NB Parts of this (long) proposal have intentionally been removed to make it more suitable for this class (e.g., consortium, risk assessment, management structure, justification of the project budget, and data management have all been removed).

Dutch Research Council
Science

Section 2 - Research proposal

2.1 Scientific summary

A core assumption underlying organizational security practices is that defenders are able to remediate known vulnerabilities in their systems in a timely fashion. Otherwise, attackers can just follow the breadcrumbs laid out by security advisories and exploit known weaknesses. This is indeed what happens in many large breaches. While progress has been made at the level of consumers, with automatic updates and default patching settings, this does not translate to enterprises. They face a painful dilemma: patch too soon and incur potential downtime and failures; patch too late and get compromised by attacks. As a result, organizations take a long time to patch even critical security vulnerabilities.

The central objective of THESEUS is to empower organizations to patch much faster. It aims to achieve this by **radically changing the risk governance of patching**. Changing the risk of patching for enterprises means to develop interdisciplinary breakthroughs at three interdependent levels:

- Systems: reducing risk of patching via new techniques in automatic vulnerability and patch triaging, as well as automatic patch generation with live update for cases where critical patches pose unacceptable availability risks.
- Enterprises: better quantifying risk of patching by assessing and aggregating the results of the patch triaging, as a way to estimate exploit likelihood in a coherent picture that accounts for different attacker models and functional impact.
- Governance: more effectively managing risks of patching by introducing incentive mechanisms via notifications and information sharing, sector-wide benchmarks of patching speed, and potentially legal instruments.

THESEUS sets out to (1) bring advances from the lab to real-world settings by working with a large consortium of partners from healthcare and transportation who contribute people, data, and pilots; and (2) replace the status quo, as well as counterproductive solutions like mandatory patching, with a richer set of governance interventions across different levels.

2.2 Focus of the project

✓	*a. Governance and theme line*
	b. Governance, theme line and cryptology

2.3 Project description

An unmitigated disaster

Contrary to common belief, more than 99% of attacks do not use highly advanced exploitation techniques or so-called zero-days, but rather known vulnerabilities that have often already been fixed in security patches for a long time.[1] The Rathenau institute recently advised the Dutch government that new technologies for security, such as artificial intelligence and post-quantum cryptography, will not matter if a basic issue like security patching remains unsolved.[2] For this reason, the current practice of delayed patching is arguably the most important problem in security today.

> Jim Hagemann Snabe (Maersk chair), at a World Economic Forum event after the NotPetya incident: "We overcame the problem with human resilience". Article author Andy Greenberg notes that before the incident, a "security revamp was green-lit and budgeted [... but was] never made a so-called key performance indicator" and so was not implemented.
> https://www.wired.com/story/notpetya-cyberattack-ukraine-russia-code-crashed-the-world/

The consequences of delayed patching can be disastrous. During the infamous wave of Wannacry and NotPetya ransomware in 2017, all major incidents (e.g., disruptions of Maersk, 16 UK hospitals shutting down, disruption of Renault factories) occurred several months after Microsoft had released a patch that would have prevented the outbreaks. In 2017, Equifax suffered a major data breach that leaked personal data of over 140 million customers. It became one of the biggest data breaches in history. The attackers exploited a known vulnerability in the Apache Struts servers that could and should have been patched in March 2017 when the patch was released and administrators were told to apply it through all affected systems, but the process failed and the

[1] Moore, A. (2017) Focus on the Biggest Security Threats, Not the Most Publicized. At: https://www.gartner.com/smarterwithgartner/focus-on-the-biggest-security-threats-not-the-most-publicized.
[2] Van Booheemen, P., Munnichs, G., Kool, L., and Hamer, J. (2020) Cyberweerbaar met nieuwe technologie: Kans en noodzaak van digitale innovatie. At: https://www.rathenau.nl/nl/digitale-samenleving/cyberweerbaar-met-nieuwe-technologie

vulnerability remained open in multiple systems until the end of July 2017[3]. Even high-profile security vulnerabilities such as Heartbleed remain unpatched in many organizations for months after disclosure (Durumeric et al. 2014). More recently, months after the vulnerabilities of two enterprise solutions (Pulse Connect Secure VPN and Citrix ADC and Gateway) were made public and mitigations became available, Internet-wide scans found thousands of organizations that were still running vulnerable systems.

In essence, the lack of patching hygiene is a governance problem. As the impact of breaches often radiates from the individual organization towards supply chains, customers, and third parties, the slow response to known vulnerabilities is also a major issue for society as a whole and thus the government. The Dutch Safety Board announced that they will investigate the Citrix incident to improve the "governance of digital security".[4] The Dutch minister of Justice & Security announced he was considering top-down interventions for critical industries: mandating organizations by law to implement the security advisories of the government.

A mandatory patching regime could, however, well be worse than the disease. In the last three years, 48,373 vulnerabilities were reported to the NVD database, of which 11,628 were considered to have high severity. To mandate patching of all of these vulnerabilities would potentially expose organizations to thousands of planned and unplanned interruptions, even though most of these vulnerabilities are never exploited in the wild (Allodi & Massacci, 2014). That being said, the fact that the government is announcing such drastic measures underlines the fact that we are dealing with a governance challenge. The societal impact of long vulnerability windows is too severe to leave the problem to the trial and error of individual organizations.

Minister Grapperhaus (Justice and Security) after the Citrix incident: "We have to be able to say to a company: 'if you don't take care of it, we will come and do it for you'."
https://fd.nl/ondernemen/1318504/justitie-wil-ingrijpen-bij-bedrijven-die-digitale-beveiliging-niet-op-orde-hebben

The challenge is daunting as organizations lack both the means and the incentives to mitigate relevant vulnerabilities in a timely fashion. The CISO of a Dutch hospital recently told us that patch deployment in their organization takes three months, on average. After seeing how Wannacry affected other health institutions, she initiated a crisis management procedure in order to patch that specific vulnerability faster than normal. Even with this peak effort, it would take the organization one month to deploy the patch. The question is: why does it take a month to patch a single vulnerability, despite the organization's full effort and best intentions?

Patch your software or stay vulnerable? What is riskier?

Technically speaking, deploying a security fix is a problem that should not exist. Empirical studies (Dashevskyi et al. 2018) have shown that, in almost all cases, the lines of code that need to be fixed can be counted on the fingers of one hand. The ideal fix process would be that a developer would patch these lines, eliminate the unwanted behavior, recompile the code, and provide users with the fixed version. Users, be they individuals or organizations, could replace the old version with the fixed one with close to zero downtime and zero functional tests, as the two versions would essentially be functionally identical (except for the unwanted behavior).

Unfortunately, the other substantial empirical evidence shows that this is not happening. A study on mobile applications (Huang et al. 2019a) showed that seemingly trivial updates in the underlying libraries would have broken the application in almost 50% of the updates: toss a coin and, in case of 'head', your application breaks. It is no surprise, then, to find that companies may delay the deployment of the not-vulnerable version by months. We might blame these companies, and security experts typically do, but we claim that there is a more fundamental obstacle: a vendor should fix only what is needed and nothing else. This option is often not available for a large population of products used by companies. Software used by companies has a long life in the field (Figure 1 in the cited work by (Dashevskyi et al. 2018) show a ***mean*** lifetime of eight years and a long tail of 14 years). The product evolves and the security fixes are in practice bundled by suppliers with `improvements' to keep abreast of the competition or gain new customers. Redhat Enterprise Linux started 2012 and will reach its extended end of life midway this project. Siemens WinCC V4.0 (software for controlling industrial control systems) was supposed to have died in 2004 (after seven years in service according to Siemens' plan) but the latest migration instructions from Siemens date from May 2020.

[3] Fruhlinger, J. (2020). Equifax data breach FAQ: What happened, who was affected, what was the impact? CSOonline. At https://www.csoonline.com/article/3444488/equifax-data-breach-faq-what-happened-who-was-affected-what-was-the-impact.html.
[4] https://www.onderzoeksraad.nl/en/page/17171/beveiligingslek-citrix

A Way Forward

In short, organizations face a fundamental risk tradeoff: balancing the risk of not patching versus the risk of patching. The latter incurs ongoing business continuity disruptions, the former incurs a potentially catastrophic compromise event (Ioannidis and Williams, 2012). Since patching is risky and the risk of not patching is unknown, though frequently nothing happens, the incentive is for organizations to patch slowly. It is a "devil you know versus the devil you don't" situation. With the disruptions to IT management arising from the global pandemic of 2020, patching is at risk of being further neglected even while vulnerabilities continue to be discovered.[5]

The way to get out of this catch-22 is to **radically change the risk governance of patching**. That is the objective of our proposed THESEUS project. While major advances have been made with automatic updating in the consumer space, these have not translated—and more importantly, cannot be translated—to the enterprise space. Changing the risk of patching for enterprises means to develop breakthroughs at three interdependent levels:

- **Systems**: drastically **reducing risk** of patching via new techniques in automatic vulnerability and patch triaging, as well as automatic patch generation with live update for cases where critical patches pose unacceptable availability risks.
- **Enterprises**: more efficiently **quantifying risk** of patching by assessing and aggregating the results of the patch triaging as a way to estimate exploit likelihood in a coherent picture that accounts for different attacker models and functional impact.
- **Governance**: more effectively **managing risks** of patching by introducing incentive mechanisms via notifications and information sharing, sector-wide benchmarks of patching speed, and potentially legal instruments.

Our proposed project sets out to (1) bring advances from the lab to real-world settings by working with a consortium of partners that contribute people, data, and pilots to the project; and (2) replace the infeasible and counterproductive idea of mandatory patching with a larger and more sophisticated set of governance interventions across different levels: system, enterprise, and sector. Our target is that enterprises can have a secure environment within which people can go about their work, without obstruction or the need for the organization to carry intractable risks. Progress in this area not only helps to thwart major impacts, but also generates benefits, for example in terms of making it easier and more secure to network within supply chain systems. Furthermore, it generates benefits for security providers, also present among our partners, who can market solutions to enlarge their client base.

To illustrate how we bring these different disciplines into a more holistic answer to this problem, let us take a look at the dominant risk assessment methodology for rating vulnerabilities, the so-called Common Vulnerability Scoring System (CVSS). This methodology is not effective in prioritizing vulnerabilities, as it rates too many vulnerabilities as 'high' or 'critical'. This reflects the underlying incentives of vendors and security firms who may overstate the potential impact of a vulnerability ("we warned you!"). Only a tiny fraction of these are ever exploited in the wild (Allodi & Massacci, 2014). Therefore, organizations are forced to dilute their patching efforts across hundreds or even thousands vulnerabilities, slowing down patching processes across the board. Our project Theseus will improve on this situation by developing and testing with our partners (1) high-assurance automated patching solutions for critical vulnerabilities, to protect the organization while testing the vendor patches, (2) better automated tools to triage vulnerabilities and patches, (3) a risk assessment approach that consumes and aggregates the results from the triaging to estimate exploit likelihood in a coherent picture for patch prioritization, (4) an information-sharing mechanism for vulnerabilities that are exploited in the wild and (5) a governance mechanism and potential regulation to provide better incentives and checks and balances around vendor vulnerability ratings and notifications (e.g., allocation of liability in case of over-rating or under-rating), such as requiring that high or critical security vulnerabilities are addressed in a dedicated patch and/or follow certain formats to allow live updating, and liability allocation (or reversal of the burden of proof) where organizations do not implement known critical patches.

> "Applying these patches typically requires rebooting the kernel, which results in downtime and loss of state [...] Rebooting can lead to momentary interruption or cause unexpected complications, which means that reboots are commonly specially scheduled and supervised. Since rebooting. Since rebooting is disruptive, many system administrators delay performing these updates, despite the greatly increased security risk—more than 90% of attacks exploit known vulnerabilities." **Prof.dr. Frans Kaashoek, MIT**
> https://bit.ly/3hx4ZUF

The common thread of these breakthroughs is that we currently do not have the right tools and approaches to incentivize and defend organizations against known vulnerabilities within an actionable

[5] PricewaterhouseCoopers (PwC). "Managing the impact of COVID-19 on cyber security". March 2020. At: https://www.pwc.co.uk/cyber-security/pdf/impact-of-covid-19-on-cyber-security.pdf

timeframe. This project approaches the problem holistically from multiple disciplines and with various societal partners to provide and test solutions to overcome the different interconnected causes of slow patching.

State of the Art

Our breakthroughs build on the state of the art in domains across different disciplines: (1) vulnerability and patch prioritization; (2) automated patch generation; (3) patching behavior; (4) vulnerability measurement and notification; (5) enterprise risk management; (6) governance and incentive structures; and (7) the legal framework and public oversight.

Vulnerability and patch prioritization

At the technical level, the state of the art lacks the ability to automatically determine the *actual* (rather than the theoretical) exploitability of a vulnerability, the ability to assess the effectiveness of a patch, and the availability of defenses that provide practical alternatives for patching. State-of-the-art research (Khazaei, 2016) mostly focuses on text analysis from databases such as the Common Vulnerabilities and Exposures (CVE) database. This means that the original vulnerability analysis itself is manual, which involves a substantial risk of overlooking particular exploit vectors. Moreover, the patch is not analyzed at all, which means we cannot tell whether the patch is effective (i.e., actually addresses all possible ways in which the vulnerability can be exploited).

Automated patch generation and live updating

In cases where patches are not yet available, are ineffective (as shown by our vulnerability classification system), or have too much of an impact on availability, organizations need an alternative to be able to quickly address security holes. While sanitizers are available to automatically mitigate many classes of vulnerabilities using compiler instrumentation, their high performance overhead is unacceptable for many production settings (Song et al., 2019). State-of-the-art performance-oriented defenses consist of ASAP (Wagner et al., 2015) and Senx (Huang et al., 2019b). ASAP selectively applies compiler instrumentation based on a run-time profile, but this approach cannot adapt to a particular known vulnerability, and does not offer any security or availability guarantees because the selection is purely performance-based. Senx is more targeted, but relies on manually-specified properties and is only effective for particularly simple software in which the data flow can be statically traced. Another needed advance is live updating. Installing patches means the programs lose all state. This causes a cascading effect on other services depending on this program, increasing the effort and the availability impact even for very small patches. Live updating can solve this problem, but it is a very challenging task. General-purpose live updating for C and C++ programs is currently only possible for programs that have a well-defined structure and with the help of manual annotations for each patch (Giuffrida et al., 2016). Even production solutions that specifically target security patches such as Ksplice (Arnold et al., 2009) require per-patch manual effort from the vendor. Moreover, all the existing live update solutions provide limited update safety guarantees. All in all, we need to innovate to develop more practical live update solutions.

Patching behavior

Most empirical work on how and when actors patch has been conducted at the level of consumers, which also includes the lower end of SMEs[6] (Canali et al., 2014; DeKoven et al., 2019). The end-user models explaining patching behavior are not suitable for describing patching behavior at the organization level. For example, Li et al. (2019) investigated patching practices of system administrators who manage organizational infrastructures and found that "*the particular factors considered and the actions taken by system administrators are significantly different across all stages of the update process [compared to end-users].*" There are also more in-depth studies of organization patching processes like the one conducted by Pandey and Mishra (2014). The authors discovered that the main reasons for the failing process are "*user resistance to change, lack of ownership, failure to educate and communicate the importance of change management, and power dynamics within the IT department*". On the other side, Gerace and Cavusoglu (2009) surveyed 114 practitioners from various sectors. The practitioners thought that the identification of vulnerabilities and network scan pre-deployment activities were more critical to successful patching than testing patches prior to deployment or having senior management support.

Large-scale vulnerability measurement and notification experiments

Various studies have innovated on internet-wide vulnerability scanning and then notified the affected actors with information about the found vulnerabilities. Different channels have different degrees of effectiveness in incentivizing the recipient to patch (Stock et al, 2018). Kotzias et al. (2019) conducted a large-scale and long-term enterprise security study. Over three years, they collected data from 28,000 enterprises across 67 industries that own 82 million client hosts and 73 million public-facing servers. According to the study, "it takes over six months on average to patch 90% of the population across all vulnerabilities in the client-side application," and it takes up to 9 months on average

[6] Elissa Redmiles. "Why Installing Software Updates Makes Us WannaCry". 2017. At: https://www.scientificamerican.com/article/why-installing-software-updates-makes-us-wannacry/

to patch 90% of the enterprise servers. However, the organizational factors affecting this process are still unclear, making it challenging to find ways to reduce this time.

Enterprise cyber risk management

At the enterprise level, a security risk is a well-structured process. Yet, Information System architects often lack the necessary security knowledge to identify all appropriate security risks, at a lower system level in particular. Even experts sometimes forget to treat risks which might be relevant for the system. To mitigate this issue, industrial methods and standards are equipped with catalogues of threats and security controls. Essentially, catalogues are a form of knowledge reuse (Souag et al., 2015) created at the community level and made available to individuals. They can be seen as knowledge reservoirs designed to transfer knowledge across the community instead of the classical source of competitive advantage between firms (Argote & Fahrenkopf, 2016). However, the catalogues are mostly expressed at a high organizational level. For example, the PCI Security Standard Council just states in the PCI DSS Quick Reference Guide: ''*Entities should apply patches to less-critical systems as soon as possible, based on a risk-based vulnerability management program.*'' Yet, it is unclear how this should be transformed into a concrete patching schedule for an organization. Existing frameworks for management and governance in IT like COBIT, Global Technology Audit Guide, or ITIL provide detailed guidance on how the patching process should work and why it is beneficial to organizations. However, as recent empirical research shows (Durumeric et al., 2014; Kotzias et al., 2019), patching deployment is still a big issue in organizations.

Security governance and incentive structures

The acceptable level of security for organizations may vary between different sectors and governmental views. The research literature suggests a variety of incentives for organizations to invest in their cybersecurity measures: sectoral regulations (e.g., PCI DSS, SOX, HIPAA) and general cybersecurity regulations (e.g., GDPR requires adequate security when processing personal data, NIS directive sets requirements for critical infrastructures) (Gordon et al., 2015), customer's requirements, reputational harm resulting from disclosed cyber incidents or successful cyber attacks, information sharing initiatives, public rating/assessment systems (Naghizadeh and Liu, 2019), and cyber insurance policy requirements[7]. Despite the wide range of available incentives, we still face incentives misalignment which results in more vulnerable systems being used (Park, 2019)

Legal frameworks and public oversight

There are many (mostly sectoral) laws around the world requiring certain industry sectors or critical infrastructures to ensure adequate cyber security. In the EU the GDPR further requires organizations that process personal data to secure their data processing systems adequately. None of these laws contains any prescriptive requirements on patching practices. It is left to the relevant organizations to assess whether a relevant patch is required to ensure the required adequate level of cyber security. Not patching may, therefore, constitute a violation of cyber security requirements under law, a breach of contractual requirements, or may constitute negligent behavior (a tort) when the relevant organization was aware of the vulnerability and did not promptly implement the available security patch. The most noteworthy example here is the Equifax breach mentioned above, where Equifax failed to patch a critical security vulnerability after being alerted for at least four months, which resulted in a data breach exposing the names and physical addresses, dates of birth, Social Security Numbers, and other personal information that could lead to identity theft and fraud of over 140 million customers. This breach resulted in regulatory investigations which ended with a $575 million settlement with the US Federal Trade Commission and several states' attorneys general[8]. Equifax was also sued before the US Court[9], which found that Equifax violated state data security regulations. For similar examples under Dutch case law, we refer to Guidance issued by the Dutch Cyber Security Council on the duty of care of organizations regarding cyber security.[10] Fact is, however, that despite organizations being exposed to claims for damages (whether under specific regulations or under tort law to prevent foreseeable harm to others), patching practices remain 'patchy'. In this research, we want to investigate what legal instruments can be deployed to improve patching practices of companies preventing such potential third party damages (rather than addressing liability after the fact).

> Quote from the judgement of Court:
> "(...) Equifax knew for months it needed to patch its open-source code in order to keep its databases secure (...) and that it failed to do so. These allegations plausibly suggest that Equifax breached its legal duties to address all reasonably foreseeable risks to its data security (...), and to implement reasonably up-to-date patches to its software (...)."
> https://www.law360.com/articles/1030065/equifax-can-t-skip-mass-ag-suit-alleging-security-failures

[7] CYBECO D7.1: CYBECO Policy Recommendations. At: https://www.cybeco.eu/images/items/CYBECO-WP7-D7.1%20Policy%20Recommendations-v1.0.pdf.
[8] https://www.ftc.gov/news-events/press-releases/2019/07/equifax-pay-575-million-part-settlement-ftc-cfpb-states-related
[9] Commonwealth v. Equifax, Inc., No. 1784CV03009BLS2, 2018 WL 3013918, at *1 (Mass. Super. Apr. 3, 2018)
[10] Handreiking zorgplichten. Dutch Cyber Security Council. At: https://www.cybersecurityraad.nl/binaries/Handreiking_Zorgplichten_NED_DEF_tcm107-314470.pdf

Scientific, economic and societal breakthroughs

We propose a holistic approach to increase the speed of patching by combining breakthroughs in three areas: systems, enterprises, and governance. For each of the areas, we pose two overarching research questions (introduced below Table 1) representing knowledge gaps in the state of the art. Each question corresponds to a PhD project. Beyond these six questions, we articulate the main questions for valorization, where we integrate the findings from the six projects and test them in real-world environments. The table shows how these knowledge gaps relate to causes of delayed patching experienced by our stakeholders, and specifies which research questions analyze or address particular causes of delayed patching.

Table 1: Relationship between work packages, research questions, and causes of delays in patching[11]

	WP1 System		WP2 Enterprise		WP3 Governance		WP4/WP5 Valorization	
Causes	RQ1: patching	RQ2: triaging	RQ3: quantify	RQ4: analyze	RQ5: incentivize	RQ6: liability	RQ7: portfolio	RQ8: pilots
A. Vendor updates may unexpectedly break existing running services	🔨			✎	✎	🔨	💼	⚖
B. Updates to solve vulnerabilities may not be available in time	🔨			✎		🔨	💼	⚖
C. Real risk unclear as most vulnerabilities will not be exploited		🔨	🔨	✎			💼	⚖
D. Planned software maintenance misaligned with timing of patching		🔨	✎	✎			💼	⚖
E. Misaligned incentives: benefits of patching are not immediately visible		🔨	🔨	✎	🔨	🔨	💼	⚖
F. Unclear responsibility for vulnerable systems				✎	🔨	✎	💼	⚖
G. User/management resistance to patching	🔨	🔨		✎	🔨	🔨	💼	⚖
H. Cost of breaches (not patches) are externalized to third parties			✎		✎	✎	💼	⚖
I. Lack of regulatory oversight, also on software vendors				✎	✎	🔨	💼	⚖
J. Vulnerability information is not effectively shared or received			✎	✎	🔨	🔨	💼	⚖

Legend: ✎ - this cause will be analyzed in the corresponding RQ; 🔨 - this cause will be addressed by the RQ; 💼 - the portfolio of solutions will address this cause; ⚖ - to be decided together with the project partners.

Systems

At the systems level, we can solve some of the technical issues causing organizations not to patch, as well as support organizations in setting policies by generating more relevant information about patches. Two of the most pressing concerns are how to address a vulnerability if no suitable patch is available, including cases where the patch has unacceptable availability impact, and how to determine how harmful a vulnerability actually is in situ, allowing triaging to be adapted to the organization's computer systems.

RQ1: How can we protect systems without disruption when a patch is not available, not effective, or has unacceptable availability impact?

Current solutions have performance overhead that is unacceptably high for production settings (Song et al. 2019). We develop a new approach to apply targeted hardening for the specific vulnerability that needs to be mitigated (T1.4). In contrast to the state of the art in Senx (Huang et al. 2019b), we propose a safe and automated approach based on production sanitizers, while still checking only one particular vulnerability. As such, we advance the state of the art

[11] In this table we used icons by FontAwesome.

and offer automated vulnerability mitigation with better performance and availability than is currently possible, which we will demonstrate with new measurement techniques (T1.6). We also propose to offer patches that support safe live updates, greatly improving availability and reducing risk (T1.5). As a consequence, organizations will be able to patch sooner, reducing the window of vulnerability.

To address this research question, we build on the automated exploit generation (AEG) techniques developed in T1.2. We use the generated exploits from T1.2 and vulnerability impact from T1.3 to evaluate our solutions. We use the results from T2.2 and T2.3 to know which problems are most important to address in patching.

RQ2: How can we automatically predict which vulnerabilities have the greatest security impact in situ?

We propose to analyze the security impact of the patch automatically. In particular, we previously built the ParmeSan fuzz testing system (Österlund et al., 2020) that can efficiently find security vulnerabilities of many types by leveraging the many sanitizers currently available (Song et al., 2019). As a starting point, we develop network scanning techniques to provide real-world input for our prioritization system (T1.1) We apply the system to both the patched and unpatched software versions and develop automated exploit generation techniques (Avgerinos et at., 2011) to determine whether both versions can be exploited (T1.2) and what the impact of these exploits is (T1.3). We additionally extend the current state of the art by specifically focusing on bypassing mitigations by means of our prior experience in creating ParmeSan. We determine whether a vulnerability threatens availability and/or confidentiality, and to what extent. Such automated classification is not currently possible. The availability of a proof-of-concept exploit for the unpatched version clearly demonstrates the importance of applying the patch without delay, while the inability to find an exploit after patching gives some confidence that the patch makes exploitation significantly harder (though it cannot prove the absence of a much more complex exploit bypassing the patch). If, on the other hand, the patched version is also shown to be vulnerable, we present alternative strategies. As such, our automated vulnerability classification system greatly advances the scientific state of the art by building on our prior experience, while also having societal impact by greatly improving patch triaging, which despite its major role in maintaining both availability and security currently involves a lot of uncertainty. T1.3 is done in tight collaboration with T2.5, which identifies attacker models to be used when estimating potential impact

Enterprise

There is a general agreement in the risk governance literature for enterprises that a risk assessment requires to provide both impact and likelihood on the enterprise as a whole (Aven & Renn, 2010). To provide actionable decision support we need to further compare the current assessment with the speculative execution of the alternatives (patching in our case). Next to improved risk assessment methods, we need better empirical data on risk decisions around patching in enterprise environments.

RQ3: Can we quantify the enterprise overall cyber risk for immediate vs delayed action given the enterprise functional requirements and realistic threats?

Modern standards use ordinal scores for both impact and likelihood. Scientifically ordinal scores provide only limited information in terms of risk assessment (Cox et al., 2005) and may possibly mislead users into rank reversal of priorities. Most importantly, such scores provide only information on ordinal impact in the ideal, isolated analysis by the experts in charge of the rating. Further, vulnerability databases are riddled with `errors of unknown size' including Christmas and Holiday clearings of reports (Schryen, 2009). Several network-wide security metrics have been proposed by researchers (See Wang et al. (2017) for a comprehensive survey) but very limited work has been carried in terms of actual experimental effectiveness. The reason is simple: one can remove the patched systems from the sample (i.e., the excel file) produced from the scan, but this does not produce any experimental confirmation as it depends on the very expert judgement that produced the first assessment. It does not prove that the new system can withstand an attack of an adversary. In this respect we build on our recent work on using big data to estimate likelihood of a successful attack by a realistic adversary (Allodi & Massacci, 2017) towards a proper interaction model of a real enterprise network for which only theoretical models of interdependence exist. Yet, to be able to do so we need lower level data: can we build an experiment, even if speculative, in which we test independently what happens if we apply the patches?

The research will build upon the data collection activities that identify the key quantitative drivers for patching (T2.1-3) will be then used by the PhD student to identify a suitable model for an attacker (T2.4) and mechanisms for Risk Aggregation that must combine both uncertainty (e.g Bayesian, Dempster-Shafer or Stoke's logic of inaction) and goal models (T2.7). The models would then be empirically verified (T2.8) by combining the results of system analysis in terms of security and exploitability (T2.6). The final result would then be used to understand the collected data from the pilots (T5.2).

RQ4: What are the causal drivers at the organizational level of patching speed and how can these be leveraged to improve patching speed?

Most of the steps that organizations go through for patching are based on community best practices and heuristics, rather than empirical evidence. Patching processes can be slowed down by a lack of clarity in ownership and responsibility of the affected systems, as well as incentive problems. A risk-averse approach to testing a patch, to avoid breaking mission-critical systems, rewards longer testing phases. The longer vulnerability window that is the result of this approach, on the other hand, has no direct cost to any organizational unit. There is then a need to engage systems administrators, systems users, and those involved in internal oversight and audit. The project recognises the activities within organizations to secure their systems to mitigate known or suspected vulnerabilities, even if a patch is not applied. We will propose different ways for organizations to signal that a system is secure. We take the positive approach to recognize both the intention and achievement of organization efforts to secure their systems, parallel to externally-recognised approaches such as patching.

To address this research question, we will collect data about existing patching practices at the sector level (T2.1), and supplement it with quantitative analysis of large datasets of security patching cases within organizations (e.g., drawn from asset management systems) (T2.2). On top of it, we conduct qualitative interviews with organizations having different patching processes (T2.3) to understand their decision-making process around the remediation or patching of vulnerabilities. Based on the outcomes of these studies, the research will investigate the key driving factors explaining the variance in patching speed across vulnerabilities, systems and organizations (T2.4). The results of this work will also be used to identify a suitable model for impact (T2.6) and appropriate models of attackers (T2.5), their further empirical validation (T2.8), and to inform the development of suitable regulatory instruments and governance mechanisms (T3.7 and T3.8). Some of the developed solutions will be validated in a field study with involved organizations (T5.2) with a preparatory adaptation of the selected solutions to organizational settings (T5.1). We will also empirically investigate the best communication policies to present the proposed solutions to stakeholders (T5.3).

Governance

Beyond factors at the organizational level, there are governance mechanisms that influence patching behavior. Regulations and laws for specific industries and critical infrastructure require them meet defined standards of cyber security protections. However, there are no particular requirements regarding patching practices at the legislative level. Organizations decide for themselves how to manage patching, albeit under the pressures of the potential legal liability for damages due to non-compliance with the legal cyber security requirements and the various incentives structures and factors we discussed in the state of the art. In the governance theme, we will investigate how range of governance mechanisms support effective patching practices within organizations (incentives, evidence-driven sector-level guidance, organizational benchmarking, public-private partnerships and regulatory instruments).

RQ5: What governance mechanisms can be proposed to facilitate and incentivize faster patching?

Acknowledging the diversity of factors which can potentially slow down the patching process (as identified in RQ4), we will develop governance mechanisms to help organizations make improvements to their processes. This may range from better information sharing on the criticality and prioritization of vulnerabilities based on actual attacks, notification of observed vulnerable systems in their networks and responses on whether mitigations are in place, benchmarks on patching speed across organizations, and evidence-based protocols for patching. Cyber-insurance may also play a role. If the complexity of patching is too high or uncertain, an organization may choose to transfer the risk through insurance. The key is to empower organizations to make positive steps and find ways to address blockers to operating securely.

In this work package, we will use the data-driven approach to explore the level of vulnerability response and patching behaviour of various organisations with vulnerability scanning (T3.1). Using the data collected with vulnerability scans, we will develop a benchmark helping organisations to get an overview of their maturity level in patch management and vulnerability response (T3.2). Further, we will investigate the relevant incentives (T3.3) that can be used with our benchmark and facilitate the adoption of solutions developed in this project. Also, we will investigate the effect of vulnerability notification on patching and vulnerability response behaviour in organisations (T3.4). The next task will explore existing governance mechanisms (sectoral and general regulations, auditing, information sharing, and others) (T3.4). The PhD research will leverage the findings of the analysis of the key driving factors for improving patching process (T2.4, with input from T2.1, T2.2, T2.3) and propose new or improve existing governance mechanisms to help organisations manage to patch risks (T3.6). The proposed solutions will be integrated into the portfolio of solutions (T4.1) and verified with the involved stakeholders and project partners (T4.2). Based on the feedback from the stakeholders, the solutions will be improved (T4.3). Some solutions will be selected for the field study (T5.2) and validated taking into account organisational settings according to collected data (T5.1).

RQ6: How are (current and draft) legal frameworks regulating cyber security and handling potential liability to third parties from security incidents resulting from unpatched systems and what legal instruments can be deployed to improve patching practices of companies preventing such potential third party damages (rather than regulating liability after the fact)?
This research line will build upon the exploration under RQ5 into existing governance mechanisms. The PhD research will inventory regulatory examples on cyber security requirements and patching practices (T3.5), as well as reviewing court cases involving violation of cyber security requirements and potential liability based on negligent behavior (a tort), in order to decide at which level and what type of regulatory intervention would be optimal. Some of the legal governance mechanisms are tied to insurance. We will analyse how post-breach insurance claim handling deals with the lack of patching of compromised systems (T3.6). The research will deliver concrete recommendations to legislators, both at the national and European level (T3.7). As these interventions would be inextricably linked to (i.e. complement) the review of governance practices under RQ5, these lines would have to cooperate closely and culminate in an overall portfolio of governance options (T3.8, in close collaboration with T2.4 and T3.3).

Valorization
After finding the main issues hindering the patching process for our stakeholders, and developing and integrating our solutions, the final issue is to make sure that our portfolio of solutions can and will be deployed in practice. In this final step, we again work closely with our partners. We first integrate all proposed breakthroughs in a coherent portfolio for our partners and wider set of stakeholders. Then, we initiate workshops in which our partners select solutions that they want to pilot in their organizations.

RQ7: How do we integrate the proposed solutions into a portfolio?
This project includes researchers from different fields, ranging from governance to computer systems, and we involve a wide range of societal stakeholders, which provides a unique opportunity to build an integrated solution for timely patching that provides technical solutions while also considering adoptance at all levels of organizations. The work to find causes and develop solutions at the individual levels is integrated into a portfolio of best practices, which will be tested by the stakeholders in our project and publicly disseminated through journal and conference publications, and any underlying programs will be made publicly available open source to allow organizations to freely use it and build on it.

RQ8: How do the solutions perform in real-world field tests?
The final step in our project is to perform a validation of the effectiveness of our approach in the field. Our consortium includes stakeholders in the critical sectors of healthcare and transportation, where improving patching processes is especially urgent for the security of the country as a whole. This provides the unique opportunity to test the validity and effectiveness of our breakthroughs in the field. This has a major impact both scientifically, because the project offers us access to real people and real systems in critical organizations for validation, and practically, because our approach is immediately deployed to improve security in real organizations. We anticipate that performing field validation can only be done for a subset of all solutions in the portfolio, because of the significant effort associated with field experiments, both on the side of the research team and on the side of our partners.

2.4 Approach/methodology
This project is unique in the sense that we depart from the idea that failure to install patches in time is either irrational organizational behavior or merely a technical problem. Instead, we recognize that there is an interplay between inadequate governance, organizational dynamics, and technical limitations, and that these issues must be addressed simultaneously to lead to effective patch deployment strategies. In light of the diversity of approaches of the various disciplines and the page limits of the proposal, we are constrained to high-level description of the research methodologies for each work package.

Project organization
The project is organized in five work packages (WPs): three research WPs where various disciplines are brought together for a holistic approach to the issue of patching and two WPs where the identified solutions are further developed, disseminated, and field tested with stakeholders. The research WPs coherently synthesize the work of 6 PhD researchers with backgrounds and methodological training from the fields of governance, law, system security, human factors, and risk assessment (see also Table 1).

In Systems (WP1), we design new methods to (1) automatically classify vulnerabilities in their real environment for triaging, (2) automatically determine the effectiveness of patches, (3) perform targeted hardening as an alternative for patching, and (4) deploy patches without the need for a restart. These methods will build on and extend state-of-the-art techniques for fingerprinting vulnerable software, fuzz testing, automated exploit generation, and software hardening. These methods together support organizations in deciding how to handle newly known

vulnerabilities, and provide alternatives where a patch is necessary but too risky. In particular, we may recommend to (1) delay patching if there is no meaningful risk in situ, (2) apply the patch if it is necessary, effective, and has limited availability impact, or (3) deploy automated hardening with live update in other cases.

In Enterprise (WP2), we combine methods from risk assessment with empirical methods from social science and the field of information systems to understand and support organizational-level risk decision making. In particular, we will use both surveys and in-depth interviews in case studies to understand organizational experiences with important patch deployments. This will lead to an overview of the main factors that determine whether patch deployments will be successful, an important input for developing solutions. Additionally, we leverage our access to their systems to investigate patch levels and the severity of unpatched vulnerability in vivo, as a basis for a technical investigation of causes. We can then use this information to provide a method for risk assessment at Enterprise level.

In Governance (WP3), we combine empirical methods from computer science (vulnerability scanning) with methods from social science (governance, public administration) to analyze actor behavior and incentives and with legal analysis to identify how legal factors influence the behavior of organizations in various sectors. The vulnerability scans across organizations in different sectors allow us to see patching speeds. These can be connected with the characteristics of the organizations (e.g., size of IT footprint). Similar to prior work by the applicants, this allows us to develop empirical patching benchmarks that informs actors how their patching behavior compares to others in their sector and beyond. Also in line with our prior work, we will conduct information sharing experiments where we notify organizations about observed vulnerable resources in their networks. In parallel, legal analysis will identify relevant regulatory frameworks and incentives in an international comparative approach. This inventory will then help design potential legal changes realigning the incentives of manufacturers, enterprise users and third parties.

In WP4 (Field Transfer and Knowledge Utilization) and WP5 (Field Validation), our consortium engages with the research findings. A wide range of stakeholders bring expertise from the patch deployment process across different roles, including enterprise users, vendors, consultants, and government agencies. These stakeholders provide a valuable source of input to determine which causes of delayed patch deployment are most pressing, and which solution directions are most viable in practice. A set of these solutions will be selected by partners for field tests in their own environment. The field tests will be measured in a structured manner, to gather evidence of their effectiveness in real-world settings from a combination of technical and socio-technical perspectives..

Again, we do not limit ourselves to governance policies or to technical solutions. We recognize that problems in one domain may need a solution from another, and our interdisciplinary consortium supported by a wide range of stakeholders puts us in a unique position to be able to achieve this. For example, we propose technical solutions to automatically classify vulnerabilities, supporting enterprise risk classification, which in turn serves as a basis for better decision making procedures at the organizational level.

Stakeholder involvement

Stakeholder involvement in every step is critical for the success of this project. We have ensured commitments from 15 stakeholders for in-kind contributions valued at 661.588 euro. At the start of the project, stakeholders will be involved to find out the causes of slow patch deployment and their needs with regards to our final results, to ensure effective knowledge transfer at the end. While developing solutions, we need stakeholder involvement to provide input data for our research as well as a testbed to evaluate our results, including the use as a basis for a digital twin. In the final steps of the project, we decide together with stakeholders which solutions arising from the field transfer step are most promising for quick deployment, and together with them perform validation in the field of those solutions. Given the heavy involvement of stakeholders, we consider the building of our consortium and collaboration with our stakeholders to be an important part of our approach.

To allow involvement of a broad range of different types of stakeholders without creating an unworkable situation, we involve stakeholders in two different roles: a smaller group of partners that is directly involved in our research, and a larger group of advisory board members that helps ensure our research properly addresses the problems with patch deployment experienced in the field. Partners are further divided based on the contributions they offer to our research:

1. **Stakeholder board member**. The partner actively participates in a stakeholder board established after the award of the project, including review of interim results and providing feedback on project direction.
2. **People**. The partner helps recruit employees involved in the patching process for interviews and other forms of information transfer. The partner's expertise will help us in studying common practices in patching at the organization, enterprise, and system level. In addition, the partner participates in the project meetings and will have access to the project results, in the form of reports, benchmarking solutions, software, improved procedures, data, and presentations.

3. **Data**. In addition to the first role, the partner also provides the researchers data about the patching process, for instance in the form of logs, incident reports, etc.
4. **Pilot**. In addition to the second role, the partner plays an active role in the process and participates in the studies of current activities and selects one or two pilots where we evaluate the solutions developed in the project in real-world settings.

For each of these roles, partners will be involved in our research on an as-needed basis, as outlined below in the sections on specific domains of causes and solutions. The advisory board will be updated with the progress in our project on a biweekly basis, and have the opportunity to comment on our plans. Partners are also invited to these advisory board meetings. This heavy stakeholder involvement allows us to benefit from the stakeholders' experience with patch deployment from their various perspectives and ensure that the system will be suitable for deployment in practice.

Integration

In this project, we address the patching problem from several angles, and, at this step, we will combine the developed domain-specific solutions in an integrated portfolio of measures. This portfolio will serve as an inventory of measures where an organization can select the ones that fit organizational context or can be adapted to it. The solutions will come with instructions and materials necessary for their implementation.

Each organization is different, and a single integrated patching protocol and technical toolchain is not realistic. Instead, we develop a range of interventions that allows organizations to pick and choose based on their particular needs. Our diverse group of stakeholders includes organizations with in-house systems that need patching, companies that manage other's networks, and government organizations that regulate computer security. As such, we will be able to identify a broad range of concerns with patching processes that will be shared by organizations outside our stakeholder group. In our methodology, we will map these concerns to the solutions we provide at each of the various levels considered in this project. Our methodology eventually results in a workflow tailored to the concerns of a particular organization.

Field deployment

Together with our industrial partners, we conduct a mapping exercise to map the developed solutions to the organizational contexts and goals of our partners. This mapping contest will help us to shortlist the most applicable solutions and test them in real settings by conducting several field studies. The feedback from the field studies will be used to finalize the instructions and design on the tested solutions and address some common issues throughout the portfolio (e.g., more specific examples, level of details in instructions, representation style, etc.) In this way, we will also create real application cases that can motivate other organizations to test our portfolio and find a suitable solution for their case. We apply our methodology to design a tailored patching portfolio for the stakeholders participating in the field validation including the selected solutions and implement it in their organizations. To evaluate our approach, we document difficulties in the implementation process and we measure patch deployment speeds as well as security incidents both before and after implementation to determine and improve ease of implementation and effectiveness of our approach. Additionally, we interview the people involved in the patch deployment cycle to determine which obstacles they encounter. Feedback from the field validation process allows us to further improve our approach, as well as to compile evidence to convince societal stakeholders that it is safe and worthwhile to implement and improve their patch deployment process.

2.5 Knowledge utilisation

2.5.1 Project deliverables (Output)

This project will result in a diverse range of output, with research value and practical value. The research itself and the interactions with our stakeholders will result in reports describing the findings and proposed solutions. The reports will be shared with our stakeholders within the consortium, and to publicly disseminate our results will also be published in top-tier scientific and professional conferences, removing confidential details where necessary. In addition to these reports, our output consists of data sets and source code. Where possible, data sets gathered during the project will be published after removing confidential and privacy-sensitive information. We will write source code for tools to assist the patching process, designed as part of our systems-related research. This source code will be published along with the related conference papers.

A concrete list of all deliverables can be found in Section 4.2 as part of the work package specifications.

2.5.2 Project Outcome & Impact

The stakeholders involved in our project as partners and advisory board members are eager to improve patch deployment processes for themselves and/or their customers. We will develop new governance models for patch

deployment and supporting tools, and the main outcome of this project will be our stakeholders supporting these models and consequently achieving faster deployment of critical patches, resulting in better security, as well as less patch-related downtime, resulting in higher availability and lower costs.

The eventual impact of this project will be widespread adoption of our new governance model and tools, whether completely or in part. In particular, government agencies and consultancy firms involved in our project will base their advice on best patching practices on our governance models, spreading the impact beyond the partners involved in the project. This is facilitated by the fact that we make all necessary tools available as open source software, easing adoption both nationally and worldwide. This will result in critical infrastructure with software that is more up-to-date on security patches, reducing the risk of devastating cyber attacks. In addition, organizations will be able to provide higher availability at lower cost. The governance model and tools will be complemented by concrete recommendations to legislators -- both at the national and the European level -- informing what legal instruments (including self-regulation) can be deployed to improve patching practices, in order to prevent potential third party damages (rather than addressing liability after the fact).

2.5.3 Utilisation plan

Initially we will conduct a workshop to gather the relevant stakeholders and collect their requirements in the context of our project. The involvement of stakeholders from the beginning until the final stage is the key point for the project. We will use the knowledge developed in our research to promote the capability of next-generation professionals and researchers within the project domain. We will also organize regular meetings with stakeholders. These will also be used to shape research summary briefs aimed at professionals working in both practice and policy.

To ensure the relevance and practicability of the research process and results, we will involve an advisory board consisting of high-quality professionals and stakeholders. It will also serve as an additional, practice-driven channel for promoting the project results. At the end of the project, we will organize a workshop to showcase the outcomes of the project (e.g., explain the portfolio and guidelines and share the experience of companies that participated in our field studies) and work with organizations to identify where portfolio measures can be applied in practice.

In our project, we regard open science and open-source code as the default rule. We will provide open free access to any created document, presentation, and paper published in scientific and professional conferences, journals, or magazines. The source code that we develop in the project will be made available to a wide audience. Moreover, the outcome of the project, the portfolio of measures, and accompanying guidelines explaining how to implement and use those measures will be freely available at the project web site.

Concerning publication venues, we will target all the major scientific conferences and journals in the fields of Governance (e.g., Governance, Government Information Quarterly) and human factors (e.g., SOUPS, WEIS, CHI), Enterprise (e.g., Risk Analysis, European Journal of Operations Research, Information Systems Research), and Computer Systems Security (e.g., IEEE Security & Privacy, ACM CCS, NDSS, USENIX Security, OSDI, SOSP, TSE) and empirical software engineering (TSE, ESEJ). We will also target all the relevant national and international professional conferences and magazines such as the ONE Conference, dcypher Symposium, SURF events, ENISA events, KPN Digital Dutch Event, and KPN Journal.

Toward the latter half of the project we will carefully develop openly-accessible case studies of challenges and successes for practitioner communities, combining research evidence and experiences of our stakeholders. This will support practitioners to relate enabling research to their own activities. Proposed regulatory instruments will be documented in a white paper to support longer-term dialogue with key regulators, the Dutch Cyber security Council, cyber regulatory experts and stakeholders, both on national and European level.

To promote knowledge sharing, we will maintain our presence on the Internet via the project web site, in social media, and an open access repository of our research products and practitioner-focused outputs. (We have budgeted for professional support in these activities.) This will support lasting accessibility for our open science outputs and tools, and summary briefings and case studies. We will also rely on our network of contacts to foster media coverage of THESEUS and disseminate our findings to other academic communities and the general public.

The consortium also has a proven track record of transferring research results to industry and national/international organizations. For example, our research on regulatory & governance led to the guidance issued by the Dutch Cyber Security Council on a “duty of care for cyber security”, which provides guidance for companies as to their patching obligations as well as the potential liability in case of lack of patching of known vulnerabilities as well as to the “White

Paper - Towards harmonized duties of care and diligence in cyber security".[12] This served as the basis for the discussions between the EU member states at the European High Level Conference on Cybersecurity in 2016, as chaired by The Netherlands. Companies like Microsoft, Google, and Mozilla have incorporated our proposed security defenses in their products. Companies like Intel are running VU Amsterdam's MINIX 3 in all their CPUs and have revisited their patching practices in light of our recent vulnerability disclosure efforts. Organizations like FreeBSD and Oracle's Ksplice have explicitly sought implementation of some of the research ideas from our prior live update work. The Dutch government and other national organizations are using our developed security benchmarks for hosting and ISPs. Our prior findings influenced the current version of the CVSS (Common Vulnerability Scoring System) industry standard which is used by the US Federal Government and all credit card companies as a mandatory standard for software assessment.

2.6 Research through Design (if applicable)

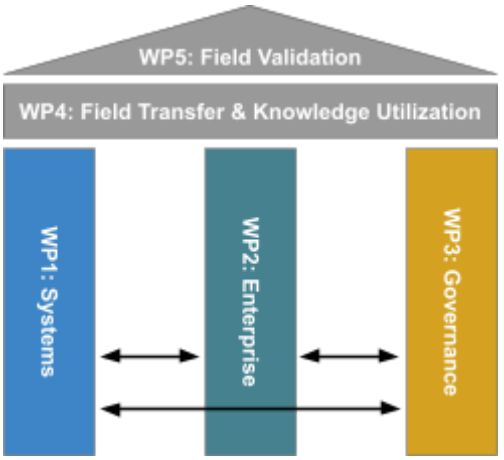
Not applicable

Section 3 - The consortium

[12] Dutch Cyber Security Council, Handreiking zorgplichten. Online at:
https://www.cybersecurityraad.nl/binaries/Handreiking_Zorgplichten_NED_DEF_tcm107-314470.pdf
Verbruggen et al., *Towards Harmonised Duties of Care and Diligence in Cybersecurity*
Cyber Security Council, European Foresight Cyber Security Meeting 2016, pp. 78-107
https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2814101

Section 4 - The work plan

4.1 Overall work plan



Our project starts from the idea that patches are not being applied for a variety of highly diverse reasons, all of which need to be investigated and addressed to be able to solve the problem of delayed patching. We therefore take a multidisciplinary three-pronged approach, where we first investigate causes and proposed solutions within three distinct domains: systems (WP1), enterprise (WP2), and governance (WP3). To achieve our ambitious goals within the six-year time path, we organize the tasks such that these research directions can start in parallel. However, there are interdependencies because research activities in one work package support the work in another. For example, the results of the vulnerability scanning in WP1 (T1.1) feeds into Attacker Models Identification in WP2 (T2.5). T1.1's toolchain also feeds into the development of vulnerability scans across organizations in WP3 (T3.1), which forms the basis for patching benchmarks (T3.2). These relationships are described in the Work Package descriptions below. Further integration is achieved by PhDs in one group taking on a task in another WP. For example, PhD1 (VUSec) is taking on T2.6.

Using the results from WP1-3, we integrate the solutions from the various domains into a coherent portfolio for field transfer and knowledge utilization (WP4) and we validate the effectiveness of a selection of our integrated solutions in real organizations (WP5).

4.2 Work packages

Work Package 1 - Systems factors

Work package leader:	VUSec	**Involved partners:**	TUD, F&ECyber@VU
Start date:	M1	**End date:**	M60

Objectives of the work package:
Keeping software up-to-date is important for security, but comes with a risk of downtime. In the context of critical sectors, systems must be available 24/7. As such, organizations triage patches: those that introduce too much (uncertainty about) downtime compared to their security benefits will be postponed, possibly indefinitely. This extends the window of vulnerability. Unfortunately, assessing the severity of a vulnerability is hard and we often find vulnerabilities considered harmless to be dangerous after all. In this work package, we develop techniques to determine whether technical factors prevent timely patch deployment. We develop techniques to determine how much security a patch offers, how it may affect availability, and if viable (temporary) alternatives to patching exist. The results serve as a basis for new patch deployment strategies in the later work packages, and the techniques can be used by system administrators to improve patch triaging and prevent risky delays.

Description of activities:
T1.1: Use network scanning to generate an overview of vulnerable software (M1-M6) [Leader: VUSec]
We will develop a system to scan the network of stakeholders to find the software versions in use by searching for fingerprints. We then apply this approach to evaluate the extent of the problems in practice for the stakeholders in our consortium, by determining which patches are missing. This information serves as input to the later tasks.
T1.2: Generate exploits for vulnerable software (M7-M24) [Leader: VUSec]
We will develop automated exploit generation (AEG) techniques and adapt them to the specific characteristics of healthcare and transportation. We then use them to find exploits for the missing patches of T1.1.
T1.3: Evaluate severity of unpatched vulnerabilities (M25-M36) [Leader: VUSec]
We classify the characteristics of the exploit generated in T1.2, including the type of vulnerability, the impact (confidentiality/integrity/availability), and the extent of data that can be leaked. The output of this task serves to prioritize patches based on security impact and provides input to the work of WP2 together with the organizational findings of T2.2-T2.4 to characterize explicit threat models.
T1.4: Develop targeted hardening as a potential alternative for patching (M25-M42) [Leader: VUSec]
Zero-day vulnerabilities do not have a patch available immediately, and extensive patches may be unsuitable for quick deployment due to availability concerns. Hardening techniques mitigate vulnerabilities by having the compiler insert checks, but are rarely deployed due to their performance impact and the risk of reducing availability when triggered by harmless undefined behavior. We develop mitigations specifically targeting known vulnerable code, using the results from T1.2 to test. Due to selectiveness, performance and availability will be better than traditional approaches. This serves as a baseline to compare patch impact, and as an alternative when a patch is not available or cannot be applied. ***T1.5: Develop live update for targeted hardening (M43-M54) [Leader: VUSec]***
We develop live update techniques which build on the characteristics of our targeted hardening patches from T1.4 to be able to apply these patches to a running system to perform live updates.
T1.6: Measure and compare effectiveness of patches and targeted hardening (M55-M60) [Leader: VUSec]
To triage patches, we must know whether they effectively mitigate a vulnerability. We further develop AEG from T1.2 to circumvent defenses and apply them to patches identified as important in WP1 to evaluate their effectiveness. We compare this against the unpatched impact of T1.3 as well as the impact of the hardening methods of T1.4 and T1.5 as a basis for deciding whether a patch is worthwhile, and whether targeted hardening is a viable alternative.

Expected output:
D1.1: Network scanning tool (T1.1; M6)
D1.2: Network scan report for stakeholders (T1.1; M6)
D1.3: Vulnerability classification prototype (T1.2/1.3; M36)
D1.4: Vulnerability classification report (T1.3; M36)
D1.5: Targeted hardening prototype (T1.4/1.5; M42)
D1.6: AEG prototype (T1.6; M60)
D1.7: AEG report (T1.6; M60)
D1.8: Targeted hardening report (T1.4/1.5; M60)

Work Package 2 - Enterprise factors

Work package leader:	F&ECyber@VU	**Involved partners:**	TUD, VUSec
Start date:	M1	**End date:**	M72

Objectives of the work package:
The purpose of this WP is to close the chasm between the technical results of what can (or cannot) be patched and the organizational requirements in terms of what is desirable or feasible in terms of organizational and functional requirements. We first plan to observe empirically the driving factors of patching processes and then to develop innovative techniques to transform the exploitation indicators at the level of the individual machines (whose ensemble has been collected by WP1) to provide the decision makers with a global risk picture.

Description of activities:

T2.1: Analysis of stakeholder views on patch management in the target sectors (M1-M6) [Leader: TUD]
We will conduct a stakeholders survey to investigate their views on existing internal governance mechanisms regarding patching management and vulnerability response in the target sectors. The survey will explore existing patching practices, potential cross-sectoral information sharing, and vulnerability response. Our survey will be based on one of the leading frameworks defining software development and management processes (e.g., CMM, ITIL, COBIT).

T2.2: Quantitative study of security patching cases and co-factors (M7-M30) [Leader: TUD]
In this task, we will quantitatively analyze a large dataset of various security patching cases in regard to patching speed and other relevant parameters. We will investigate existing patching patterns and possible factors contributing to faster or slower patching processes. The outcomes of this task will serve as an input for T2.3-T2.4.

T2.3: Qualitative study of security patching cases and co-factors (M19-M42) [Leader: TUD]
Based on the findings of T2.2, we will select and interview organizations with different patching processes. We will conduct a semi-structured interview to investigate how a patching process looks for the stakeholders, e.g., if they use other remediation strategies to mitigate the security issue instead of patching.

T2.4: Analysis and definition of the driving factors for the effective patching process (M31-M48) [Leader: TUD]
In this task, we will analyze the outcomes of T2.1-T2.3 and identify the factors behind the effective patching process. These factors will be used as an input for T2.8 and T3.8.

T2.5: Attacker Models Identification (M25-M48) [Leader: F&ECyber@VU]
This task will work on the result of T1.1-T1.3 and T2.3-T2.4 to abstract what is the type of vulnerabilities actually present in the network and whether the engineering skills needed to attack them are actually compatible with a model of the potential attacker. This will provide a formal foundation to the preliminary threat models identified in the initial studies.

T2.6: Measure and compare availability impact of patches and targeted hardening (M25-M48) [Leader: VUSec]
To determine availability impact of mitigations, we need to know whether real workloads trigger the modified code. For security patches, only potential attacks should be affected, not legitimate workloads. We generate traces in the production environment and test them with both a patched system and a system protected using targeted hardening. Using the traces as a ground truth, we determine the extent to which both solutions affect availability of the system.

T2.7: Risk Aggregation (M37-M72) [Leader: F&ECyber@VU]
The purpose of this task is to create a model for the aggregation of risk and decision making that builds upon the theory of Hierarchy Analytical Process that has been used in the past to elicit qualitative decisions from stakeholders when requirements cannot be totally ordered. The purpose is to adapt them to the case for security and risk decisions.

T2.8: Empirical Experiments on Risk Aggregation (M37-M72) [Leader: F&ECyber@VU]
The purpose of this task is to identify a suitable set of natural (case-control) and randomized experiments in patching that can be used to capture the enterprise risk and the odds ratio to be communicated to the stakeholders. There will be three dimensions for experimentation in the models. On the attacker side the models identified on T2.7 should be run against the experimental findings of T2.2 and against the requirements and processes identified by T2.1 and T2.3.

Expected output:

D2.1: Results of stakeholders survey (T2.1; M6)
D2.2: Quantitative study report (T2.2; M30)
D2.3: Qualitative study report (T2.3; M42)
D2.4: Main analysis findings (T2.4; M48)
D2.5: Attackers models for network scanning (T2.5; M48)
D2.6: Preliminary network scan and resulting impac
(T2.6; M36)
D2.7: Results of concrete risk aggregation for the identified networks/end user (T2.7; M72)
D2.8: Report on the empirical validation (T2.8; M72

Work Package 3 - Governance factors

Work package leader:	TUD	**Involved partners:**	TILT, F&ECyber@VU
Start date:	M19	**End date:**	M72

Objectives of the work package:
The goal of this work package is to investigate mechanisms at the governance level influencing patching behavior. There are some examples of the requirements for patch management in regulations for specific industries and critical infrastructure, and no special requirements at the legislative level. Thus, we will explore existing governance mechanisms related for patching and design potential regulatory instruments and governance solutions for managing patching risks. To support this process and make it evidence-based, we will 1) study the patching behavior across different organizations with network scans, 2) develop a benchmark based on the outcomes of these scans, and 3) empirically investigate potential incentives that can be coupled with the benchmark to motivate organizations to adopt best practices for effective patch management and vulnerability response.

Description of activities:
T3.1: Data collection from vulnerability scans (M19-M36) [Leader: TUD]
We will collect data from the Internet vulnerability scans and from the participating organizations (when possible we will do internal scans). These scans will explore the level of vulnerability response in organizational networks.
T3.2: Develop benchmark based on the outcomes of vulnerability scans (M31-M66) [Leader: TUD]
This task will analyze the outcomes of T3.1 and translate them into the benchmark for organizations. This benchmark will help organizations to get a clear view on their maturity level in patch management and vulnerability response.
T3.3: Investigate potential incentives based on the benchmark performance (M49-M72) [Leader: TUD]
We will conduct an empirical study with organizations on potential incentives that can be linked to the benchmark performance of organizations (T3.2). These incentives may range from increasing organizational awareness about their vulnerability response maturity up to regulatory requirements.
T3.4: Empirical experiment on vulnerability notification effect on vulnerability patching (M25-M60) [Leader: TUD]
We will conduct a controlled experiment to investigate the effect of different vulnerability notification approaches on the vulnerability response and patching in organizations.
T3.5: Explore existing governance mechanisms for patching and vulnerability response (M19-M30) [Leader: TILT]
We will investigate existing governance mechanisms (sectoral and general regulations, auditing, information sharing, and others). Moreover, we will review regulatory examples on cybersecurity requirements and patching practices and case law on liability due to not or delay in patching of known vulnerabilities.
T3.6: Investigate the role of cyber insurance in patching and vulnerability response (M37-M60) [Leader:TILT]
We will explore the role of cyber insurance in managing risks related to patch management and vulnerability response through interviews with insurance providers, brokers, and companies. We will also analyze to what extent the patch management is required by the existing cyber insurance policies available on the Dutch market.
T3.7: Design regulatory instruments for managing patching risks and vuln. response (M25-M48) [Leader: TILT]
Based on the findings T2.1, T3.5, and T3.6, the research will deliver concrete recommendations to legislators (both at national and European level) on what regulatory instruments (including self-regulation) can be deployed to improve patching practices in order to prevent third party damages (rather than addressing potential liability after the fact), these recommendations will complement the recommendations on internal governance of T3.8.
T3.8: Develop governance solutions for managing patching risks and vuln. response (M25-M66) [Leader: TILT]
To have an actual impact on the individual organizations, we design a set of governance solutions to help organizations to reduce the vulnerability time window and improve their patching processes. This task will be based on the results of T2.1, T2.2, T2.3, and T3.7. These solutions will help the stakeholders to build a picture on the impact and possible risks of the identified potential vulnerabilities and make them aware of possible trade-offs around patching decisions (e.g., system exposure time vs. patching speed, possible service outage, alignment with organizational change management policy). Selected solutions will be validated in a field study with the participating organizations in WP5.

Expected output:
D3.1: Dataset with the results of vulnerability scans (T3.1;M36)
D3.2: Draft vulnerability patching benchmark (T3.2;M48)
D3.3: Final benchmark (T3.2;M66)
D3.4: Report on the incentives for benchmark (T3.3;M72)
D3.4: Report on the incentives for benchmark (T3.3;M72)
D3.5: Notification experiment report (T3.4;M60)
D3.6: Review of governance mechanisms (T3.5;M30)
D3.7: Preliminary version of regulatory instruments and governance solutions (T3.7/3.8;M36)
D3.8: Final package of selected regulatory instruments and governance solutions (T3.7/3.8;M66)

Work Package 4 - Field Transfer & Knowledge Utilization

Work package leader:	VUSec	**Involved partners:**	All
Start date:	M37	**End date:**	M72

Objectives of the work package:
The idea of this work package is to gather the solutions developed in domain-specific work packages (WP1-WP3) and put together in an integrated portfolio of measures that can be used by individual organizations. The work package focuses on 1) preparing the solutions package with supplementary materials and instructions on how to implement it and 2) matching the most promising solutions to our industrial partners based on their possibilities, preferences, and organizational and technical context.

Description of activities:
T4.1: An integrated portfolio of solutions from WP1-WP3 (M37-M48) [Leader: VUSec]
We will iteratively combine the solutions developed in WP1-WP3 in an integrated portfolio aiming at improving patching processes from the organizational and technical perspectives. We include both the preliminary mitigation (which guarantees safety and availability) and the final patch to be delivered by the vendor (which may require availability interruptions and testing for safety). These solutions will be supplemented with application instructions and illustrated with practical examples. Partner NCSC is an end user for the portfolio that is the outcome of this task. They need the portfolio to improve patching among their constituency, just like they currently provide security advisories.
T4.2: Mapping participating organizations to the solutions ('beauty contest') (M49-M60) [Leader: F&ECyber@VU]
Once having the integrated portfolio of developed solutions, we will check how these solutions map to the partnering organizations in terms of necessary settings, relevant problems, opportunities and organizational willingness to try a corresponding solution in practice. Based on this 'beauty contest' of our solutions with industrial partners we will select the short list of solutions to be validated in WP5.
T4.3: Refined version of the validated solutions (M55-M72) [Leader: TUD]
We improve our solutions based on the outcomes of WP5 and develop the refined package of solutions that can be used by individual organizations.

Expected output:
D4.1: An integrated portfolio of organizational, technical and enterprise level solutions (T4.1; M48)
D4.2: Report on the mapping results and short listed solutions (T4.2; M60)
D4.3: Refined version of the solutions package (T4.3; M72)

Work Package 5 - Field Validation

Work package leader:	TUD	**Involved partners:**	All
Start date:	M49	**End date:**	M72

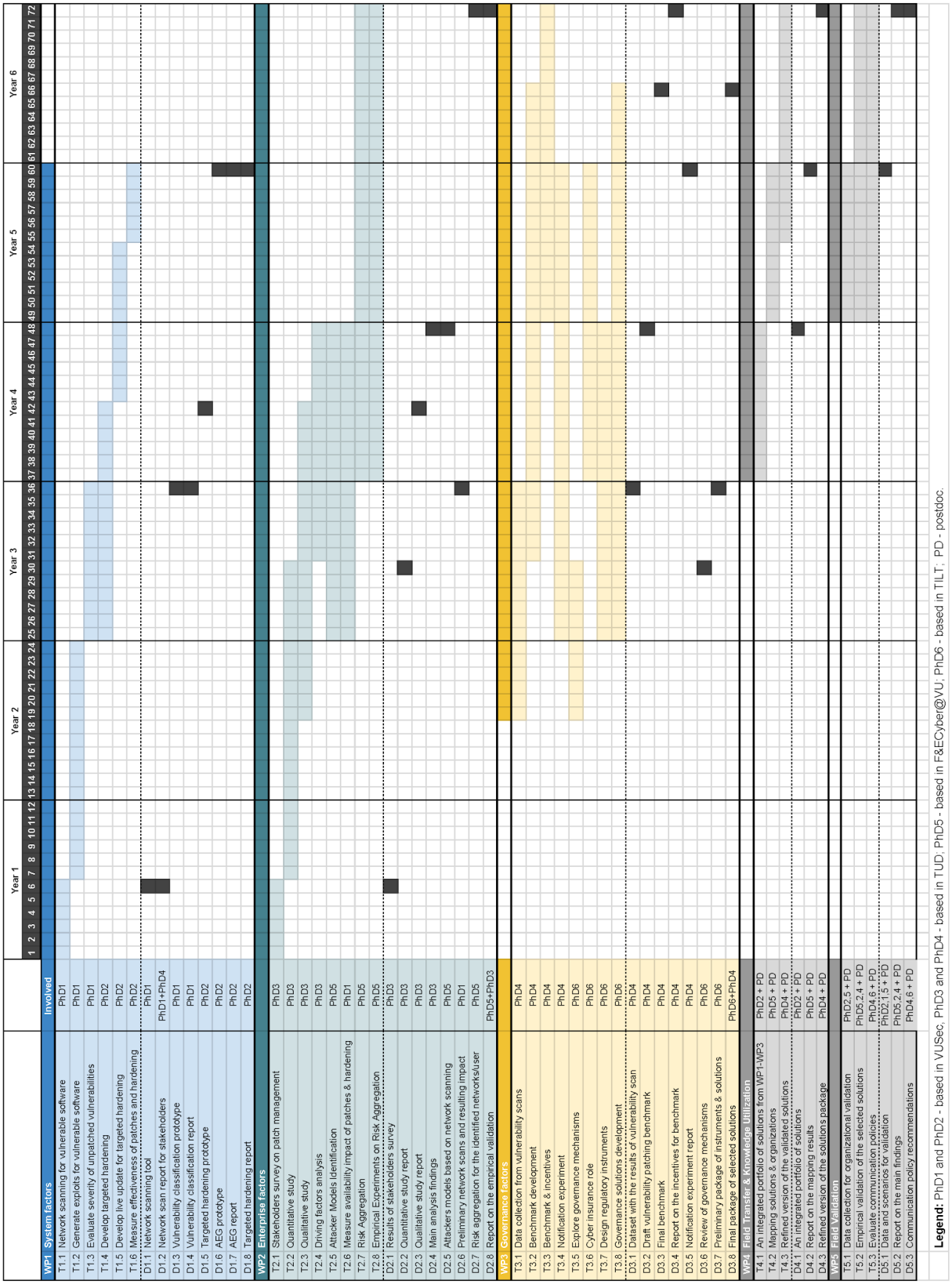
Objectives of the work package:
This work package focuses on the field validation of the selected solutions developed in the previous work packages. We will conduct a comprehensive study of the solutions in the real-world settings together with our industrial partners. The sustainability of our solutions and the real impact on the patching processes in the organizations will be in the focus of our field test. We will improve our solutions based on the feedback received from the field validation study. Additionally, we will research a suitable policy to communicate the recommendations to the involved stakeholders in an effective way.

Description of activities:
T5.1: Data collection for organizational validation (M49-M60) [Leader: VUSec]
To set the grounds for the field validation, we collect the necessary data per each case of solution-organization pairs. For example, we will need the initial data about organizational and technical settings to understand how the solution should be deployed in the organization and what adjustments might be necessary for the implementation.
T5.2: Empirical validation of the selected solutions (M49-M72) [Leader: F&ECyber@VU]
In this task, we conduct a field study with the organization(s) involved to validate the effectiveness of specific solutions developed in WP1-WP3. In T4.2 we select a short-list of solutions to be validated based on the specific settings and conditions of the participating organizations. Each team will conduct a comprehensive study of the selected solutions in the field settings wrt its specific expertise (e.g., VUSec will be responsible for the network scans and collecting systems metrics, etc.). We will combine a quantitative approach focusing on specific metrics at the organizational, technical and enterprise levels with a qualitative exploration of stakeholders' perception and experience through a series of semi-structured interviews (Analyzed by TUD). The outcomes of this validation will serve as an input in the corresponding work packages, where we take into account this feedback and develop the refined set of systems (WP1), enterprise (WP2), and organizational level solutions (WP3).
T5.3: Evaluate communication policies for the proposed recommendations (M49-M72) [Leader: TUD]
We will research suitable policies to communicate the recommendations with the necessary stakeholders in such a way as to clarify the risks, the benefits of the mitigation and/or patch.

Expected output:
D5.1: Data and scenarios for validation (M60)
D5.2: Report on the main findings of the field studies and implications (M72)
D5.3: Communication policy recommendations (M72)

4.3. Timeline, Milestones and Output

Figure 1 presents a preliminary THESEUS Gantt chart with the work packages, deliverables, and activities implemented by the first-stage and early-career researchers that will be involved in the THESEUS project. The detailed description of the tasks and outputs is available in Section 4.2. PhD1 and PhD2 are based at VUSec, PhD3 and PhD4 are based at TU Delft, PhD5 is based at F&ECyber@VU and PhD6 is based at TILT at Tilburg University. PhD researchers collaborate across work packages.



Section 5 – Literature references

5.1 Literature references

Allodi, L. & Massacci, F. (2014), Comparing Vulnerability Severity and Exploits Using Case-Control Studies. ACM Transactions on Systems Security. Vol.1. Also in industry: How CVSS is DOSsing your patching policy (and wasting your money). Black Hat USA 2013.

Allodi, L. & Massacci, F. (2017), Security Events and Vulnerability Data for Cybersecurity Risk Estimation. Risk Analysis, 37: 1606-1627.

Argote, L., & Fahrenkopf, E. (2016). Knowledge transfer in organizations: The roles of members, tasks, tools, and networks. Organizational Behavior and Human Decision Processes, 136, 146-159.

Arnold, J., & Kaashoek, M. F. (2009). Ksplice: Automatic rebootless kernel updates. In Proceedings of the 4th ACM European Conference on Computer Systems (EuroSys) (pp. 187-198).

Aven, T., & Renn, O. (2010). Risk management. In Risk Management and Governance (pp. 121-158). Springer, Berlin, Heidelberg.

Avgerinos, T., Tze Hao, B.L., & Brumley, D. (2011). AEG: Automatic Exploit Generation. In Proceedings of the 18th Network and Distributed System Security Symposium (NDSS).

Canali, D., Bilge, L., & Balzarotti, D. (2014). On the effectiveness of risk prediction based on users browsing behavior. In Proceedings of the 9th ACM Symposium on Information, Computer and Communications Security (CCS) (pp. 171-182).

Cox, L.A., Babayev, D. & Huber, W. (2005), Some Limitations of Qualitative Risk Rating Systems. Risk Analysis, 25: 651-662.

DeKoven, L. F., Randall, A., Mirian, A., Akiwate, G., Blume, A., Saul, L. K., ... & Savage, S. (2019). Measuring Security Practices and How They Impact Security. Proceedings of the 19th Internet Measurement Conference (IMC) (pp. 36-49).

Dashevskyi, S., Brucker, A. D., & Massacci, F. (2018). A screening test for disclosed vulnerabilities in foss components. IEEE Transactions on Software Engineering, *45*(10), 945-966.

Durumeric, Z., Li, F., Kasten, J., Amann, J., Beekman, J., Payer, M., ... & Halderman, J. A. (2014). The matter of heartbleed. In Proceedings of the 14th Internet Measurement Conference (IMC) (pp. 475-488).

Gerace, T., & Cavusoglu, H. (2009). The critical elements of the patch management process. Communications of the ACM, 52(8), 117-121.

Giuffrida, C., Tamburrelli, G., & Tanenbaum, A. S. (2016). Automating live update for generic server programs. IEEE Transactions on Software Engineering, 43(3), 207-225.

Gordon, L. A., Loeb, M. P., Lucyshyn, W., & Zhou, L. (2015). Increasing cybersecurity investments in private sector firms. Journal of Cybersecurity, 1(1), 3-17.

Huang, J., Borges, N., Bugiel, S., & Backes, M. (2019a) Up-To-Crash: Evaluating Third-Party Library Updatability on Android. In Proceedings of the 4th IEEE European Symposium on Security and Privacy (EuroS&P).

Huang, Z, Lie, D., Tan, G, & Jaeger, T. (2019b). Using Safety Properties to Generate Vulnerability Patches. In Proceedings of the 40th IEEE Symposium on Security and Privacy (S&P).

Ioannidis, C., Pym, D., & Williams, J. (2012). Information security trade-offs and optimal patching policies. European Journal of Operational Research 216.2 (pp. 434-444).

Khazaei, A., Ghasemzadeh, M., & Derhami, V. (2016). An automatic method for CVSS score prediction using vulnerabilities description. Journal of Intelligent & Fuzzy Systems, 30(1), 89-96.

Kotzias, P., Bilge, L., Vervier, P. A., & Caballero, J. (2019). Mind Your Own Business: A Longitudinal Study of Threats and Vulnerabilities in Enterprises. In Proceedings of the 26th Network and Distributed System Security Symposium (NDSS).

Li, F., Rogers, L., Mathur, A., Malkin, N., & Chetty, M. (2019). Keepers of the machines: examining how system administrators manage software updates. In Proceedings of the 15th USENIX Conference on Usable Privacy and Security (pp. 273-288). USENIX Association.

Naghizadeh, P., & Liu, M. (2019). Using Private and Public Assessments in Security Information Sharing Agreements. IEEE Transactions on Information Forensics and Security, 15, 1801-1814.

Österlund, S., Razavi, K., Bos, H., & Giuffrida, C. (2020). ParmeSan: Sanitizer-guided Greybox Fuzzing. In Proceedings of the 29th USENIX Security Symposium.

Pandey, A., & Mishra, S. (2014). Understanding IT Change Management Challenges at a Financial Firm. In Proceedings of the Information Systems Educators Conference (Vol. 2167, p. 1435).

Park, S. (2019). Why information security law has been ineffective in addressing security vulnerabilities: Evidence from California data breach notifications and relevant court and government records. International Review of Law and Economics, 58, 132-145.

Schryen, G. (2009) "A Comprehensive and Comparative Analysis of the Patching Behavior of Open Source and Closed Source Software Vendors," In Proceedings of the 5th International Conference on IT Security Incident Management and IT Forensics, Stuttgart, 2009, pp. 153-168,

Song, D., Lettner, J., Rajasekaran, P., Na, Y., Volckaert, S., Larsen, P., & Franz, M. (2019). SoK: sanitizing for security. In Proceedings of the 40th IEEE Symposium on Security and Privacy (S&P).

Souag, A., Mazo, R., Salinesi, C., & Comyn-Wattiau, I. (2016). Reusable knowledge in security requirements engineering: a systematic mapping study. Requirements Engineering, 21(2), 251-283.

Stock, B., Pellegrino, G., Li, F., Backes, M., & Rossow, C. (2018). Didn't you hear me?—Towards more successful Web vulnerability notifications. In Proceedings of the 25th Annual Symposium on Network and Distributed System Security (NDSS).

Wagner, J., Kuznetsov, V., Candea, G., & Kinder, J. (2015). High system-code security with low overhead. In 2015 IEEE Symposium on Security and Privacy (S&P).

Wang, L., Jajodia, S., & Singhal, A. (2017). Network Security Metrics (pp. 1-207). Switzerland: Springer.