Introducing responsibly self-healing into the incident management lifecycle

Alexandros Papanikolaou
a.papanikolaou@innosec.gr
Innovative Secure Technologies P.C.
Thessaloniki, Greece

Christos Ilioudis
iliou@ihu.gr
International Hellenic University
Thessaloniki, Greece

Vasilis Katos
vkatos@bournemouth.ac.uk
Bournemouth University
Poole, UK

ABSTRACT

In this paper we propose an approach for adopting the self-healing paradigm in complex networking environments. We argue that a straightforward application of self-healing capabilities may have an adverse effect on incident response due to the ill-understanding of the state of the system under protection. We sketch how the use of the Cynefin framework leverages the understanding of complex systems at the appropriate level of detail. In particular, we show how the framework can help to understand how the environment operates and to identify ways to improve its resilience and ability to recover from failures.

CCS CONCEPTS

• **Information systems** → *Information integration.*

KEYWORDS

self-healing networks, human-in-the-loop, cybersecurity incident handling

ACM Reference Format:
Alexandros Papanikolaou, Christos Ilioudis, and Vasilis Katos. 2023. Introducing responsibly self-healing into the incident management lifecycle. In *Proceedings of the 16th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '23), July 05–07, 2023, Corfu, Greece.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3594806.3594837

1 INTRODUCTION

Self-healing is a promising proposition for responding to network failures, outages, but also malicious threats. Self-healing networks have drawn researchers' interests in the last few years due to the explosive development of machine learning and artificial intelligence, which are key ingredients of self-healing capabilities [2, 7, 8].

As with all machine leaning enabled technologies, one of the dimensions of self-healing is the level of supervision [10]. Fully supervised self-healing systems require a high degree of engagement of the system administrator, whereas unsupervised systems do not. Within this range, there are also the semi-supervised self-healing systems that require human intervention in some parts of the workflow. However, when it comes to deploying self-healing networks for countering malicious cyberthreats in particular, such solutions are unlikely to perform and offer the envisaged level of robustness and resilience. This is due to the inherent uncertainty such systems maintain, originating from the variability of information as well as from lack of knowledge. Moreover, excessive expertise on a domain when not managed correctly, may have adverse effects. A representative example is described in [6], where a game-theoretic approach revealed that in the case of many security experts being involved in protecting a network, the overall security is expected to drop. It is noteworthy that this result was derived following the assumption of limited information, which in fact in modern complex systems not only holds, but can be exacerbated due to misinformation.

In this paper it is argued that a self-healing capability in a complex system will need to be introduced along with the transformation of the incident handling functions that will not only monitor the target network assets, but the self-healing system itself. As self-healing is not solely about infrastructure monitoring but also (automated) response, traditional incident response based on cyber threat intelligence and network incidents is not appropriate, as the self-healing agents will be missing contextual information to take the appropriate decisions. In what follows, we briefly describe the main aspects of self-healing and propose the directed and adjusted involvement of a human operator, which should be performed following a decision-making framework. This is in order to allow the self-healing to process and gradually take over the management of the "straightforward" security incidents, but include the human in the loop for the complex and unrecognised incidents.

2 CONTEXT-SUPPORTED SELF-HEALING

Self-healing models are based on distinguishing three main phases: *detection*, *diagnosis* and *recovery*. Detection refers to the identification of suspicious activity. Diagnosis includes root cause analysis and associates this with a relevant self-healing policy. Recovery is the process of returning the system to its normal operational state, following paths and planned adaptations whilst respecting the constraints of the system [10]. Self-healing architectures are comprised of the following aspects:

- Network monitoring and intrusion detection. This refers to the anomaly or misuse detection techniques for identifying network attacks [3, 5]. There exists a considerable body of research and state-of-the-art leveraging and evaluating the performance of machine learning techniques in this domain, see for example [1].
- Root cause analysis. This refers to the activities during the diagnosis stage, aiming to identify the root causes that led to security control failures.
- Failover. Together with root cause analysis, failover enriches the diagnosis of the self-healing component in order to allow

PETRA '23, July 05–07, 2023, Corfu, Greece

ACM ISBN 979-8-4007-0069-9/23/07.
https://doi.org/10.1145/3594806.3594837

Table 1: Context-driven incident response

	Characteristics [11]	Self Healing	Human-in-the-loop	Example incidents (indicative)
Simple	Repeating patterns and consistent events Clear cause-and-effect relationships evident to everyone; right answer exists Known knowns	Rule based well-defined (P) (D) (T) (C) (E)	no engagement	known virus, anti-virus definitions exist known IoCs*
Complicated	Expert diagnosis required Cause-and-effect relationships discoverable but not immediately apparent to everyone; more than one right answer possible Known unknowns	"traditional" AI unsupervised (P) (D) (T) (C) (E)	(P) (I) (PI)	DDoS Spam
Complex	Flux and unpredictability No right answers; emergent instructive patterns Unknown unknowns Many competing ideas A need for creative and innovative approaches	advanced AI supervised (P) (D) (T)	(P) (T) (C) (I) (E) (PI)	Advanced Persistent Threat (Successful) ransomware
Chaotic	High turbulence No clear cause-and-effect relationships, so no point in looking for right answers Unknowables Many decisions to make and no time to think	AI-mediated (D) (T)	(P) (D) (T) (C) (I) (E) (PI)	Zero days Targeted supply chain attacks

*Indicator of Compromise. Incident response phase: (P) Preparation (D) Detection (T) Triage (C) Containment (I) Investigation (E) Eradication (PI) Post-Incident

the construction of an actionable and practical recovery path [12].

- Self-optimisation and automated recovery. This is where information from root cause analysis and the failover plan can be combined and a course of action is executed for automatically restoring the system. The course of action comprises system operational activities such as restore from backup, firewall and network re-configurations, access control updates and so forth. Automated recovery and self-healing in general is in line with the recent Software Defined Networking (SDN) paradigm [9].
- Adaptability. Although self-healing aims to offer effective and efficient responses to threats by improving the recovery window, the added value is the acquired knowledge from an incident which can feed back to the system and make it more resilient. From an incident response perspective, this aspect is related to the post-incident activity and is considered to be one of the most important part of incident handling [4].

Against the self-healing aspects presented above, we can identify two emerging themes that are pervasive across all aspects: sense making and decision making. While the type and level of abstraction of information may differ between these aspects, the success of the respective component is related to selecting the right response from a set of actions. This set may contain more than one correct and incorrect responses, which can be dictated by the *Domain*, as specified by the Cynefin framework below.

2.1 Cynefin for self-healing supported incident response

Cynefin is a framework for understanding complex systems through establishing the prevailing context which in turn will enable the identification of the best choice [11]. Cynefin has gained popularity in the leadership and (human oriented) management domain as it allows a team to decide on the best course of action based on the complexity and uncertainty of a given situation. Cynefin has five so-called dimensions or contextual definitions namely: *simple*, *complicated*, *complex*, *chaotic*, which are surrounding the *centre of*

confusion or *disorder*. Briefly, *simple* is when the problem is clear and well-defined and there is a clear solution; this is the area where best practices can be applied. It is the case of "known unknowns" and a fully automated recovery process can be applied, without even requiring any AI involvement. The *complicated* dimension is where the problem is characterised of "known unknowns", requiring expertise to establish the cause and effect. This dimension is where AI can perform at its best and offer added value.

The next dimension in line is *complex*, referring to the "unknown unknowns"; there are no clear right answers and cause and effect can be retrospectively established. As such, the system needs to probe and observe the results in order to determine the best course of action. We could argue that advanced AI such as Generative Adversarial Networks could operate at this point, with some human supervision. Moreover, a digital twin could offer a safe space to run *what if* scenarios and explore the best strategy forward for mitigating adverse actions. The fourth dimension is the *chaotic* contextual state, where cause and effect cannot be established even through probing and observation. At this point AI cannot operate and self-healing algorithms will most likely cause more problems than contributing to solutions.

The fifth dimension of *disorder* is simply a placeholder representing the situation where there is not enough information to take an informed decision on identifying which dimension we are in.

Table 1 summarises how the Cynefin framework can be used in self-healing supported incident response. In addition, given that a cyber attack is a dynamic event, we consider the time to be a key factor for progressing from one state or dimension to the other. As such, when an incident is initially detected, the risk averse approach would be to start in *disorder* and transition to another state depending on the reliability and amount of information gathered.

3 CONCLUDING REMARKS

We introduced an approach for managing the incorporation of self-healing in a structure manner and driven by uncertainty and context. Recognizing that there is no one-size-fits-all when deploying self-healing capabilities, we propose a mapping of the self-healing involvement on the different Cynefin dimensions. As future research, this mapping can be used to develop appropriate and more granular incident response playbooks to streamline the interplay between self-healing and human in the loop.

ACKNOWLEDGMENTS

Co-financed by the European Regional Development Fund of the European Union and Greek national funds through the Operational Program Competitiveness, Entrepreneurship and Innovation, under the call RESEARCH - CREATE - INNOVATE (project code: T2EDK-01469).

REFERENCES

[1] Mohammad Almseidin, Maen Alzubi, Szilveszter Kovacs, and Mouhammd Alkasassbeh. 2017. Evaluation of machine learning algorithms for intrusion detection system. In *2017 IEEE 15th International Symposium on Intelligent Systems and Informatics (SISY)*. IEEE, Subotica, Serbia, 000277–000282.

[2] Guy Amit, Asaf Shabtai, and Yuval Elovici. 2020. A self-healing mechanism for internet of things devices. *IEEE Security & Privacy* 19, 1 (2020), 44–53.

[3] Monowar H Bhuyan, Dhruba Kumar Bhattacharyya, and Jugal K Kalita. 2013. Network anomaly detection: methods, systems and tools. *Ieee communications surveys & tutorials* 16, 1 (2013), 303–336.

[4] Paul Cichonski, Tom Millar, Tim Grance, Karen Scarfone, et al. 2012. Computer security incident handling guide. *NIST Special Publication* 800, 61 (2012), 1–147.

[5] Ozgur Depren, Murat Topallar, Emin Anarim, and M Kemal Ciliz. 2005. An intelligent intrusion detection system (IDS) for anomaly and misuse detection in computer networks. *Expert systems with Applications* 29, 4 (2005), 713–722.

[6] Benjamin Johnson, Jens Grossklags, Nicolas Christin, and John Chuang. 2010. Are security experts useful? Bayesian Nash equilibria for network security games with limited information. In *Computer Security–ESORICS 2010: 15th European Symposium on Research in Computer Security*. Springer, Athens, Greece, 588–606.

[7] Linda Joseph and Rajeswari Mukesh. 2018. Detection of Malware Attacks on Virtual Machines for a Self-Heal Approach in Cloud Computing using VM Snapshots. *Journal of Communications Software and Systems* 14, 3 (2018), 249–257.

[8] Tao Ma, Shaukat Ali, and Tao Yue. 2021. Testing self-healing cyber-physical systems under uncertainty with reinforcement learning: an empirical study. *Empirical Software Engineering* 26 (2021), 1–54.

[9] Leonardo Ochoa-Aday, Cristina Cervelló-Pastor, and Adriana Fernández-Fernández. 2020. Self-healing and SDN: bridging the gap. *Digital Communications and Networks* 6, 3 (2020), 354–368.

[10] Harald Psaier and Schahram Dustdar. 2011. A survey on self-healing systems: approaches and systems. *Computing* 91 (2011), 43–73.

[11] David J Snowden and Mary E Boone. 2007. A leader's framework for decision making. *Harvard business review* 85, 11 (2007), 68.

[12] Alexander Stanik, Mareike Höger, and Odej Kao. 2013. Failover pattern with a self-healing mechanism for high availability cloud solutions. In *2013 International Conference on Cloud Computing and Big Data*. IEEE, Fuzhou, Fujian, China, 23–29.