A Pragmatic System-failure Assessment and Response Model

Jassim Happa[1], Graham Fairclough[2], Jason R. C. Nurse[1], Ioannis Agrafiotis[1],
Michael Goldsmith[1] and Sadie Creese[1]

[1]*Department of Computer Science, University of Oxford, Oxford, U.K.*
[2]*Department of Politics and International Relations, University of Oxford, Oxford, U.K.*

Keywords: Situational Awareness, Emerging Attack Methods, Triage, Future Threats, Cyber Resilience.

Abstract: Several attack models exist today that attempt to describe cyber-attacks to varying degrees of granularity. Fast and effective decision-making during cyber-attacks is often vital, especially during incidents in which reputation, finance and physical damage can have a crippling effect on people and organisations. Such attacks can render an organisation paralysed, and it may cease to function, we refer to such an incident as a "System Failure". In this paper we propose a novel conceptual model to help analysts make pragmatic decisions during a System Failure. Our model distils the essence of attacks and provides an easy-to-remember framework intended to help analysts ask relevant questions at the right time, irrespective of what data is available to them. Using abstraction-based reasoning our model allows enterprises to achieve "some" situational awareness during a System Failure, but more importantly, enable them to act upon their understanding and to justify their decisions. Abstraction drives the reasoning process making the approach relevant today and in the future, unlike several existing models that become deprecated over time (as attacks evolve). In the future, it will be necessary to trial the model in exercises to assess its value.

1 INTRODUCTION

Quick decision-making is vital during incidents in which reputation, finance and physical damage can have a crippling effect on people and organisations. Such attacks can render an organisation paralyzed, and it may cease to function. A recent example would be the Sony "The Interview" attack that rendered the operational aspects of Sony inoperable. We refer to such an incident as a "System Failure" in which hardware or software faults, misconfigurations, or the intentional work of malicious actors are the reason behind the system no longer functioning.

In this paper we propose a novel conceptual model to help analysts make pragmatic decisions during a System Failure. Our model distils the essence of attacks and provides an easy-to-remember framework intended to help analysts ask relevant questions at the right time, and adopt to the data that is immediately available – allowing our model to be as relevant today and in the future, unlike several existing models that become deprecated as attacks increase in complexity. Our approach expresses different levels of granularity on an ad-hoc basis, and complements existing models as opposed to competing with them.

Our model is derived from a number of existing models on the topic of attack and response. Abstraction drives the reasoning process through a series of "*Aspects*" (including "*Impact*", "*Vector*", "*Motives*" and "*Attribution*") and "*Nuances*" (individual properties) of attacks, enabling the model to be inclusive about what is needed to be considered during the incident response decision-making processes. Our model assumes analysts have little time to explore all theoretical considerations and have to make the most achievable decisions possible with scarce data (about the attack) and resources. Due to abstraction, our model may not yield the best decisions in all circumstances, but it is likely to lead to more pragmatic decisions, and more importantly: actions that can be justified after the attack. Our method has the advantage of:

- **Attempting to "Know the Unknowns":** helping map out which pieces of the attack puzzle may be missing.
- **Decision Documentation** of the incident response reasons more straightforwardly.
- **Prioritising Actions** during a System Failure. In our paper we present a use case example in which our model could have been used and discuss its advantages and limitations.

Happa, J., Fairclough, G., Nurse, J., Agrafiotis, I., Goldsmith, M. and Creese, S.
A Pragmatic System-failure Assessment and Response Model.
DOI: 10.5220/0005795105030508
In *Proceedings of the 2nd International Conference on Information Systems Security and Privacy (ICISSP 2016)*, pages 503-508
ISBN: 978-989-758-167-0

2 REFLECTING ON ATTACK AND RESPONSE MODELS

Many attack models and classification schemes tend to describe cyber-attacks in one of two ways; either as hierarchical structures or as linear processes. Hierarchical structures (e.g., attack trees) have the advantage of describing attacks in terms of their different properties, but often neglect the temporal component e.g. AVOIDIT (Simmons et al., 1997), CAPEC (MITRE, 2015), VERIS (VERIS, 2015), NIST (NIST, 2015), SANS (SANS, 2015). Linear processes capture the temporal element since they assume that actions happen sequentially (Howard and Longstaff, 1998; Hutchins et al., 2011), however, may fail to describe lateral movement or cases where attacks occur in parallel.

Many prior works attempt to outline attacks comprehensively or provide explanations of the direst consequences when an attack succeeds. In addition, they describe ideal solutions, see for instance several of MITRE's efforts (2015), FIRST's efforts (2015) and VERIS efforts (2015). While these efforts show substantial progress in tackling cyber-attacks, they may not be feasible for all circumstances, particularly when decisions have to be made with limited resources (regarding information available and time constraints, e.g. during an electric blackout), technical and operational common sense has to prevail when making decisions and incident responses quickly.

To the best of our knowledge, no truly *pragmatic* approach to facilitate understanding of attacks and to provide a framework to ensure technical and operational sanity exists. It is worth noting here that we do not consider practical in terms of convenience, but in terms of necessity and efficiency (due to limited resources). No model uses easy-to-grasp reasoning to aid understanding and response to cyber-attacks that is able to abstract the technical details of an attack and simply consider its properties. Other models that we have considered but are not included above due to space limitations include (Bishop, 1995); Lough, 2001); (Ten et al., 2010), but were still considered in our model.

2.1 Commonalities Across Models

From the models we have reviewed, we were able to identify a number of noteworthy differences and common factors. For instance, at the core of each of the attack models, they detail the specific activities leading to the compromise of some security feature (whether it be confidentiality, integrity or availability) of an asset. While some (e.g., the Killchain) place more emphasis on the types of attack steps and characterising what goal each step is seeking to reach, others (such as VERIS (2015)) adopt more general steps and focus on the wider problem. In terms of attack modelling, possibly the most representative model is that of Howard's taxonomy to specify incidents. It captures several of the actions within an incident but also sheds light on the reason for an attack (e.g., for financial gain, to cause system damage, or for political gain).

While attack models allow for a detailed analysis of an attack, incident response models consider what attack has been launched, but especially how to appropriately respond to it. In the NIST model above (NIST, 2015) for instance, we see a requirement to detect an attack, but a majority of the life cycle is on responding to it. Some of the key questions in incident response target why and how an attack occurred, and who caused it. Almost identical questions can be found in the SANS model and process flow for incidents.

Across the more attack-focused models and those more geared to incident response, there are notable commonalities. To start, there is an aim to understand incidents and clearly define what has been impacted and the activities that have led to a breach of an asset's security. Key questions on motivation may also inform the choices of actions after attacks.

Our approach shares commonalities with business continuity/cyber resilience models (for an overview, see (Gibson and Tarrant, 2010) and (Caralli et al., 2010), with the key distinction being that our efforts are mainly attack focused and intended to be used by Security Operations Centres (SOCs) and Computer Emergency Response Teams (CERTs).

3 A PRAGMATIC SYSTEM-FAILURE ASSESSMENT AND RESPONSE MODEL (SAM)

Our System-failure Assessment and response Model (SAM) is a directed human-reasoning approach to incident handling that uses abstraction as part of the reasoning process. The decision-making process that the model promotes is based on deduction and experience.

A series of high-level observables from very basic questions are able to provide first-pass indicators of how to respond. For instance, in the case of attempting to identify impact of an attack, and understand what needs to be fixed immediately. An overview is shown in Figure 1.

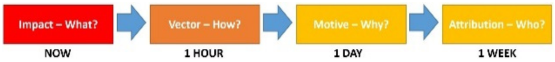

Figure 1: SAM outlines what questions to consider first, as well as an example timeframe we suspect they have to be addressed. Note: the timeframe will be relative to each attack.

Our method is based on identifying common factors between other attack and incident models led to the creation of a model that examines cyber-attacks in a pragmatic manner. In particular, by asking the question of "what matters the most when disastrous attacks occur?" It can be considered as a series of aspects to ask questions about to help identify what the best incident response strategy might be. At its core, the structure of our model follows directed interrogative pronouns in order of importance: *What* is being attacked, what's affected, how is it affected? (Here known as Impact); *How* is it being attacked including when and for how long (namely, attack Vectors); *Why* is it being attacked (or Motive); and, *Who* conducted the attack (i.e. Attribution)? These are listed as what we believe to be their importance during attacks. This approach uses the common factors identified in prior attack-modelling literature outlined in Section 2.

By Impact we mean the consequence of an attack, one that is achieved through a Vector. Vectors describe the means to achieve a Motive. Several attack vectors exist in the cyber domain: some are technically driven (e.g., exploits such as buffer-overflow attacks, code injection, or use Trojans, viruses, worms, etc.), others social-engineering (e.g. phishing attacks). Motives describes the intents of an attacker. Attribution should describe who is responsible for the attack.

The first question (i.e., the "*What*") fills in the remainder of the other questions, and is crucial to be able to answer first. If we do not know what has been attacked, it is difficult to consider anything else. While the consequences of an attack however, may not be immediately clear, identifying what has been attacked should be the first priority, as it is only after this that we may be able to determine the level of damage made. Similar to triage, an estimation of the damage must be made to assess what can be done afterwards. It is worth noting that time criticality will depend on the incident, and the timeframe shown in Figure 1 should be considered relative. Also, should another incident happen before reaching the end we likely have to start again at the Impact aspect, or re-review whether the initial assessment was correct.

The model attempts to guide analysts and decision makers alike by providing a framework of key general questions about attacks. These questions abstract out the technological component but provide the basis for which technical aspects can be applied. By asking basic, easy-to-grasp questions, our framework can be used to communicate incident-response decisions quickly to non-technical audiences as well, such as business managers, lawyers or policy makers, some of whom will be responsible of organisation operations. What exactly these questions are, can be driven by existing standards and models, however, the choice in standards available needs to be well-understood by everyone on the team before any major incident.

Naturally, precision is lost in abstraction. However, it also means that analysts can choose a classification scheme or detailed attack model they believe is most appropriate to describe the attack. In practice this means concepts such as CIA or models like Killchain can be used to describe nuances of the Impact, while CAPEC and Common Vulnerability Scoring System (CVSS) can be used to describe details (Nuances) about the Vector, see Figure 2.

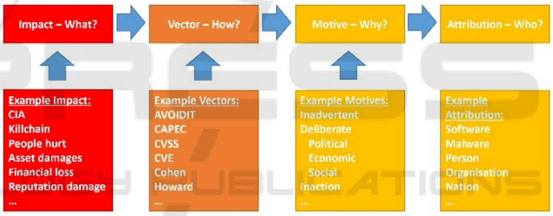

Figure 2: SAM's use of abstraction, the uses of more detailed models or inputs can feed into the main components of SAM.

4 APPLYING THE MODEL

SAM is an abstraction-based approach to understand attacks to achieve "*some*" understanding, quickly, to help act upon them. The model distils all components of any attack to their bare minimum, and allows decision makers to insert details – using their standard of choice (whether of necessity or preference) – where appropriate. Examples could include assigning costs to the Impact of attacks, identifying what machines have been affected, whether people have been hurt by the attack (and to what extent), or CIA to describe the area of the attack (as part of Impact, e.g. "was integrity affected of any components"?). CAPEC could be used to describe the attack Vector. Using SAM, we care about understanding as many aspects about the attack with as limited resources as possible. This means that details may not matter (or even be available).

4.1 As a Mnemonic

Our model represents a first stage in constructing a mnemonic for dealing with a cyber incidents, one we expect will be refined in the future. In First Aid there is A (Assess and Airways), B (Bleeding and Breaks), C (Circulation and Consciousness) and D (Deal). In dealing with a physical security incident there are the 5Cs: Confirm, Clear, Cordon, Control and Close. However, unlike the health sector, in cyber systems there are other considerations to take into account; some of these are mentioned below.

Triage assumes that a person is being treated, and determines priority of who to treat first. Despite a system is being healed, it does not mean is has been healed permanently. Healing can happen much faster for cyber systems, but attacks can also persists after initial incident. Whilst formulating the aspects, the order of them raises considerable debate. As an analogy of our model to the real world, we can express these aspects in the form of responding to a hostile situation; such as a person who has just been harmed, in which we might say:

Impact: Identify what has happened, then "*stop the bleeding*" and prevent further immediate damage to happen.

Vectors: Identify what the attack weapon was and prevent the attack weapon from being used again.

Motives: Identify the likely reason why the person was attacked, and attempt to demotivate the attacker to want to find another weapon if the attack was with intent.

Attribution: Identify who the attacker is and disable the attacker completely.

If the attacker is malicious, one might say that preventing the attack weapon from being used again is more important than stopping the bleeding and more harm from happening. Having said this, in the cyber domain, we argue Impact should be addressed first in most situations because the rate of which Impact can happen. In most cases in the cyber domain, we may be able to cut of connection or isolate the damage relatively quickly (e.g. by disconnecting the monitored system from the Internet or local network), and it should be done first to limit the damage. Then, analysts can investigate the attack vector, whether it is outward facing or an insider attack, it will now be considered an outward facing component to a device or a system. Similarly, the Motive and Attribution aspects may be swapped – it may not be possible to understand motives of an attacker before one understands who the attacker is, however, in our use cases, our ability to identify motives was far greater than the ability to identify who the attacker was. The exact order of aspects, or means to validate the order is subject to further research.

4.2 In Operational Environments

SOC-like environments are often overwhelmed by network and intrusion-detection system alerts every day, and strive to understand the threat landscape. There is an important distinction we make between this model and cyber-incident handling more generally. This model is intended to be used during System Failures only, i.e. attacks that have crippled the monitored system to a point in which even cyber observables are limited. Our model in an operational environment follows a feedback loop, as shown in Figure 3.

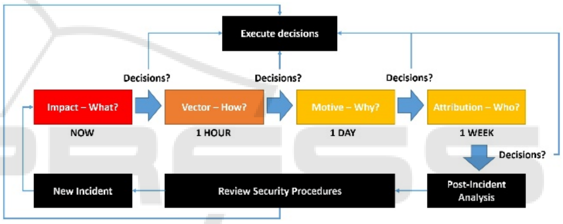

Figure 3: SAM in an Operational Environment.

Once a New Incident has happened, an assessment of the Impact is made by asking the appropriate questions pertaining to what has been observed. Decisions are made and executed based on observable evidence and common sense reasoning related to the Impact of the system.

After decisions have been made related to the Impact, we progress in a similar fashion and reasoning structure until decisions have been made (and executed) related to Attribution. After an incident, Post-Incident Analysis can be conducted (if any are in place), and existing procedures can be reviewed and implemented, until a New Incident comes in again, and the loop is repeated. Each answer gained from the model should allow elucidation of how best to respond to the attack at that juncture. There are particularly difficult corner cases such as reputational damage (i.e. reputational damage is difficult to measure the long-term impact of), which we are contemplating for future work.

5 DISCUSSION

Our model is a first version of an approach that is able to describe attack behaviour in terms of four constituent components by distilling observables to their essence and making decisions from this understanding. The proposition of "first what, then how, then why, then who" as a pipeline is not a controversial one. Indeed, we argue that most security experts naturally respond in the order presented in SAM anyway. However, with this pipeline happening across the board informally, and no one having identified it as such is an observation we believe is worth sharing to enable analysts to more effectively communicate through a shared incident-response mental pipeline model. This helps communication during incidents, but also helps for planning red teaming exercises (through storyboarding using our pipeline), but also allows analysts to identify when nuances of attacks distinguish themselves from past incidents. This is helpful in the immediate future as we're able to classify those nuances into different domains describing the attack.

5.1 Strategic Role of SAM

Our model serves as first-pass mechanism to respond to incidents quickly, by reviewing key, generic aspects of an attack and being able to ask the right questions about the attack at the right time, regardless of type of attack; whether it be a zero-day exploit or an APT.

The main purpose of SAM is to support the immediate response to a cyber event the model also fulfils important strategic functions. Time is of equal importance to the strategic decision maker as it is to the tactical cyber analysis trying to deliver a solution. Both are involved in the damage limitation process. At this point, focus is not on the detail; it is on the message that something is wrong. This message is not just for those at the coalface. Other actors including system users, senior decision makers and externally those who may face a similar threat need to be made aware. This is a task for which SAM is well suited. Its format provides the means to alert internal and external audiences. Providing adequate insight to activate the necessary emergence response within the challenging time constraints and without overwhelming recipients with superfluous detail that is arguably not important at that juncture.

SAM's dynamic construct allows details to be obtained, in a structured manner over time. This accumulation of facts is critical in enabling informed decision-making. As time progresses the immediacy of dealing with the reality of the attack transfers to dealing with its consequences. These consequences are increasingly both internal and external. Internally, these might include the extent to which the system can continue to operate or the level of damage that has been done. Externally, consequences might involve mitigation to protect reputation or the need to inform partners of a potential threat in order to safe guard their interests. SAM's ability to establish this structured feed of detail, whilst maintain simplicity and the spread of erroneous information is a key contributor in ensuring appropriate decision-making. Of great importance will be which recovery and resilience measures should be activated.

5.2 Red Teaming

Helping cyber-security analyst defend their systems better during particularly devastating attacks. SOCs and CERTs are some of the few intended audiences we have in mind for the model as it currently stands. We believe, however, it is possible to use the model to also design Red Team activities (i.e. ethical hacking to improve the system). Applications of this might help pen testers storyboard attack scenarios in a structured, reproducible manner. This may perhaps be best done by starting the model in the reverse order: beginning with who is attacking, moving on to why they are attacking, to how the attack is implemented (as a means to achieve the why the attack is taking place), and finally outline the intended impact. As an activity, it can:

- Provide insight into the threat and hence inform decisions over the level of risk;
- Act as a mechanism to exercise and assess an organisation's emergency response planning,
- Be used to generate options not previously considered in times of disaster.

5.3 Scoring Systems and Nuances

Conceivably, the model could be implemented similarly to vulnerability scoring systems such as CVSS (FIRST, 2015). By having a form that asks a series of straightforward questions relating to the four main aspects it may be possible to extrapolate direct relationships between what has been observed, and possible remedies (first-pass indicators about what to do next). In the case of attempting to identify "Impact", and what needs fixing "now", questions about "Nuances" might help the decision process. Each of these Nuances should help answer: "What actions can we do to stop or limit the (aspect)?" Conceivably, each of these aspects could be formalised and better defined to better process them in automated systems or deliver precise metrics that help describe the attacks.

Examples of nuances might be asking questions about the **Measurability** of the attack Impact (i.e. what has been observed as affected? (listing people, physical items, hardware, software, finance, reputation etc.)), its **Influence** (i.e. how is the impact affecting the system? (e.g., listing CIA per affected Impact)), the **Duration** of the Impact or **Transparency** of the Impact (i.e., how visible the Impact is). In the case of Vector, we might be interested in exploring Nuances such as **Implementation** (i.e., "what technologies (hardware and software) and protocols were involved to implement the attack?"), **Deployment** (i.e., "how was the attack likely deployed?" For example, network vulnerability exploit or similar.), **Distribution** (i.e., how distributed is the attack? For example, what physical location in the real-world are affected?), or **Repeatability** (i.e. how repeatable is the attack?). From the **Patterns** in the Vector or by identifying the intended **Target**, it may be possible to obtain Motives, and finally, we may obtain (likely) Attribution through an analysis of all of the aforementioned Aspects and Nuances. Future work will explore exactly which Nuances to consider, and how the model could be implemented in a tool.

6 CONCLUSIONS

In this paper we provided a way to organise thinking and prioritise effort when dealing with system failures. SAM is intended to be a model through which technical and non-technical decision makers can easily communicate and make better decisions collaboratively during major incidents in which time is of the essence and information is lacking significantly.

The model enforces common-sense reasoning, enables justifiable decision making that are based on empirical evidence where available. Future assessment is necessary to say with confidence whether our model effectively achieves its aims. We intend to conduct studies with SOC analysts under a variety of different use case scenarios.

REFERENCES

Bishop, M. (1995). A taxonomy of unix system and network vulnerabilities. Technical report, Technical Report CSE-95-10. Department of Computer Science, University of California at Davis.

Caralli, R.A., Allen, J and White, D. W., (2010). CERT Resilience Management Model: A Maturity Model for Managing Operational Resilience. Addison-Wesley Professional.

Cohen, F. (1997). Information system defences: a preliminary classification scheme. Computers & Security, 16(2):94–114.

FIRST. (2005). Common Vulnerability Scoring System. http://www.first.org/cvss.

FIRST. (2015). FIRST Security Library. https://www.first.org/library.

Gibson, C.A. and Tarrant, M. (2010). A 'conceptual models' approach to organisational resilience.

Howard, J. D. and Longstaff, T. A. (1998). A common language for computer security incidents. Sandia Report: SAND98-8667, Sandia National Laboratories, http://www.cert.org/research/taxonomy 988667.pdf.

Hutchins, E. M., Cloppert, M. J., and Amin, R. M. (2011). Intelligence-driven computer network defense informed by analysis of adversary campaigns and intrusion kill chains. Leading Issues in Information Warfare & Security Research, 1:80.

Lough, D. L. (2001). A taxonomy of computer attacks with applications to wireless networks. PhD thesis.

MITRE. (2015) Common Attack Pattern Enumeration and Classification. http://capec.mitre.org/MITRE 2015. Computer Security. http://www.mitre.org/publicationkeywords/computer-security.

NIST. (2012). Computer security incident handling guide. Technical report.

SANS (2004). Global information assurance certification paper. Technical report.

Simmons, C., Shiva, S., Dasgupta, D., and Wu, Q. (2009).
AVOIDIT: A cyber attack taxonomy. University of Memphis, Technical Report CS-09-003.

Ten, C.W., Manimaran, G., and Liu, C.-C. (2010). Cybersecurity for critical infrastructures: attack and defense modeling. Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on, 40(4):853–865.

VERIS. (2015). Vocabulary for Event Recording and Incident Sharing. http://veriscommunity.net/