Automotive Vulnerability Grace Period Evaluation (AVGPE)

1st Robin Bolz
Institute of Energy Efficient Mobility
University of Applied Sciences
Karlsruhe, Germany
robin.bolz@h-ka.de

Abstract

With UN regulation R155 coming into force in 2021, the implementation of processes for incident response management has become relevant for type approval for all OEMs in many parts of the world. Accordingly, the implementation of processes capable of detecting cyber incidents and responding to vulnerabilities must be demonstrated. The regulation demands response within a 'reasonable time frame'. However, precise specifications as well as an evaluation basis for adequacy is not given. Due to the growing attack surface for vehicles, cyber incidents will increase and rapid prioritization and remediation of detected vulnerabilities will gain in importance. Time is of the essence when reacting to vulnerabilities. Efficient vulnerability handling and patching require cooperation along the supply chain with as little friction as possible. An industry-wide common understanding of the term 'reasonableness' is necessary to implement a corresponding internal resource allocation. From our point of view, there is a lack of sufficient experience and data in the automotive sector. This paper presents a metric to make assessment of the reasonable remediation time for vulnerabilities quantifiable. Thus, we introduce an objectifiable metric that can harmonize vulnerability management and make internal resource allocation more efficient. Specifically, we see potential use cases in the prioritization of vulnerability remediation actions, as well as in the assessment of the appropriate grace period in the context of coordinated vulnerability disclosure (CVD) processes.

Index Terms

Automotive vulnerability handling; Patch prioritization; Patching time evaluation, Incident Response Management

I. Introduction

Regulation UN R155 first introduced specific approval-related requirements for automotive OEMs in 2021, demanding for processes to ensure the detection of cyber threats and adequate responses to them. formulates the requirement that processes must be in place to enable a response within a reasonable timeframe. The regulation thus passes on to the manufacturers the responsibility for more precise specification of the concept of 'time adequacy', as well as concrete, valid assessment criteria for this. ISO/SAE 21434, also published in 2021, provides manufacturers with recommended actions in this regard. Our cross-industry research has shown that blanket, strict grace periods for disclosing information of vulnerabilities are often specified. These grace periods are deadlines for manufacturers to fix or mitigate vulnerabilities. If manufacturer fails to do so within the deadline (partial)

information is published by the coordinating entity which knows the vulnerability information. Within the IT industry Zero Day Initiative (ZDI) [4], CERT/CC [3], Rapid7 [1] and Google Security Team [5] or Project Zero [2], respectively, are the most dominant entities. Common deadlines are between 45 and 120 working days. In a worst-case scenario, details may then be published even before a mitigation or fix is available. This wide range shows the difficulty in determining a reasonable grace period. Plenty of research has been done, dealing with ideal grace periods for IT products and services (e.g. [6], [7]). To our knowledge, there is only one corresponding independent entity in the automotive sector that coordinates the disclosure of vulnerabilities in vehicles. The globally represented association Automotive Security Research Group (ASRG) [8] defines a 90-day default grace period. The deadline is not strict, so it can be extended under certain conditions. In view of the particular safety criticality of some driving functions, we consider this flexibility to be appropriate. However, we must criticize the fact that no valid reasons for an extension are specified. Setting strict deadlines for grace periods for disclosure of vulnerability information like common in IT is not recommended for the automotive sector for mentioned safety implications. In addition, remediation distribution can be of high complexity, when over-the-air functionality is not implemented. When it comes to vulnerability detection prevention of escalations such as 0-day disclosure or even exploitation is a key goal. Harmonizing the understanding of the appropriate remediation time or the associated grace period can prevent escalation and make company-internal incident response management more efficient. From our point of view the definition of common criteria which are accepted industry wide as valid for assessing reasonable remediation and grace period, respectively would help to achieve this harmonization. In this article, we first discuss existing and well-established criteria and metrics to measure vulnerability risk and severity. We separate qualitative and quantitative criteria and highlight their requirements and applicability. In the following, we introduce our proposal for valid criteria to assess the appropriateness of remediation times and argue our decisions. Based on these criteria, we then present our concept for a calculation metric that enables security experts to calculate the appropriate remediation time for a given vulnerability.

II. Related Work

This paper builds on and adds further detail and progress to previous publications on our work. In particular, an update of the set of formulas from [9] is discussed. In particular, the influence of the scope on the evaluation result is presented in more depth and the set of formulas is adapted accordingly. In the following, essential contents from our previous work, which are necessary for the understanding of this paper, are briefly reflected. The work [9] highlights existing state-of-the-art solutions and tools for company internal prioritization of security incidents and vulnerabilities in terms of applicability within the automotive industry. Thus, two main use cases can be identified for the metric proposed in the following to evaluate the appropriate patching and grace time for automotive. One is the quantification of the ’adequacy’ of OEMs’ response to vulnerability reports in terms

of UN155. The other is to improve the efficiency of internal company prioritization of the remediation of vehicle-specific vulnerabilities. To identify possible applicable criteria for evaluating the appropriate remediation and grace time, the work [9] examines existing approaches from the literature for assessing vulnerability severity, risk, and exploitability. Numerous works have been able to show that many of the solutions lead to inefficient vulnerability remediation because they focus too much on the severity rating of vulnerabilities according to the CVSS Base Score metric [10]- [15]. As a result, significant factors such as time-varying, and company- and product-specific influences are omitted from the assessment. Existing quantitative approaches that use regression models [16], [17], as well as data extraction models and machine learning algorithms [18]- [21] are also discussed. While these quantitative approaches are shown to be very useful within IT due to their objectifiability and automatability. However, the dependence of these approaches on appropriately large, public datasets of vulnerability information and context make their application to vehicle vulnerability assessment impossible for the foreseeable future, as such datasets do not exist. In [9], numerous existing heuristic metrics, such as the CVSS score, are cited. It is explained why our approach to evaluating reasoable patching times largely relies on evaluation criteria that are essentially derived from or extend the CVSS Score.

A. CVSS metric as a basis

Valid criteria are the basis for an accurate evaluation. In order to make the evaluation results usable throughout the industry, it must be possible to express them in numerical values. For this purpose, scores must be defined for all criteria according to their evaluation, which are then calculated within the framework of a superordinate set of formulas in the form of a numerical evaluation result. This is the state-of-the-art for industry-wide established metrics, such as the CVSS Base-, Temporal- and Environmental- Score. Adequate specification of both this set of formulas and the numerical values for the individual evaluation criteria is not trivial, as the overall result must correctly reflect the real situation. This complexity is also evident in the CVSS, which in its currently published version v3.1 has already undergone the fourth optimization of its calculation procedure during its 18-year existence. We believe that the CVSS is the most mature assessment metric for risk and severity of security vulnerabilities and also enjoys the greatest usage and acceptance across industries. For this reason, we base our metric on both criteria and scores, as well as key approaches from the CVSS v3.1 formulary, making adjustments where we believe it is necessary. Table 1 also shows the origin of our criteria.

III. Criteria for Appropriate Grace Period Evaluation

In this section, we briefly present the evaluation criteria we first specified in [9] and associated evaluation categories with numerical scores in the form of Table 1. Section 4.G. clarifies in more detail how we assigned these numerical scores.

Dimension	Kriterium	Bewertung aus...	Abb. Metrik	Ausprägung (Metrisch/Numerisch)				
Vulnerability Assessment	Vulnerability Risk Value (Impact Rating & Attack Feasibility)	ISO/SAE 21434 -	RV	Medium	High	Very High		
				0,7272	0,8636/0,95	1		
	Scope	CVSSv3.1 Base Score	S	Unchanged	Changed			
				1	Polynom			
Incident Analysis Auch GP Berechnung für Teilveröffentlichung bei vorübergehende Mitigation und späterem update denkbar	Exploit Code Maturity Quelle 4,5	CVSSv3.1 Temporal Score	EM	Not Def.	High	Functional	Proof-of-Concept	Unproven
				1	1	0,97	0,94	0,91
	Remediation Level Quelle 4,5	CVSSv3.1 Temporal Score	RL	Not Def.	Unavailable	Workaround	Temporary Fix	Official Fix
				1	1	0,97	0,96	0,95
	Report Confidence	CVSSv3.1 Temporal Score	RC	Not Def.	Confirmed	Reasonable	Unknown	
				1	1	0,96	0,92	
	Exploit Code Dissemination	none	ED	Unclear	External - Black	External - PoC	External - not public	
				1	1	0,96	0,91	
	Incident Scale/ Target Distribution (wie 'Scope' behandeln?!)	none/ CVSSv2 Environmental Scor	ISC	Unclear	High	Medium	Low	
				1	1	0,97	0,95	
Process Complexity	Supply chain scale	none	SCS	Low	Medium	High	Very High	
				0	0,037057	0,074113	0,11117	
	Remediation Dissemination (wie 'Scope' behandeln?!)	none	R	None R=	No access req.	Immediately R=	not determined	determined
				0	0	0	R=0,11117	Zeitwert Y
	Certification & approval	UN R156	APP	None	Extension	Renewal		
				0	5 Woche	Zeitwert		
	Contractual agreements (CA) - Interface agreement - CAL	ISO/SAE 21434 - 15. Distributed CS Activities	*CA*	None	CAL1	CAL2	CAL3	CAL4
				0	0	0	0,0693	0,11117
	Contractual agreements (CA) - CS Incident Response Plan	ISO/SAE 21434 - 13.3 CS Incident Response	*to be def.*	None	....	...	...	...
				to be def.				
	Contractual agreements (CA) - ...	...	*to be def.*	None	....	...	...	...
				to be def.				

Fig. 1. Criteria with associated evaluation categories and scores

The temporal urgency of remediation actions is strongly linked to the criticality of a vulnerability. Deriving criticality only from severity or risk (e.g., CVSS Base Score) ignores the potential dynamics of the threat situation. We therefore introduce the time criticality factor (TCF) to account for the pressure to act due to inherent vulnerability criticality. The remediation process is affected by extrinsic factors and circumstances that may delay the remediation time. We account for this through the process criticality factor (PCF). The evaluation of the appropriate grace period or remediation time must consider both urgency and process complexity.

A. The Process-Criticaliy Factor PCF

The PCF evaluates the (vulnerability-independent) time required by the responsible actors due to the technical prerequisites in the target system and the organizational circumstances. The greater the process criticality, the longer the appropriate grace period should be. The evaluation of the factor can change over time if the state of knowledge changes. The greater the process criticality, the longer the appropriate grace period. The factors that influence the PCF are presented below. These factors are assessed either by assigning a score (as in CVSS) and, in some cases, by assigning specific time when possible. While inadequate quality or quantity of human resources or insufficient processes have a significant impact on process duration, they are invalid factors and are not considered for this metric because they are the responsibility of the manufacturer.

1) Supply Chain Scale SCS: Multi-stakeholder processes for disclosing and remediating a vulnerability exhibit rising complexity as the number of vendors involved increases. Software from one supplier can be unique in each system implementation of the different customers, even if it maps the same function. A vulnerability can potentially affect multiple vendors in different ways. The assessment of risk

and thus vulnerability prioritization may vary from vendor to vendor. Increased intra- as well as inter-communication needs, and increased potential for conflict, can delay the process. In case of doubt, the process efficiency is as high as its weakest link. This metric does not consider whether pre-existing measures as part of contractual agreements exist. These factors can greatly facilitate and specify the estimation of process criticality, but are not always present. Therefore, they are recorded separately. The impact of the metric increases linearly with the number of stakeholders. The real given complexity is thus approximately modeled by assuming an only bilateral communication. Each intra- and interaction is assumed to have the same mean duration.

2) Remediation Dissemination R: This metric evaluates the reasonable grace period with respect to the impact of the responsible vendor's organizational-strategical patch or update management. Technical considerations that affect patch development or verification are specifically excluded from this metric. Such aspects, if they are part of agreements between manufacturers, can be included in the evaluation of the process criticality PCF with the Contractual Agreements (CA) metric. The assessment may change over time under certain circumstances. For example, if a temporary mitigation or hotfix is planned first and a more profound patch is planned for a later date. For clarity, the different definitions for a patch and an update are pointed out. An update is defined as a measure that is characterized by the intention to implement extended or optimized functionality. It is usually deployed on a cyclical basis. A patch, on the other hand, is defined as a measure characterized by being triggered by a specific incident. Deployment can be event-triggered or cyclic (synchronized with the update cycle), depending on given circumstances. Manufacturers usually pursue an efficient update management. Especially when no over-the-air (OTA) functionality is available, the bundled release of patches (e.g., Microsoft's "Patchday") makes sense for many reasons. Key factors influencing patch and/or update management can be economic considerations, system availability and safety and liability risks (monetary or reputational). Keeping the system up-to-date with recently released patches results in higher operational costs, while patching the system infrequently for its vulnerabilities leads to higher damage costs associated with higher levels of exploitation. A basic distinction can be made between a time-driven approach and an event-driven approach. In addition to the time at which a manufacturer provides patches, patch implementation management of the vehicle owner also plays a key role. The behavior of the vehicle owner lies outside the manufacturer's responsibility and has no influence on the evaluation of the appropriate grace period. If no concrete time specification can be made, the evaluation is carried out by a dimensionless score that is included in the PCF. If a concrete time specification can be made, the evaluation is done by a concrete time specification in days.

3) Certification & Approval APP: This metric evaluates the reasonable grace period with respect to delays which may arise with necessary approval obligations. Regulations for type approval vary worldwide (e.g. UN 156 regulation). Depending on country-specific regulations, vehicle manufacturers may be required to provide

processes as part of their internal software update management that can be used to verify whether and how a software update will modify (alter, remove, add, enable, disable) any parameters or functions of type approved systems to be updated/patched or parameters used to type approve those systems. Furthermore, this also applies to modifications of parameters or function that are defined within legislation or that will affect any other system required for the safe and continued operation of the vehicle. The responsibility for the correct assessment of the relevance of an update/patch for approval lies with the vehicle manufacturer. Whether an extension or renewal of the approval is necessary must and can therefore be assessed by the manufacturer. The extent of the delay in patch dissemination due to approval processes should be specified on the basis of empirical values and in consultation with the approval authority or technical service.

4) Contractual Agreements CA: This metric can be used to include aspects into the evaluation of the reasonable grace period that may have been defined as part of agreements between the customer and supplier and can help to refine the assessment for the given specific vulnerability. These may be agreements between partners within a company or beyond. In particular, this metric reflects the RQ-07-04 and RQ-07-05 requirements of ISO/SAE 21434 and provides the ability to make the evaluation more adaptable to given real circumstances. Since such agreements generally lack an objective judgment, this metric may be unsuitable for the 'Coordinated Vulnerability Disclosure' use case. However, it may be suitable for use in the internal patch or vulnerability prioritization process. Example 1: Patch development and verification Technical requirements of patch development and verification could serve as criteria. If qualitative or quantitative agreements on the organizational and technical management of patch development and verification are defined between costumer and supplier, these could be used to adapt the assessment. (e.g. required security testing activities) Example 2: Incident Response Plan Agreements defined in the context of an Incident Response Plan (criteria for closure of information, remedial actions). Example 3: Cybersecurity Assurance Level - CAL Given a CAL which is assigned to the vulnerable or affected component/item (resp. its security goals), special requirements has to be fulfilled by a supplier when it comes to the need to handle the vulnerability (e.g., patching, communication, providing information). The possible values to rate this metric could reflect the gap between the pre-defined assurance levels (e.g., 1-5) and the assigned 'Risk Value' or the assessment could be based on a pre-defined CAL classification scheme to determine the level of rigour of necessary cybersecurity measures to provide the required assurance. It may also be a combination of several such aspects.

B. The Time-Criticality-Factor TCF

The TCF evaluates the action or time pressure based on the criticality given by the vulnerability. The evaluation of the factor can change over time if the state of knowledge changes. The lower the time criticality, the longer the appropriate grace period should be. The factors that influence TCF are presented below. These factors are evaluated by assigning a score (as in CVSS).

1) Vulnerability Risk Value RV: This metric is based on the 'Risk Value' in accordance with ISO/SAE 21434. Accordingly, the value assigned to the 'Risk Value' is determined from the risk assessment defined in the company (e.g.: 'Risk Matrix'). This risk assessment considers the 'impact' and 'feasibility' assessments in a certain weighting. Impact and feasibility assessments are also defined in ISO/SAE 21434. If, in the context of a TARA, 'Risk Values' have already been defined in the product development phase for the 'Threat Scenario' that is now to be evaluated in real terms on the basis of the vulnerability, these values can be adopted here (the same applies to 'impact' or 'feasibility' evaluation). According to ISO/SAE 21434, a 'Risk Value' is determined for each SFOP impact category (Safety, financial, operational, privacy) of a threat scenario. The relevant impact categories for this metric are safety and operational. If multiple threat scenarios (with associated attack paths) can be derived from the vulnerability to be assessed, then the threat scenario that has the highest risk value for the safety impact category is relevant. If this value is assigned to multiple threat scenarios, then the threat scenario that also has the highest risk value for the operational impact category is relevant. The 'Risk Value' reflects the company-specific ('environmental') consideration of the impact. It should therefore not be equated with the CVSS Severity Base Score that is usually found in CVE and/or NVD entries.
2) Scope S: This metric is adapted from the Common Vulnerability Scoring System v.3.1 (CVSS v3.1) The Scope metric captures whether a vulnerability in one vulnerable component impacts resources in components beyond its security scope. If a vulnerability in a vulnerable component can affect a component which is in a different security scope than the vulnerable component, a scope change occurs. Intuitively, whenever the impact of a vulnerability breaches a security/trust boundary and impacts components outside the security scope in which vulnerable component resides, a scope change occurs. Formally, a security authority is a mechanism (e.g., ECU, HSM, CGW) that defines and enforces access control in terms of how certain subjects/actors (e.g., drivers, processes) can access certain restricted objects/resources (e.g., data, car interior, functions) in a controlled manner. All the subjects and objects under the jurisdiction of a single security authority are considered to be under one security scope. The security scope of a component encompasses other components that provide functionality solely to that component, even if these other components have their own security authority. TCF is greater when a scope change occurs.
3) Exploit Code Maturity EM: This metric is adapted from the CVSS v3.1 It measures the likelihood of the vulnerability being attacked, and is typically based on the current state of exploit techniques, exploit code availability, or active, "in-the-wild" exploitation. Public availability of easy-to-use exploit code increases the number of potential attackers by including those who are unskilled, thereby increasing the severity of the vulnerability. Initially, real-world exploitation may only be theoretical. Publication of proof-of-concept code may follow. Further dependencies arise with the assessment of Attack Feasibility in the context of determining the Vulnerability Risk Value according to ISO/SAE 21434. The threat posed by the

vulnerability to be assessed is considered over time. The 'risk value' that is taken from TARA documentation does not fully consider the maturity of an actual real-world exploit at the time of the initial assessment of a real vulnerability. Even if 0-days have to be evaluated differently compared to already publicly known vulnerabilities, it is true for both that the evaluation may differ from the one at the time of TARA. The more easily a vulnerability can be exploited, the higher the score.

4) Remediation Level RL: The Remediation Level of a vulnerability is adapted from the CVSS v3.1. It assesses the reduction of the threat posed by the vulnerability to be assessed over time and is an important factor for vulnerability prioritization. Unlike in IT, the typical vulnerability in automotive is usually already patched before public awareness. Nevertheless, the remediation level influences the time criticality for measures like workarounds, hotfixes or final patches, since it reflects a decreasing urgency as remediation becomes final. When a possible attack scenario is reported by a white hat, this report can contain several related vulnerabilities (attack vector). Since each vulnerability is assessed individually, assessment and prioritization can differ. Mitigation or remediation can simultaneously lead to mitigation or remediation of related vulnerability. This metric captures such effects that lead to changes in exploitability over time. It must therefore be applied to the entirety of all threat scenarios (or attack paths). If multiple threat scenarios (with associated attack paths) can be derived from the vulnerability, the effect of a measure on only the threat scenario that has the highest 'Risk Value' is relevant. Therefore, if an existing measure has no effect the threat scenario that has the highest 'Risk Value', then this measure is considered as 'Unavailable'.

5) Report Confidence RC: This metric is adapted from the CVSS v3.11. It measures the degree of confidence in the existence of the vulnerability and the credibility of the known technical details. Sometimes only the existence of vulnerabilities is published, but without specific details. For example, an impact may be identified as undesirable, but the root cause may not be known. The vulnerability may later be confirmed by research that suggests where the vulnerability may lie, although the research may not be certain. Finally, a vulnerability may be confirmed by confirmation from the author or vendor of the affected technology. Thus, the urgency of a vulnerability may vary over time as new discoveries lead to the assumption of a vulnerability's existence with altered certainty. This metric also indicates the level of technical knowledge and equipment available to potential attackers. Thus, overlaps may exist with the 'Risk Value' according to ISO/SAE 21434 (Attack feasibility rating). The more a vulnerability is validated by reputable sources, the higher the score.

6) Exploit Code Dissemination ED: This metric measures the availability of exploit code. The maturity of the exploit does not need to be evaluated (since this aspect is already evaluated by the metric 'Exploit Code Maturity'). Scoring is based on statistical significance from studies of historical data on vulnerabilities in the traditional IT environment [22]- [23] The limits of the validity of quantitative evaluation methods for automotive were given in this work and are driven by

the following evidence-based assumptions: i. Considering the existence of exploit code for a given publicly known vulnerability as a risk factor for actual exploitation in the wild can increase prediction rate up to 45% better than only considering the CVSS Score. [23] ii. Considering the existence of exploit code for a given vulnerability on black market as a risk factor for actual exploitation in the wild can increase prediction rate up to 80% better than only considering the CVSS Score. [23] The lower the prediction rate given by a risk factor, the lower the statistical exploitation probability on average and TCF decreases. If there are no findings or if the prediction rate is as high as possible (ED=1 for external - black market), this metric has no effect on TCF.

7) Incident Scale ISC: The evaluation of the scalability of damage is not fundamentally considered in the 'Risk Value' according to ISO/SAE 21434. This metric therefore does not represent an overlap. It factors the amount of impacted/vulnerable components deployed in the field. Since the assessment can change over time, it should be reviewed and, if necessary, adjusted, if the scalability of damage has already been considered in the development phase for a specific manufacturer (e.g., expected and actual deployment deviates significantly). The metric does not necessarily reflect only the company-specific ('Environmental') view. The environment in which components could be exposed to a threat due to the given vulnerability must initially be defined. The environment under consideration is not necessarily only the environment of one or more companies, but potentially the entire world due to the highly transient character and the connectivity of vehicles. If a specification of the exact number of deployed impacted/vulnerable components cannot be made with certainty, the maximum number to be assumed should be used. This can be the case, for example, if precise assembly documentation (e.g. via vehicle identification number (VIN)) is available, but an exact assessment is not possible due to incomplete tracking of the software status of the impacted/vulnerable component. In this case, all components would have to be treated as if they had the impacted/vulnerable software version. It should be noted that this metric does not evaluate possible significant scaling effects resulting from the vulnerability potentially enabling, for example, physical or functional bypass of an attacker to adjacent components, items, domains, or functions along an attack path. Such effects are accounted for using the 'scope' metric. For example, this may refer to a vulnerability within a telematic control unit that is located right at the interface to the backend server and could potentially allow the attacker to scale the attack to the backend.

IV. Appropriate Grace Period Evaluation Framework

The criteria specified in Chapter 3 are used to mathematically determine the appropriate grace period. The mathematical framework for this is explained in this chapter.

A. The Reasonable Grace Period T_{GP}

A look at the calculation equation makes the basic idea clear. It can be seen that the calculation of T_{GP} is done by scaling an initial time value T_{GP} with the factor GP. We calculate the reasonable grace period T_{GP} as follows:

$$T_{GP} = T_0 \cdot GP + APP + [r \cdot R] \tag{1}$$

The differences between dimensionless values (factors) and time values (summands) are discussed as well as the other individual terms of this equation are further explained in the following.

B. The Default Grace Period T_0

The default grace period T_0 is the (vulnerability-independent) minimum time that the manufacturer must be granted to make a measure available for the vulnerable or affected component. For a scenario with maximum time criticality and minimum process criticality, the appropriate grace period would be the default grace period T_0. Industry representatives and 3rd party security service providers in the IT sector usually anchor a valid grace period in their own disclosure policy, as a strict, non-adjustable deadline applies. Special criticality is also not considered. This not infrequently leads to threats from 0-day disclosure. As of today, no such strict requirements are known in the automotive sector. Due to the special criticality in cyber-physical systems, we do not think this is appropriate either. Instead, we advocate anchoring a default grace period as a non-strict deadline. If there is no particular criticality, this deadline is valid. If a stakeholder has doubts about the appropriateness of this default deadline, he should be able to indicate this. An adjustment can then be made comprehensibly and transparently on the basis of the TCF and PCF factors mentioned. The evaluation is uncomplicated with our tool.

C. The scaling factor GP

The scaling factor quantifies the overall criticality. Given the minimum possible default grace period T_0, GP scales the grace period to the actual appropriate length. Scaling is based on the calculated values for TCF and PCF

$$T_1 = T_0 \cdot (1 + PCF) \tag{2}$$

$$T_2 = T_(1) \cdot (1 - TCF) \tag{3}$$

$$GP = (1 + PCF) \cdot (2 - TCF) \tag{4}$$

For a scenario with maximum time criticality TCF and minimum process criticality PCF, GP would be 1. Thus, there would be no scaling of the given default grace period.

D. Calculating TCF

Like described in Section 3 the TCF is influenced by the factors Vulnerability Risk Value RV, Exploit Code Maturity EM, Remediation Level RL, Report Confidence RC, Exploit Code Dissemination ED and Incident Scale ISC. These factors are multiplied. Given the scores of the individual factors, the possible range of values for TCF is as follows:

$$TCF = RV \cdot EM \cdot RL \cdot RC \cdot ED \cdot ISC \tag{5}$$

$$0,5 \leq TCF \leq 1$$

The factor Scope S plays a special role for the value of TCF, since it essencialy influences the evaluation of risk.

1) The influence of Scope on TCF: To appropriately reflect the extent of the influence of Scope in our metric, we again align with the CVSS Base Score. The extent to which the Scope score has on the appropriate grace period is intended to be analogous to the extent on the CVSS Base Score. In the following, we therefore analyze the calculation procedure of the CVSS v3.1 and how it is influenced by the Scope criterion. We then derive insights for our own metric from this.
The CVSS v3.1 Base Score is calculated according to the basic formula:
If Impact $\leq$ 0:

$$BaseScore = 0 \tag{6}$$

If Scope is 'Unchanged'

$$BaseScore = Impact + Exploitabiity \tag{7}$$

If Scope is 'Changed'

$$BaseScore = 1.08 * (Impact + Exploitabiity) \tag{8}$$

The calculation forms for the CVSS criteria Impact and Exploitability have been mathematically optimized over the years based on experience gained in applying the metric. By calculating all possible 2592 (1296 for scope change and 1296 for scope unchanged) Base Score results for the current version v3.1, sorting them in ascending order and plotting them, the metric can be analyzed. Figure 2 shows the result of this process. There are two correlating graphs, since a different formula is used for Scope Changed than for Scope Unchanged. According to the formulas (2) and (3), one would expect the difference of both graphs with a factor of 1.08 to apply to all 1296 values in each case. However, the calculation of the criterion Impact also changes depending on the Scope. We also want to reflect the special role that scope plays in the calculation of the CVSS base score in our metric. For this purpose, we quantify the percentage deviation of the graphs of the Base Score for Scope Changed or Scope Unchanged for each value in the sorted

progression, smooth it, and graph it using 4th degree polynomial p4(x) (Figure 3). The definition range of p4(x) results from the number of all possible 1296 values that the Base Score can assume for Scope change or Scope unchanged, respectively. The smoothing is done using 1st order exponential smoothing with smoothing parameter g=0.975. The polynomial represents the deviation of the two CVSS base score curves. This change in risk assessment due to Scope Change can now be applied to our metric. The Base Score of the CVSS metric combines the Exploitability and the Impact and thus expresses the severity or risk of a vulnerability. The risk in our metric is determined by the Risk Value and is reflected in the TCF value. If the Scope is unchanged (S=0), this leads to a minimization of TCF. The minimization depends on TCF itself and is analogous to the change by Scope change (S=1) of the Base Score. TCF can assume 4800 values, instead of 1296 like Base Score. So, The x-axis of the graph shown in Figure 3 has to be scaled by factor 3,7037037 to get the function deltaTCF(z) which is representing the impact of Scope change on TCF. deltaTCF can be represented as a 4th degree polynomial with parameters a=-0,0000000000006, b=0,0000000062642, c=-0,000021993568, d=0,0315280325515 and e=-0,2484586037732, with a coefficient of determination $R^2 = 0,9049$ (see Figure 4). In analogy to the CVSS Base Score, the Scope score influences the TCF in the following way:

$S = 1 : TCF = RV \cdot EM \cdot RL \cdot RC \cdot ED \cdot ISC$

$S = 0 : TCF = RV \cdot EM \cdot RL \cdot RC \cdot ED \cdot ISC \cdot (1 - deltaTCF)$

with $deltaTCF(z) \approx deltaBaseScore(x) = p4(x), x \in \mathbb{Z}, x = [1; 1296]$ and $z \in \mathbb{Z}, z = [1; 4800]$

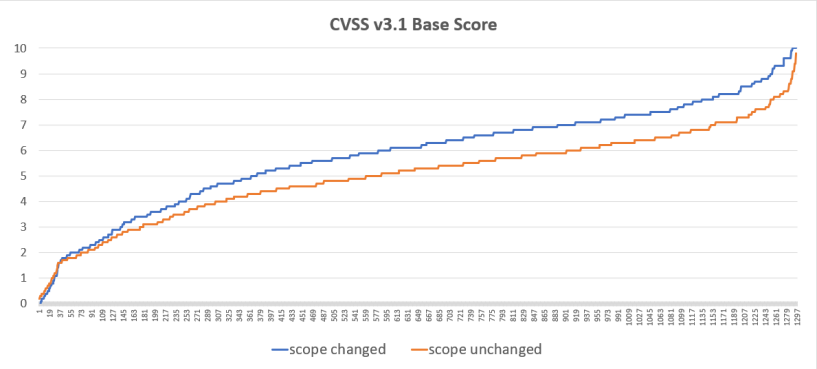

Fig. 2. The whole possible value range of the Base Score for Scope Changed and Scope Unchanged, respectively

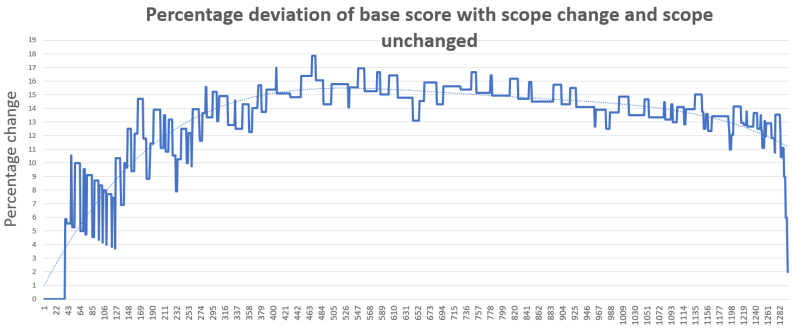

Fig. 3. Percentage deviation of Base Score with Scope Changed and Scope Unchanged

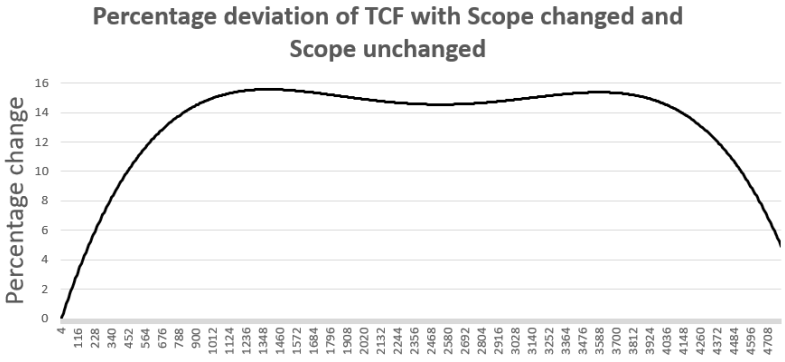

Fig. 4. Percentage deviation of TCF with Scope Changed and Scope Unchanged

E. Calculating PCF

Like described in Section III the PCF is influenced by the factors Supply Chain Scale SCS, Remediation Dissemination R and Certification & Approval APP. These factors are multiplied. Given the scores of the individual factors (see Figure 3), the possible range of values for PCF is as follows:

$$PCF = SCS + [r_2 \cdot R] \tag{9}$$
$$0 \leq PCF \leq 0,3335$$

R can assume both a dimensionless and a time value. If a time value is assigned (r_1=1, r_2=0), this value is not included in PCF, but is included in equation (1) and the appropriate grace period T_{GP} is extended by this time value in days. If, on the

other hand, a dimensionless value is assigned (r_1=0, r_2=1), this value is included in the calculation as a factor. PCF is scaled by the assigned value. The CA (see Section III) value is intended to describe the impact on process criticality resulting from any agreements between manufacturers. Whether such agreements can and should be considered, or whether they exist at all, cannot be assessed in a general framework. Accordingly, CA serves to individualize our metric by potential users which applies our tool. Since both the documentation and the source code for the calculation tool are open source, the score can be changed at any time to suit the needs of any entity that applies it for further consideration, we assume the value 0 for CA. We thus ignore its influence on PCF. For more intuitive handling, the numerical values can be assigned to categories of Low, Medium, High. Figure 3 illustrates the categorization.

F. Overview

The whole methodology is illustrated in Figure 5. It illustrates the cascaded scaling of T_0 with PCF and TCF, as well as the time values.

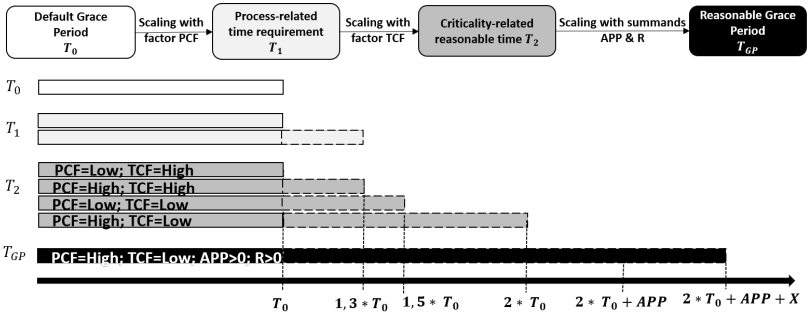

Fig. 5. Cascaded computational logic, which shows the scaling impact of the individual terms of the equation, T_0, PCF, TCF, R and APP on the result T_{GP}

G. Determination of the scores

Scores for criteria were determined based on two constraints.

1) For the scaling factor GP the following must apply: $1 \leq GP \leq 2$
   For maximum time criticality (TCF=max.) and minimum process criticality (PCF=min.) GP=1 shall apply and thus the standard grace period shall not be changed by GP. For minimum time criticality (TCF=min.) and maximum process criticality (PCF=max.) GP=2 shall apply and thus the standard waiting time shall be extended by GP by a factor of 2.
2) For all criteria adapted from CVSS v3.1, the scores are adopted as defined there.

V. Conclusion

The requirement in UN R155 to respond to vulnerabilities in a reasonable timeframe raises questions regarding the notion of reasonableness. In this work, we address this question by defining valid criteria to assess adequacy in this context. Sufficient empirical values and empirical data on vulnerability lifecycles can be very useful here, but are scarce in the automotive industry. This makes the application of quantitative methods, such as those increasingly used in IT, largely impossible. This was contrasted with the problems of qualitative evaluation procedures, which generally lack objectivity. Using essentially qualitative application procedures, we then established criteria with which an evaluation of the appropriate remediation time resp. grace period can be made. We present a scoring scheme for the defined criteria as well as a calculation formula. Our evaluation metric can be used within CVD processes to transparently define a reasonable grace period and thus harmonize the process flow. The grace period factor GP can serve as an industry-wide indicator and, in our view, should be part of any published vulnerability advisory. In an enterprise context, our metric can help to clarify the prioritization of vulnerability remediation and thus improve resource allocation within the enterprise. For this purpose, our metric as well as our tool is publicly available and highly customizable.

Author Contributions

This research work has been funded by the German Federal Ministry of Education and Research (BMBF) in the context of the project SecForCARs.

References

[1] The Rapid7 Disclosure Policy. Available online-11/14/2022: https://www.rapid7.com/security/disclosure/
[2] Project Zero, Policy and Disclosure: 2020 Edition, Jan. 2020. Available online-11/14/2022: https://googleprojectzero.blogspot.com/2020/01/policy-and-disclosure-2020-edition.html
[3] The CERT/CC Disclosure Policy. Available online-11/14/2022: https://vuls.cert.org/confluence/display/Wiki/Vulnerability+Disclosure+Policy.
[4] The Zero day Initiative Disclosure Policy. Available online-11/14/2022: https://www.zerodayinitiative.com/advisories/disclosure_policy/
[5] Google Security Team, Rebooting Responsible Disclosure: A focus on protecting end users, 2010, https://security.googleblog.com/2010/07/rebooting-responsible-disclosure-focus.html, online available 05/28/2020
[6] McQueen, M., Wright, J., Wellman, L., Are Vulnerability Disclosure Deadlines Justified?, Proeedings of the Third International Workshop on Security Measurements and Metrics, Third International Workshop on Security Measurements and Metrics, IEEE Xplore, doi: 10.1109/Metrisec.2011.9, 2011.
[7] Arora, A., Telang, R., Xu, H., Optimal Policy for Software Vulnerability Disclosure, Management Science Vol.54, No.4, p. iv-862, doi:10.1287/mnsc.1070.0771 2008.
[8] The Automotive Security Research Group Disclosure Policy. Available online-11/14/2022: https://asrg.io/disclosure/

[9] Bolz, R. Evaluating reasonable patching times for security product vulnerabilities in the automotive field, in Reports on Energy Efficient Mobility - Volume 3, Zenodo, 2023. Available online-11/14/2022: https://zenodo.org/record/7573669
[10] C. Fruhwirt und T. Mannisto, Improving CVSS-based vulnerability prioritization and response with context information, 3rd International Symposium on Empirical Software Engineering and Measurement, S. 535-544, 2009.
[11] Debabrata Dey, Atanu Lahiri, and Guoying Zhang. 2015. Optimal policies for security patch management. INFORMS Journal on Computing 27, 3 (2015), 462–477. Available online-11/14/2022: https://doi.org/10.1287/ijoc.2014.0638
[12] W Shuguang, X Chunhe, G Jinghua und J Qiong, "Vulnerability evaluation based on CVSS and environmental information statistics", 2015 4th International Conference on Computer Science and Network Technology (ICCSNT), S. 1249-1252, 2015.
[13] L. Gallon, "On the Impact of Environmental Metrics on CVSS Scores", 2010 IEEE Second International Conference on Social Computing, S. 987-992, 2010.
[14] Bozorgi M, Saul LK, Savage S, Voelker GM. Beyond heuristics: Learning to classify vulnerabilities and predict exploits. Pp. 105– 114 in Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining, 2010.
[15] N. Munaiah, A. Meneely. 2016. Vulnerability severity scoring and bounties: why the disconnect? In Proceedings of the 2nd International Workshop on Software Analytics (SWAN 2016). Association for Computing Machinery, New York, NY, USA, 8–14. https://doi.org/10.1145/2989238.2989239
[16] Frei, S. et al., Large-scale vulnerability analysis Available online-11/14/2022: https://dl.acm.org/doi/10.1145/1162666.1162671
[17] C. Fruhwirth and T. Mannisto, "Improving CVSS-based vulnerability prioritization and response with context information," 2009 3rd International Symposium on Empirical Software Engineering and Measurement, 2009, pp. 535-544. Available online-11/14/2022: https://ieeexplore.ieee.org/document/5314230
[18] Ansari, A., Ameen Alimohideen, M., & P.C., K. (2021). Deep Learning Based Real Time Vulnerability Fixes Verification Mechanism for Automotive Firmware/Software. Available online-11/14/2022: https://api.semanticscholar.org/CorpusID:233604439
[19] Jay Jacobs, Sasha Romanosky, Idris Adjerid, and Wade Baker. 2019. Improving vulnerability remediation through better exploit prediction. In Proceedings of the Workshop on the Economics of Information Security
[20] Jay Jacobs, Sasha Romanosky, Benjamin Edwards, Idris Adjerid, and Michael Roytman. 2021. Exploit Prediction Scoring System (EPSS). Digital Threats 2, 3, Article 20 (September 2021), 17 pages. Available online-11/14/2022: https://doi.org/10.1145/3436242
[21] Chen, H., Jing Liu, J., Liu, R., Park, N., Subrahmanian, V. S., VEST: A System for Vulnerability Exploit Scoring & Timing, Proceedings of the Twenty-Eighth International Joint Conference on Artificial Intelligence Demos. Pages 6503-6505. Available online-11/14/2022: https://doi.org/10.24963/ijcai.2019/937
[22] Luca Allodi. 2017. Economic Factors of Vulnerability Trade and Exploitation. In Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security (CCS '17). Association for Computing Machinery, New York, NY, USA, 1483–1499. https://doi.org/10.1145/3133956.3133960
[23] Luca Allodi and Fabio Massacci. 2014. Comparing Vulnerability Severity and Exploits Using Case-Control Studies. ACM Trans. Inf. Syst. Secur. 17, 1, Article 1 (August 2014), 20 pages. https://doi.org/10.1145/2630069