will calculate the quantity of data used each month (in Mb – Megabytes), an associated monthly cost (in euros) and estimate the functional coverage of the service (a ratio in %). Based on our experience, if 100% of the functions can be developed with 200 parameters, we will make the realistic hypothesis that

	Retrieve ALL the data	Retrieve 20 raw parameters	Retrieve 20 processed parameters in real-time	Retrieve 20 processed parameters in a daily batch	Retrieve only 5 critical alerts
Monthly Data Consumption (Mb)	2000	500	100	5	1
Cellular Communication Monthly Cost - International Coverage (€)	120	80	20	8	5
Functional coverage (%)	100	95	95	50	35

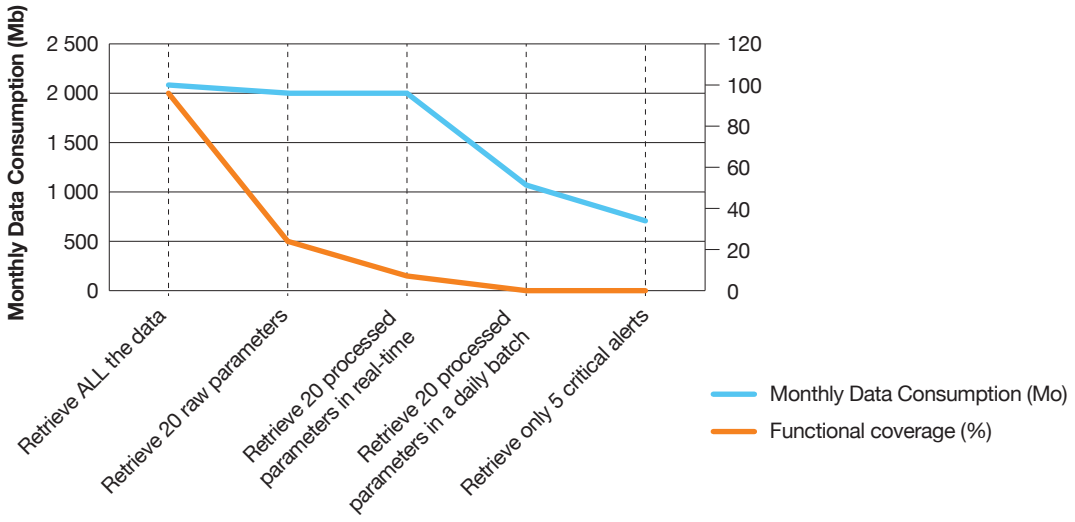

Diagram 1 - Data Consumption vs Functionnal Coverage

95% of functional coverage requires only 20 measurements. But the service has to ensure 24/7 reliability, so when a daily batch frequency is applied the functional coverage falls to 50%. And lastly, if only 5 critical alerts are retrieved, functional coverage is just 35%.

As shown in the graph above, the return on investment is the difference between the cost and the potential added value created by the solution, namely the functional coverage.

Even with only little data collected the value created is significant. Equivalent functional coverage can be obtained with scenario 3 while dividing the communication costs by 6 compared to the scenario 1. On another hand, scenario 4 divides both the communication costs and functional coverage by more than 2 compared to scenario 3. So is scenario 4 better than 3? In our use case, 24/7 reliability requirement means that we cannot provide only 50% of the service.

Retrieving 20 critical parameters in a processed format in real-time is the most cost effective strategy for this particular use case. This result shows how data transmission efficiency can have a major impact in striking the right balance between performance and profitability, thus helping to make the Internet of Things and Industry 4.0 more profitable.

Our 700 Internet of Things and analytics experts can help you in your IoT project worldwide with Datavenue, our innovative IoT and data analytics modular offer.
For more information, please contact us at contact.datavenue@orange.com

---