

**SUMMARY
REPORT**

**Cyber Security Modeling Language
(CySeMoL)**

1 CySeMoL (Cyber Security Modeling Language)

CySeMoL is the result from WP2 and is documented in deliverable D2.2. CySeMoL is used to estimate the attack resilience of the ICT-system architecture of SCADA and control systems.

Fundamental parts of CySeMoL

In summary, CySeMoL is a language (or Meta model) in which system architectures are described. The language contains general purpose entities such as services, data flows, operating systems, as well as security specific entities such as intrusion detection systems, firewalls and patch management processes. The language also defines how these concepts can be related to each other as well as some important properties (from a security perspective) of the entities, such as for instance if an operating system is using non-executable memory or if services have known vulnerabilities. With the language, users of CySeMoL are able to describe their system architectures. In principle they will draw figures looking like the architecture in Figure 1 in section 1.1. In addition to this purely descriptive part of CySeMoL, a mechanism for calculating value that roughly could be considered a security index (see next paragraph) is also included in the language. In essence, this mechanism is an attack/defense graph, which describes how different attacks and attack steps could be performed in the system architecture and its different components. So, depending on the exact configuration of the architecture, different attack processes will be possible for an attacker to accomplish. For all those attack processes, CySeMoL provides numerical estimates for how likely it is that all the different attack steps are possible to accomplish. These estimates are given as conditional probabilities (specified in Bayesian networks). As an example, the table below illustrates numerical figures of conditional probabilities of how likely it is to succeed with a semantic (rather than brute force) denial of service attack on a computer given the following two parameters of the system architecture, i.e. if the targeted host has a known vulnerability and if the attacker has access credentials to the targeted host.

Software Vulnerability Present	Attacker has Access Credentials	Expected Likelihood of Attack Success
Yes	Yes	0.72
Yes	No	0.53
No	Yes	0.60
No	No	0.38

(Vital Infrastructure, Networks, Information and Control System Management)

CySeMoL then combines a great number of such conditional probabilities of attack steps and associated countermeasures into an aggregated expected likelihood of success for the whole attack process. This aggregated value is the security index that we mentioned above. A more detailed discussion about these values is provided in the next section.

Data for the CySeMoL calculation mechanism

At the core of the CySeMoL lie the conditional probabilities used for the calculations. These figures have to a large extent been collected by asking security experts in surveys on their opinions of the impact of different countermeasures on different attacks, such as the DoS example above. For all questions, the main assumption to the respondent is that the attacker is a professional penetration tester with one week of preparation. Some of the figures are also deterministically derived, and some have been derived from previously published studies. In total, four surveys have been conducted on various parts of the CySeMoL with answers from 165 respondents as maximum and a handful of respondents as minimum. In order to identify qualified respondents (identifying which experts that really are experts) Cooke's classical method has been used. This method essentially weight different respondents depending on how good they are at answering some test questions relevant to the area of the survey questions (that the CySeMoL developers have known the answers to). This means that only a few of the for instance 165 respondents mentioned above performed good enough to be called experts. All answers, i.e. conditional probabilities, have been collected also including the respondents' opinion on the uncertainty of the answer (expressed as a three point estimation). For instance, for the first estimate in the table above, the average answer (of the respondents selected as the "true experts") is 72% as indicated in the Table. However, there is a 5% chance that the value is below 32% (again, on average) and a 95% chance that the value is below 95% (on average). Another way of expressing it would be that there is a 90% chance that the attack success value is between 32% and 95%. As we can see from this example the figure 72% here is acquainted with a large share of uncertainty. However, in the calculations we have made in the storyboards of this report only the expected mean value is used, i.e. 72% in the example above.

Intended usage of CySeMoL

The intended usage of the CySeMoL is to support security analyses of SCADA and control system architectures. It should support users that are not necessarily security experts themselves. If the user provides a system architecture, the CySeMoL can provide a security estimate in terms of attack probabilities. So, by analyzing different architectures and different attack processes the user can get a better understanding of available weak spots in the architecture. In addition, it also provides a clue on how effective different mitigation strategies (probably) are. As described above, the figures provided are often acquainted with quite big uncertainty. This imposes that the calculated percentage value figure should be treated with care. The results should be seen as a support for reasoning about different alternative scenarios or mitigations.

(Vital Infrastructure, Networks, Information
and Control System Management)

On average a scenario with attack success probability of 10% is more resilient towards the analyzed attack process than one with 30%, even included the uncertainties (that in general are the same magnitude for scenarios). Essentially the user needs to define two things: 1) the system architecture (including a number of properties), and 2) which targets they would like to analyze as well as starting points for the attacks. I.e., CySeMoL delivers results for (the most probable attack process between) pairs of a single starting point and a single target. But, in order to get a more complete and holistic understanding of the whole architecture many such pairs needs to be considered. Again, comparing scenarios without analyzing the complete set of potential pairs will provide an indication of their relative security.

Since the CySeMoL is quite large and complex, it is extremely time consuming to do the calculations by hand. In this report all examples used the Enterprise Architecture Analysis Tool to calculate the results and visualize the models. For further information about this tool see KTH Royal Institute of Technology, Enterprise Architecture Analysis Tool homepage, www.ics.kth.se/eat

Notes on the objectivity of the CySeMoL.

The correctness of results of the CySeMoL is to a large extent determined by the empirically collected conditional probabilities. A wide range of experts, both academic and practitioners, have been involved in the validation process. In the survey, Cooke's classical method, which is the state of the art method for expert knowledge elicitation, has been used. The method does not accept anyone as an "expert" but test all respondents with a test questions. Before, the surveys were sent out the parameters that were included in the survey was also validated with respect to their importance in a qualitative way by cyber security professionals. Both of these actions ensure the correctness in the details of the CySeMoL. Furthermore, the aggregated results (the attack path estimations) of the CySeMoL have also been validated through a Turing test. In the Turing test, both cyber security professionals and the CySeMoL were asked to assess a number of scenarios. All these assessments were then given to another group of cyber security professionals to judge if they believed that the results seemed reasonable (without knowing who produced the assessment). CySeMoL performed equally well in comparison to the cyber security professionals (it was neither best nor worst and everyone was performing in the same "magnitude").

Nevertheless, it is important to realize that the figures are "soft facts" rather than "hard" statistical data. This is due to the fact that no statistical data is available for the scope of CySeMoL and it is not feasible, from time perspective, to produce a sufficient new data. This means that we have ended up with a lot of uncertainty in the results. This is likely to reflect both that more detailed information about the system architecture is needed in order to make better predictions, but also that the answers to the questions that CySeMoL addresses are simply not known. (It could of course also be that the wrong people were asked, which we have tried to avoid by using a well-known scientific methods.). In many aspects CySeMoL needs to be enhanced and complemented in future works. By removing the following currently used assumptions CySeMoL could be improved:

(Vital Infrastructure, Networks, Information and Control System Management)

- The attacker has been simplified to one type of attacker with a lot of variance related to it. So how well these figures represent advance persistent threats or script kiddies needs to be further studied.
- The focus of CySeMoL is on addressing the availability and integrity objectives of the system rather than protecting the confidentiality (due to the SCADA and control system focus of the VIKING project). Thus, in the current version, countermeasures related to deterrence and recoveries are not yet considered.
- The focus of CySeMoL is mainly on technical parts of security. Thus, other type of cyber security attacks that involves non technical aspect, e.g. social engineering, information security governance and physical attacks are treated in a very rudimentary in this version.
- The conditional probabilities that produce the CySeMoL results needs to be updated from time to time as knowledge of both attackers and defenders change over time. CySeMoL reflects a snapshot of knowledge dated approximately to 2010.

Despite the above mentioned limitations, we believe that CySeMoL provides significant contribution in providing a systematic approach for analyzing the cyber security of SCADA and control systems. From an academic point of view this delivery should not be considered the end of CySeMoL but rather the beginning.