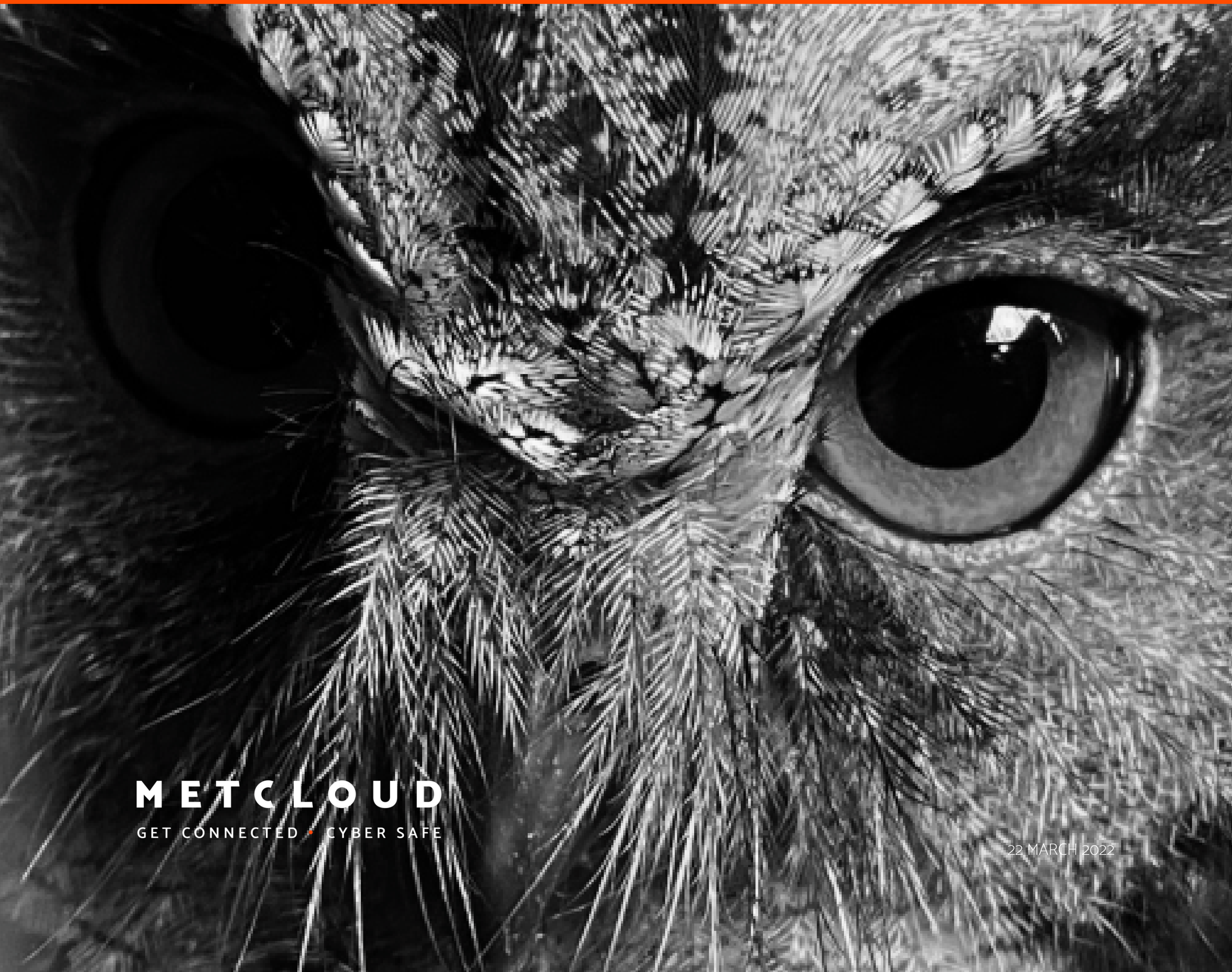
WHITE PAPER – ARTIFICIAL INTELLIGENCE IN THE SECURITY OPERATIONS CENTRE

Artificial Intelligence in the Security Operations Centre



METCLOUD

GET CONNECTED • CYBER SAFE

22 MARCH 2022

A PERFECT STORM

2021 has been a challenging year for those responsible for cybersecurity in public and private sector organisations. Ransomware and software supply chain attacks have been constantly in the headlines, claiming some high-profile scalps, including Colonial Pipeline and the many users of SolarWinds' Orion monitoring platform. Trends such as the Internet of Things (IoT), the integration of operational technology (OT) with IT systems, and the migration to cloud-native architectures are continuing apace, requiring new security approaches and technology. On top of everything else, the COVID-19 pandemic has disrupted working patterns and led to hasty introduction of support for large-scale remote and flexible working, as well as being the focus of phishing campaigns.

Organisations recognise the need for increased security investment to introduce pro-active security practices and new hardware and software, and to implement employee awareness campaigns[6]. However, one of the factors hampering progress is the severe shortage of experienced cybersecurity professionals. A recent study[7] conducted on behalf of the DCMS estimated that 50% of UK businesses lack the ability even to implement the basic cyber-hygiene measures laid down by the Cyber Essentials scheme[8]. Increasingly, organisations are outsourcing their security functions, but this merely displaces the skills gap to managed security service providers. It is reported that 47% of cybersecurity firms face problems due to lack of technical skills in existing staff or job applicants. A related report[9] provides evidence that in 2021 there was an annual shortfall of about 10,000 in the UK cybersecurity employment pool and suggests that if remedial action is not taken, the situation will only get worse. A contributing factor is the rate at which experience staff are leaving the profession. According to a global survey of cybersecurity professionals[10], consequences of the skill shortage include increasing workload (62%), unfilled posts (38%) and a high burnout rate (38%).

METCLOUD and Birmingham City University (BCU) are working together on an applied research project that addresses this problem in the context of the Security Operations Centre (SOC). Its aim is to use Artificial Intelligence (AI) and Data Science (DS) techniques to reduce the workload and amplify the skills of SOC analysts, resulting in more efficient and effective detection of, response to, and remediation of cyber-attacks. In the process it will increase job satisfaction and staff retention, and enable a more scalable business model.

[6] According to the 2021 IDG Security Priorities Study, 90% of security leaders believe they are falling short in addressing cyber-risk. Small and medium businesses plan to double their security budgets to an average of $11 million over the next year, while average enterprise budgets will increase to $123 million.

[7] Cyber security skills in the UK labour market 2021, Ipsos MORI for the Department for Digital, Culture, Media and Sport (DCMS)

[8] https://www.ncsc.gov.uk/cyberessentials/overview

[9] Understanding the Cyber Security Recruitment Pool, Ipsos MORI for the DCMS

[10] The Life and Times of Cybersecurity Professionals 2021, Enterprise Strategy Group (ESG) and Information Systems Security Association (ISSA)

LIFE IN A SECURITY OPERATIONS CENTRE

A SOC is the organisational function that is responsible for active defence against cyber-attack. SOC activities include preventing, detecting, monitoring, and responding to cyber-threats. Because of the specialist skills and infrastructure required, more and more organisations are outsourcing this function to third parties providing a 'SOC as a Service' (SOCaaS).

SOCs operate in three main concurrent modes:

1. Responding to alerts generated by security monitoring infrastructure elements, that might be indicative of a future, attempted or on-going attack.

2. Responding to reports of problems by other parts of the organisation that might be consequences of a cyber-attack.

3. Threat hunting: proactively looking for evidence of attacks that have not been detected by the first two means. This may be driven, for example, by threat intelligence reports of the widespread, stealthy exploitation of some new vulnerability. A 'threat hunt' could be initiated to investigate whether the organisation had already become a victim.

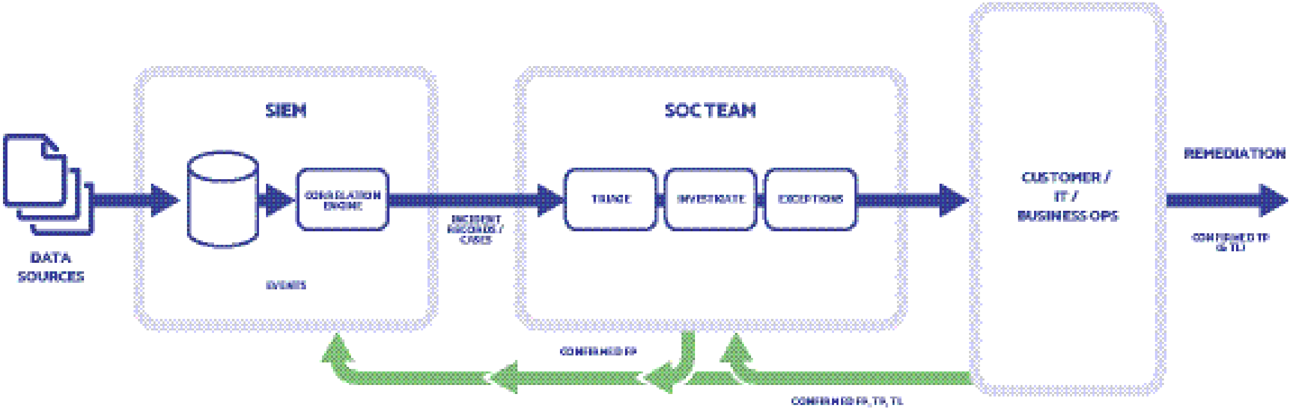


Figure 1: Recognising and classifying incidents

This report will mainly focus on the first mode. Assuming that a Security Information and Event Management system (SIEM) is used to store and process security event data collected from heterogeneous sources, the associated process can be broken down into the steps outline below and in Figure 1. The process applies to both in-house SOC and SoCaaS cases. In the latter, the SOC analysts will be employees of the service provider and the IT and business operations personnel will belong to the customer organisation. Note that terminology is not standardised and may vary depending on the SIEM used and local practice.

The SIEM looks for combinations of recent event records that match one of its correlation rules. Depending on the SIEM, this could be done by scanning the stream of new events using a Complex Event Processing (CEP) engine, or else by continuously searching recent additions to the SIEM database.

When a match occurs, a correlation event record is generated, containing references to the matched events, and potentially other similar events occurring within the same time window.

An incident record is created containing references to one or more similar correlation events and submitted to a Case Management System (CMS), which assigns the incident to a particular analyst for investigation. A human-readable explanation and remediation advice may be added to the incident record.

The assigned analyst works with the rest of the SOC team and with IT and business operations personnel to confirm and characterise the incident and, if appropriate, initiate remedial action.

Let us look at step 4 in more detail. A major goal during this stage is to assign the case to one of a small number of categories indicating whether further action is required. For the purposes of this report we adopt the classification scheme used in the 2022 Orange Cyberdefense Security Navigator report[6], in which the categories are: True Positive (TP, the incident is confirmed as compromising security), True Legitimate (TL, the incident happened, but was a result of legitimate activity), and False Positive (FP, the SIEM misinterpreted the events).

First, the assigned analyst examines the incident record to see whether it can be dismissed immediately as FP. If so, the case is closed; if not, an investigation is carried out. This will often require consultation of external resources such as threat intelligence databases or web sites. The resources consulted will depend on the nature of the incident. If the analyst is fortunate, he/she will have a 'play book' to guide the investigation, but often he/she will have to rely on experience, intuition, and advice from more senior colleagues. If the investigation concludes that security has not been compromised, then the incident is marked as FP and the case closed. Otherwise, the analyst must consider informing IT and business operations personnel. Reasons why this might not be done include the following:

- The incident has already been reported. This can happen if a correlation rule fires multiple times in response to the same underlying incident.
- The IT and business operations personnel have previously notified the analyst of some legitimate activity that would explain the incident, for example, scheduled maintenance.

[6] Available from: https://orangecyberdefense.com/global/security-navigator/

In the absence of such reasons, the analyst informs the IT and business operations personnel of the incident. If the analyst is confident that the incident is malicious, it is classified as TP, first. Otherwise, the IT and business operations personnel conduct their own investigations, taking into account information provided by the analyst leading to classification as TP, FP or TL as appropriate. This conclusion is then passed back to the SOC. Orange report that in the first ten months of 2021, 36% of incidents were confirmed as TP, 21% as TL, 40% as FP, and 3% remained inconclusive. The TPs were categorised as follows: Malware (38%), Network and Application Anomalies (22%), Account Anomalies (13%), System Anomalies (9%), Policy Violations (8%) and Social Engineering (6%). These can largely be deduced from the incident record received from the SIEM.

Once the TP status of an incident has been confirmed, response and remediation can begin, informed by playbooks and information gathered during the investigations. However, it is also important to review lessons learned, and update the knowledge base used in each step if appropriate. This includes updating correlation rules, investigation playbook, contextual and state information, and sharing knowledge among analysts. For example, suppose an incident is confirmed FP by the business operations personnel/customer for well-identified reasons. The correlation rule that generated the incident might be refined to avoid this happening again in the same situation. Similarly, the investigation playbook could be amended to advise the analysts to watch out for analogous situations.

The process describe above is both time-consuming and labour intensive. Time is precious in cybersecurity operations. The sooner a TP incident can be confirmed, the sooner remediation can begin, and the less time an attacker has to exploit a breach or vulnerability. Similarly, the earlier a FP or TL can be confirmed, the more time an analyst has to devote to investigating and remediating TPs. SOAR (Security Orchestration, Automation and Response) tools have gained popularity recently, promising to help standardise and automate intelligence gathering and response and remediation workflows. However, it is only possible to automate well-defined processes, and most of the low-hanging fruit has already been plucked. Much of the analyst's work is investigative, with subsequent actions guided by the results of previous ones and involves consultation of a variety of external information sources. Furthermore, they are often not good at introspection and articulating their thought processes. Consequently, further automation is not straightforward. A complementary approach is to provide the analyst with interactive 'power tools' that make investigations more efficient by placing information at his/her fingertips in a form that is easy to assimilate. A combination of intelligent automation and pro-active and interactive assistance may hold the key to the future of the SOC.

CAN ARTIFICIAL INTELLIGENCE SAVE THE DAY?

As an applied discipline, Artificial Intelligence (AI) is about taking inspiration from the individual and collective capabilities of humans and other living creatures and theories of how these capabilities arise, to help us create useful computational systems. AI has a history as long as that of computing, and in fact their origins are closely intertwined. It has gone through many 'hype cycles', oscillating between eras of inflated expectation and disillusion. It is a very broad discipline with many branches, which also have gone in and out of favour over the years. Nevertheless, the progress in AI and its sub-fields over the years, and the resulting practical benefits are many and real. AI is fated to be under-appreciated as once understands how to implement an intelligent capability, it no longer appears intelligent. In recent years, great advances in the branch of AI known as Machine Learning (ML), enabled by the availability of large amounts of data and cheap and plentiful storage and processing capacity. ML is now going mainstream and being applied in many application domains including cybersecurity. However, other branches of AI research also remain active and relevant to real-world problems. Furthermore, current ML techniques, while extremely useful, are still a long way short of reproducing the ability of humans to learn new knowledge continuously over the course of a lifetime and apply it intelligently.

In normal software development, a programmer designs and implements an algorithm that when presented with some input data, the correct results are output. ML comes into its own in problems where the algorithm is unknown, and so must be discovered. ML models are basically complicated mathematical functions with very many free parameters. The models are constructed so that a wide range of concrete functions can be approximated if parameter values are chosen appropriately. When ML practitioners talk about training the model, they mean applying procedures that adjust the parameter values incrementally to improve the quality of the function according to some 'objective' measure. Often, the objective measure assesses the trained function's ability to replicate/predict the results of some natural, technical, business, or cognitive process.

There are two main active branches of ML: statistical and deep learning (DL). DL models are inspired by studies of the human/animal brain. The brain is made up of huge numbers of basic structures called neurons, that are highly interconnected. On short timescales, it learns by adjusting the strengths of the connections, and on longer timescales by growing new connections. DL models are also known as deep neural networks (DNN). Artificial neural networks (ANN) are constructed by connecting together computational neurons, loosely based on simplified versions of those found in the brain. They derive their power from the way their neurons are arranged and connected, and are trained by adjusting weights that express the strength of the connections. The availability of cheap computing and the ability to collect and store large amounts of data, together with progress in learning algorithms, has led to the development of much larger ANNs made up of many layers, hence DNNs. These have been successfully applied to a wide variety of problems in recent years with considerable success. Broadly speaking, statistical ML models work well when datasets are not too large. DL requires more data to train, but also has the capacity to continue to improve performance as the size of the training dataset increases.

ML techniques have already been applied to a number of cybersecurity functions including malware detection, network intrusion detection and phishing email detection. Often the ML is being used to complement or replace signature-based approaches to detection, whereby the security software is supplied with so-called Indicators of Compromise (IoCs). These are patterns or measurements that are characteristic of particular malicious activity. For example, signature-based malware detection systems often work by applying a hash function to an executable file found on disk or attached to an email and comparing the results against a database of hashes of known malware executables. Signature-based approaches have a number of disadvantages, including the following:

- Cyber-criminals are an inventive bunch who are always coming up with new attacks. Consequently, the IoC database needs to be updated continually.
- Attackers are aware of the detection techniques in use and develop ways of circumventing them. For example, they apply random modifications to malware that do not affect its function but mean that the hash function will produce a different result. Consequently, attackers and defenders are in a never-ending arms race, with defenders usually having to respond to the attackers' innovations.
- A new attack needs to be observed, reported, and confirmed before the appropriate IoC can be added to the signature database. During this time, the attacker has a window of opportunity to exploit the technique undetected.

One of the motivations in applying ML to cybersecurity is avoid these disadvantages.

One ML approach that has been applied with some success in a security operations context is anomaly detection (AD). An anomaly can be defined as an event/incident that is inconsistent with the assumed model of a process. The AD algorithm is trained by showing it many examples of 'normal' system behaviour. In a cybersecurity context, 'normal' typically means in the absence of threat activity, although this may be hard to guarantee. In operation, the detection algorithm is presented with observations of the real process and compares them with the predictions of learned process model, flagging inconsistencies as anomalies. So-called User and Entity Behaviour Analytics (UEBA) tools are a good example of this class of approach and are now incorporated in many SIEMs. As the name suggests, these monitor the behaviour of users, servers, endpoints, routers etc. and report unusual activity, for example if a user who normally works standard office hours from her desk at company HQ, suddenly starts logging in remotely in the early hours of the morning. She could simply be catching up on a backlog of work, but she could also be downloading client lists to take with her to a new job, or else a threat agent could be using her stolen credentials.

So, UEBA and more generally, AD security techniques are not dependent on signatures, but rather on the (reasonable) assumption that malicious activity will reveal itself in changes to the behaviour of the system being monitored. They should be able to detect novel attacks, not be fooled by variations in existing attacks, and not be subject to the time lag between first use of an attack and the ability to detect it. They are complementary to existing signature/rule-based techniques, each detecting attacks that the other may miss. UEBA tools are best seen as additional data sources feeding into the SIEM. More generally, AD also could be used to provide a non-rule-based component to the SIEM's correlation engine. They do not really assist the SOC analyst with the task of investigating and classifying incident reports, in fact they may well make matters worse! In cybersecurity applications of AD, unusual activity is identified as potentially malicious. This means high FP rates can result from atypical examples of legitimate behaviour that are not included in the data set used for training.

AD methods are an example of unsupervised learning, in contrast with supervised learning methods that are trained using data that has been labelled with the correct answer. As an example of supervised learning, consider an application that is intended to detect malware in email attachments. To create a training dataset, sets of attributes are extracted from examples of different types of known malware and legitimate attachments, and labelled “malware” and “benign” accordingly. The learning algorithm tries to find criteria that would ‘explain’ why some attribute sets had been given one label, and some the other. These criteria are then applied during the operational phase to classify the attachments. Another type of unsupervised learning technique is clustering. Here, during training, the algorithm tries to discover natural groupings in the data without being told in advance what those groupings are. Then, in operation, it assigns new data to the cluster whose members it appears most like. Such an approach could be applied to the attachment classification problem, but the training algorithm would have to work out for itself that there are two (or possibly more) types of attachment, and a human would need to determine the significance of the types.

PRACTICAL CHALLENGES

Returning to the problem in hand, recall that we want to use AI to significantly improve the efficiency and effectiveness of the SOC team in performing the role indicated in Figure 1 and described in the text below it. At its simplest, the task is to take an incident report output by the SIEM an classify it as TP, FP or TL (assuming the Orange scheme is used). As there is ample historical data available that records how previous incident reports were dealt with, a supervised learning-based classifier seems to be called for. However, there are complications:

- The information on the disposition of previous cases is in the form of free text annotations added to the record by the assigned analyst. Labelling, therefore, must be performed manually or by a natural language processing (NLP) algorithm.

- There are many different types of incident report. Training a separate ML classifier for each type may not be practical. On the other hand, a single classifier covering all incident types would need to be extremely complex.

- Each incident type corresponds to a correlation rule that recognises the combination of events that characterise the type. This set of rules evolves with time: rules are edited, new ones are added, and old ones removed. This means not only that the set of incident types is dynamic, but also that the criteria for generating a particular incident report are subject to change. This is particularly likely to happen during the 'shake-down' period following a new SIEM deployment, major change, or on-boarding of a new customer by an MSSP, when rules are tuned to the deployment context. A large aspect of such tuning is to reduce the number of FP incidents generated by the SIEM. Arguably, data gathered during shake-down should not be used for training as it is not typical of normal operations. However, alterations do continue at a lower level throughout the operational phase to reflect changes in prevailing threats, ICT resources being monitored, available logs and threat intelligence sources, and business priorities.

- Much of the information on an incident is not in the incident report itself, but in the event records associated with it. These are highly specific to the source (e.g. the type of firewall) that generated them, and each source may generate many different event types. For example, the log event index for SonicWall SonicOS 6.5.4 goes up to 1626. The full, 'raw' form of one of these events contains a large number of parameters, often in Syslog format. Even with a particular source, the number and nature of the parameters can vary from event type to event type. SIEMs typically only parse the event records partially, picking out key values referenced in the correlation rules. However, it is notable that experienced analysts tend to head straight for the raw form, indicating that it contains valuable information that has not been extracted, or else that the format used by the SIEM, case management tool, etc. obscures rather than reveals the useful information. All of this means that the space of parameters available for use as features in categorising an incident is extremely large. The definition of the features and of how they are determined from the available parameters, and selection of the best set of features (known as "feature engineering") is going to be key to the success of the project, but very challenging.

- Analysts often make use of a variety of external sources of information (e.g. threat intelligence sites) when investigating an incident, the sources used depending on the nature and context of the incident. The ML algorithm therefore needs to be given access to the same or equivalent information encoded as additional features. The information is likely to vary with time (e.g. whether a hostname has been flagged as malicious), so may need to be queried in real time, or at least refreshed regularly.

THE WAY FORWARD

Because of these and other factors, we are taking a pragmatic and eclectic approach, experimenting with different combinations of ML techniques, data science, and old-school knowledge-based AI. We are working closely with analysts from METCLOUD and its partners, both to shape the requirements for the tool and to elicit their expert knowledge. It is still 'early days' but results so far are promising. We will soon have an early prototype that can be demonstrated to METCLOUD analysts and management, its partners, and customers, to create awareness and gather feedback. After further development the plan is for METCLOUD analysts to use the prototype alongside their existing tools to tune the algorithms and establish trust in its recommendations. It will then be integrated with METCLOUD's evolving and increasingly automated toolset for production use, to help the analysts work more effectively, manage their workload, and provide improved service to customers. Watch this space for further white papers documenting progress.

METCLOUD

GET CONNECTED • CYBER SAFE

- 0121 227 0730
- hello@metcloud.com
- www.metcloud.com
- @METCloud_com