Discussion paper on confidential input data for VECTO

Author Iddo Riemersma / William Todts (Transport & Environment)

Version/date 1.1 (updated draft), 12-5-2015

Introduction

The VECTO simulation tool which is being developed aims to predict the fuel consumption/CO_2 of HDVs as accurately as possible in various boundary conditions such as mission profile, vehicle loading, configuration, etc. To achieve an acceptable level of accuracy of the model output, there are two important requirements:

1. The model architecture should match the complexity of all components that play a dominant role in the determination of actual fuel consumption/CO_2
2. Sufficiently detailed input data is required on those components that play a dominant role in the determination of actual fuel consumption/CO_2

Currently, the expert groups are working hard on the model architecture to arrive at the best modelling structure possible for VECTO, which covers the first requirement. The second point is just as important as the first, as it is often said amongst simulation modelers that '**the quality of the model is as good as the quality of the input data'.** On the delivery of quality input data there is a potential conflict between on the one hand the level of detail required to feed the model, and the confidentiality of these data on the other hand. Another issue is the extent to which the input data can be subjected to (public) scrutiny.

This paper is intended as a starting point on the discussion about the confidentiality of data in an effort to bring clarity to this issue and come to a common understanding on which data should or should not be seen as confidential. It will also touch on ways to make confidential information available for VECTO without it being disclosed to third parties.

Definitions

Several definitions can be found on what is seen as 'confidential information'. Within legal contracts it is defined as follows[1]:

> *Any trade secret, know-how, invention, software program, application, documentation, schematic, procedure, contract, information, knowledge, data, process, technique, design, drawing, program, formula or test data, work in progress, engineering, manufacturing, marketing, financial, sales, supplier, customer, employee, investor, or business information, whether in oral, written, graphic or electronic form;*

[1] http://www.contractstandards.com/clauses/confidential-information

All items listed in this definition -at least the ones having a relation towards technical components – have one characteristic in common: they are not just numbers or words, but carry some kind of intellectual content in them as well. In other words, there is knowledge contained within the information.

Another source provides a more compact definition, which relates to the potential risk of disclosure of confidential information[2]:

> *Containing information, the unauthorized disclosure of which poses a threat to national security.*

Interpreted to the situation for vehicle/component manufacturers, this definition might read:

> *Containing information, the unauthorized disclosure of which poses a threat to the manufacturer's competitive position.*

Combined with the conclusion on the first definition, we might conclude that the risk of the information being disclosed is the following: the knowledge contained within the information becomes available for competitors, who can then freely use this knowledge without having to invest in obtaining this themselves.

Discussion

An element which has not been touched upon in the definitions above is how the confidentiality of data is protected or guaranteed. If a parameter's value can easily be measured by others, it probably should not be seen as confidential. Another important question is whether there is knowledge included in that measurable parameter value in the first place. If not, such data cannot be labeled confidential as there is no way to secure them from exposure.

The same goes for a straightforward control system of which the input and output parameters can be measured, and from which the control algorithm is determined. This kind of 'reverse engineering' will cost a bit more effort but it is already common practice. Manufacturers buy vehicles from competitors and strip them completely to find out what can be learned from it in terms of design, production processes, calibration, etc. Only those systems that are too complex for reverse engineering, and which cannot be determined (e.g. by decoding from the ECU) can really be kept confidential/secure from the competitor or other third parties.

[Some of the essential input parameters may classify as genuinely confidential information, e.g. the underlying control algorithms of the engine's fuel injection map. The reason that the engine manufacturer wants to keep these confidential is to protect the enormous amount of delicate calibration work he has put into them. Other manufacturers might benefit from the know-how that was put into the fuel injection map, which brings a competitive disadvantage. In some cases the competitive disadvantage may disappear if data transparency is enforced by law. A good example of this is the road load of passenger cars. In Europe this information is treated as highly confidential information. Third parties cannot get access to the road load factors, even though it is a measurable set of three parameters without any intellectual content. In the USA however, the road load of

[2] http://www.thefreedictionary.com/confidential

passenger cars is annually verified by the EPA, and the measured values are made publicly available[3]. As every manufacturer has access to the information of its competitors, it is no longer experienced as a threat if the road load data on his vehicles also enter the public domain.

In the end, we will need to weigh the relevance of the input data for the VECTO results against the risk of disclosing sensitive information. As outlined in above, input data would not classify as being confidential if any of the following scenarios apply:

- If there is no knowledge included in them;
- If the data can measured or be easily reverse engineered;
- If there is no competitive risk to disclose the information, and/or
- If the competitive risk of disclosure disappears due to the transparency provided by all manufacturers.

To make the decision if input data classifies as 'confidential' more measurable, a table could be made which defines the border case for confidentiality on any of these aspects. An example table is shown below:

	Low	*High*
Data knowledge level	Just values or general know-how is included	Data contains detailed know-how resulting from design, engineering or calibration work
Data protection level	Data is easily measured or reverse engineered	Not possible to trace for third parties without excessive research or measurements
Data disclosure risk level	No competitive risk of disclosure or risk is diminished by full transparency	Clear competitive risk of disclosure

Table 1 – Data evaluation table to check the confidentiality

Input data that consistently scores in the 'High' category of this table would consequently classify as being confidential.

Since the relevance of the input data for VECTO also plays a role in this decision, it might be necessary to use specific tables for each of the different levels of input data relevance. It is therefore recommended that the process of confidentiality classification is started by making relevancy categories for the input data, and defining data evaluation tables for each of these categories.

Only in those cases where the input data is objectively judged as being confidential, but yet needed as an essential input, it should be discussed whether this input can be replaced by default values.

Data protection

In the previous section we have identified a method to objectively identify whether input data classifies as confidential or not. If some of the essential input data are considered to be confidential, the conclusion does not necessary have to be that these inputs cannot be used for VECTO. There

[3] http://www.epa.gov/otaq/crttst.htm

might be alternative ways to make use of these data without them while protecting them to be accessible to third parties, e.g.

- Confidential input data could be encrypted for use in VECTO, and the key for decryption would only be in the hands of manufacturers and/or appointed users (e.g. type approval bodies)
- The VECTO tool for specific vehicle models could be stored locally (e.g. at the server of the manufacturer or type approval body). VECTO users may access the tool by providing the necessary inputs and boundary conditions, upon which the output files are returned to the user.
- The use of the VECTO tool could be restricted to appointed parties (service providers) who will act as a bridge between manufacturers and end users. These parties would have to provide guarantees towards non-disclosure of the confidential VECTO inputs.

So far, these are just some first ideas on this. The feasibility of such options has to be further investigated and discussed within the VECTO editing board.