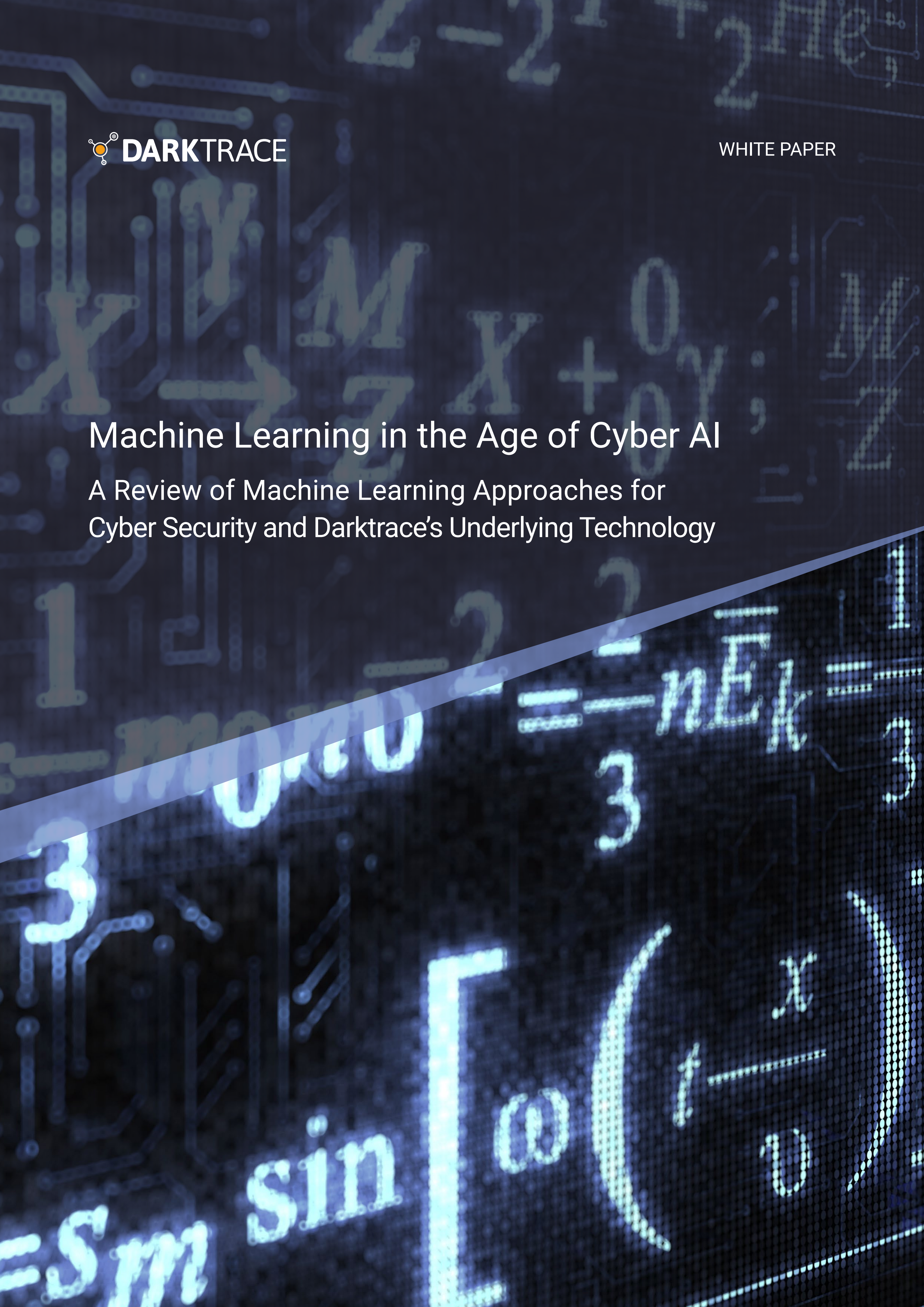
DARKTRACE

WHITE PAPER



Machine Learning in the Age of Cyber AI

A Review of Machine Learning Approaches for Cyber Security and Darktrace's Underlying Technology

Overview: A New Age of Cyber AI

A new era in cyber security has begun. In today's complex digital environments, machines are fighting machines, and advanced attackers and criminal groups are contriving sophisticated new ways to perpetrate their missions. The corporate network has become a battlefield, where the stakes are control of digital assets and, ultimately, the ability of the organization to function.

The danger today is not just the classic scenarios of data theft, or a hacked website, but the silent threat lurking beneath the surface. These attackers are quiet, creeping in unannounced, and surreptitiously changing data at will, or installing kill switches ready to be activated. Using custom code, only crossing the perimeter boundary once and never sending information outside, such threats are almost impossible to find.

Against this new reality, legacy security systems are failing, and many face extinction. This is because the traditional approach to cyber security relies on being able to define the threat in advance. Rigidly programmed to only detect known threats, this approach is no longer viable. From novel and fast-spreading attacks to insiders gone rogue, from hacked IoT devices to compromised supply chains, the threat landscape evolves in unpredictable ways and a new approach to cyber defense is urgently required.

Under this new paradigm, AI technology can identify and neutralize previously unseen cyber-threats. While machine learning has the power to transform cyber defense, the challenge of getting it to work at scale, in a variety of dynamic data environments, while detecting genuine threats in real time, without human intervention, is not trivial.

With the first AI for cyber defense proven to work across diverse digital enterprises, Darktrace is the world leader in detecting and autonomously responding to cyber-threats that legacy systems miss. Powered by machine learning and AI algorithms, Darktrace's 'immune system' technology is used by thousands of organizations worldwide.

This white paper explains Darktrace's approach to machine learning and shines a light on the unique interplay between unsupervised machine learning, supervised machine learning, and deep learning behind the world's leading cyber AI technology.

> "Darktrace's AI technology for cyber defense is a game changer - it allows us to remain resilient in the face of a rapidly evolving threat landscape."

Raspberry Pi

> "We no longer live in an era where cyber-attacks are limited to the desktop or server. Darktrace's machine learning fights the battle before it has begun."

City of Las Vegas

The Legacy Approach

Today's cyber-threats are increasingly advanced. Some are automated and fast, others slow and stealthy. Meanwhile, networks have become more and more complex.

With an ever-growing number of connections, internally and externally, it has become increasingly difficult to track all network activity, and to set parameters and signatures that will provide anything but the most basic level of protection. The perimeters of networks have essentially become redundant, while cyber-threats are evolving in unpredictable ways.

According to the traditional paradigm, firewalls, endpoint security methods and other tools such as SIEMs and sandboxes are deployed to enforce specific policies, and provide protection against recognized threats.

While these tools have a part to play in an organization's overall defense posture, they are insufficient in the new age of rapidly evolving cyber-threats. Some have become defunct as networks grow and advanced threats can increasingly bypass these controls with relative ease.

'Behavioral analytics'

Behavioral analytics is a technique that relies on correlation. For example, if an external port scan is followed by a series of failed login attempts on an external system, a correlation engine may decide that such activity looks suspicious.

The critical problem is that large systems always have some degree of correlation. Furthermore, correlation between two variables does not imply causality. If your system does not understand that, false correlations are inevitably produced.

> "Traditional tools that are programmed to spot known threats are no longer sufficient."
>
> **Heritage Education Fund**

The Limitations of the Legacy Approach

- Perimeter controls are dependent on signatures, rules and heuristics – if they miss an attack at the point of entry, they have failed and cannot take further action.
- Endpoint security depends on signatures and detecting attacks that have been previously identified, and are incapable of meeting the challenges of unseen threats.
- Sandboxes are sidestepped by modern attacks, which recognize when they are in a fake space and delay the execution of malicious activity.
- Log tools and SIEM databases require inordinate manual effort to ensure data is consistently collected across the entire organization and matched against the security team's predictions of threats. As well as this being resource intensive, it relies on the security team imagining everything that might possibly go wrong without overwhelming analysts with alarms.
- So-called 'behavioral analytics' cannot detect new threats as they emerge – they rely on the rules-based paradigm of configuring how certain job titles or devices 'should' behave and then looking for deviations. This approach fails to scale to the complexity of modern businesses.

Ultimately, legacy systems have been outpaced by modern business complexity and attacker innovation, suffering from fundamental constraints:

- They need to know about all previous attacks.
- They need to perfectly understand your business and business-specific rules.
- They need a perfect way of sharing high quality information about new attacks.
- They need to guess what all future attacks and software weaknesses look like.
- They need to be able to turn all the above insights into rules or signatures that work.

Most significantly, legacy tools require victims before they can provide solutions. The age of unpredictable, fast-moving attacks has rendered this approach deficient.

Machine Learning to Date

Supervised machine learning

The proliferation of data in the modern world means that it is not just unproductive, but impossible for humans to sift through the vast amount of information generated each minute within a typical enterprise network.

Machine learning is difficult to develop and deliver, as it requires complex algorithms and an overarching framework to interpret the results produced. When applied correctly, these approaches can facilitate machines to make logical, probability-based decisions, augmenting the capabilities of human teams and uncovering previously unimaginable insights.

The most widely-applied type of machine learning is supervised machine learning, which is used in a number of commercial and industrial fields for the purpose of classification. For example:

- Payment processing companies can use state-of-the-art machine learning techniques to build models which can identify fraudulent payments in real time.
- Online video services use algorithms to understand the viewing preferences of customers in order to provide tailored recommendations for subscribers.
- Advertising firms are able to use analyses of browsing history to determine advert placement, making targeted decisions that deliver greater success than would otherwise be possible with human marketers alone.
- Computers in cars produce huge amounts of data which can be distilled to provide engineers with a better understanding of how customers actually use the vehicle, and also assist in the prediction of part failure.
- In healthcare, data collection processes mean that wellbeing can be closely monitored, problems highlighted earlier and therefore the risk of serious situations developing can be reduced.

Supervised learning works by using previously-classified data, from which the machine learns the classification system. For scenarios where behaviors are well understood and classifications are easy to determine, the output of such systems can be highly accurate. For example, state-of-the-art image classification systems are outperforming humans in some cases. Indeed, what makes supervised machine learning so powerful is its ability to learn to deal with the errors and noise of the real world, through a statistical approach.

Thus, supervised machine learning systems are best equipped to give you an explicit answer based on prior knowledge. For example, we can feed a system with lots of examples of known ransomware and it will learn the common indicators of that malware and be able to detect similar attacks in future.

Similarly, if you want to be able to distinguish 'cats' from 'dogs' within a series of images, supervised machine learning is immensely effective, because society has a lot of existing pictures of known cats and dogs that can be fed in to train the system, and new types of cats and dogs do not often emerge in the world.

However, overfitting is a common problem in supervised machine learning, where model parameters are too finely tuned to the training data. Instead of learning the essence of a category, the machine learns a particular example – for example, it may learn to recognize a German Shepherd, but fail to understand 'dogs' as a category, and the features that make that German Shepherd pertain to the group.

Deep Learning

Deep learning is a popular subset of supervised learning that uses many layers of inter-connected mathematical processes to create non-linear decision-making engines. Deep learning tends to substantially outperform other supervised approaches because it is able to handle far more complex representations or beliefs about the world without humans having to tell the system what the data is made up of.

This powerful representation comes at a cost though, as deep learning requires computing power of a different order of magnitude to be able to train the mathematical engines.

Deep learning is expected to increasingly replace traditional algorithmic approaches across all of computing where sufficient input data, examples of the expected output, and an automated way to measure whether the algorithm is successful, are available.

Machine Learning & Cyber Security

Traditional approaches to cyber security are based on identifying activities that resemble previously known attacks – the 'known knowns'. This is usually done with a signature-based approach, whereby a database of known malicious behaviors is created. New activities are cross-referenced with the database and those that match are flagged as threats.

These solutions sometimes also use methods based on supervised machine learning, which help to classify the output of the signatures. Using this supervised approach, a system is fed a training data set in which each entry has been labeled as belonging to one of a set of distinct classes.

In the information security context, the system is trained using a database of previously seen behaviors, where each behavior is known to be either malicious or benign and is labeled as such.

New activities are then analyzed to see whether they more closely match those in the malicious class, or those in the benign class. Any that are evaluated as being sufficiently likely to be malicious are again flagged as threats.

Systems that rely entirely on supervised machine learning have fundamental weaknesses:

- Malicious behaviors that deviate sufficiently from those seen before will fail to be classified as such, hence will pass undetected.
- A large amount of human input is needed to label the training data.
- Any mislabeled data or human bias introduced can seriously compromise the ability of the system to correctly classify new activities.

Machine learning has presented a significant opportunity to the cyber security industry. New machine learning methods can vastly improve the accuracy of threat detection and enhance network visibility thanks to the greater amount of computational analysis they can handle. They are also heralding in a new era of autonomous response, where a machine system is sufficiently intelligent to understand how and when to fight back against in-progress threats.

Darktrace's unique combination of machine learning approaches

While supervised machine learning can be powerful, Darktrace was founded with the vision to build the first self-learning cyber defense platform. Using unsupervised machine learning instead allowed the system to uncover rare and previously-unseen threats, which did not rely on inherently imperfect training data sets. Data relating to historical attacks does not necessarily protect against future ones.

Having built the world's leading machine learning system for cyber security, which is based on this unique approach, Darktrace also uses deep learning techniques to supplement its AI engine with the specialized domain expertise of Darktrace's world-class cyber analysts.

Deployed extensively in thousands of real-world network environments, these new techniques are increasingly powerful, feeding our neural networks and allowing the power of unsupervised machine learning to be further augmented.

> "With Darktrace, talk about AI in cyber security has turned into action."
>
> **Ovum**

Unsupervised Machine Learning

Darktrace's unsupervised machine learning is critical because, unlike supervised approaches, it does not require labeled training data. Instead it is able to identify key patterns and trends in the data, without the need for human input. Unsupervised learning can therefore take computer processing beyond what programmers already know or can imagine, and discover previously unknown relationships.

Darktrace uses unique unsupervised machine learning algorithms to analyze network data at scale, and make billions of probability-based calculations based on the evidence that it sees. Instead of relying on knowledge of past threats, it independently classifies data and detects compelling patterns. From this, it forms an understanding of 'normal' behaviors across the network, pertaining to devices, users, or groups of either entity, and detects deviations from this evolving 'pattern of life' that may point to a developing threat.

Core principles of Darktrace's machine learning

- It learns what is normal within a network 'on the job' – it does not depend upon knowledge of previous attacks.
- It thrives on the scale, complexity and diversity of modern businesses, where every device and person is unique.
- It turns the innovation of attackers against them – any unusual activity is visible.
- It constantly revises assumptions about behavior, using probabilistic mathematics.
- It is always up to date and not reliant on human input.

The impact of Darktrace's unsupervised machine learning on cyber security is transformative. Its cyber AI technology has quickly proved itself capable of seeing hitherto undiscovered cyber events, from a variety of threat sources, which would otherwise have gone unnoticed. These include:

- Insider threat – malicious or accidental.
- Zero-day attacks – previously unseen, novel exploits.
- Latent vulnerabilities – dormant vulnerabilities that are undiscovered, often due to the lack of network visibility.
- Machine-speed attacks – ransomware and other automated attackers that propagate and/or mutate very quickly and are virtually impossible to stop and neutralize using human-dependent response mechanisms.
- Silent and stealthy attacks that lurk in networks undetected.

Reverend Thomas Bayes

The cutting-edge mathematics at the forefront of Darktrace's machine learning approach are anchored in the seminal work of British mathematician Thomas Bayes (1702–1761). His theory of conditional probability provides a mathematical bridge between objective, developed methods and the subjective world that we populate. An advanced approach to Bayesian theory, developed by mathematicians from the University of Cambridge, provides a filter to ascertain the true meaning of vague and profuse data.

Darktrace's use of Bayesian probability as part of its unsupervised machine learning approach uniquely enables Darktrace's technology to:

- Discover previously unknown relationships.
- Independently classify data.
- Detect compelling patterns that define what might be considered normal behavior.
- Work without prior assumptions when needed.

> "Machine learning can detect things that we can't predict and define. It's like finding a needle in an enormous haystack."
>
> **Steelcase**

Technical Overview

Darktrace's transformative approach to cyber defense relies on probabilistic methods developed by Cambridge mathematicians. Employing multiple unsupervised, supervised, and deep learning techniques in a Bayesian framework, the Enterprise Immune System can integrate a vast number of weak indicators of anomalous behavior to produce a single clear measure of threat probabilities.

For each unique environment, Darktrace generates millions of interrelated mathematical models which are correlated to ensure that only truly anomalous behavior is detected without a profusion of false positives. Unlike rules-based computation, the results that probabilistic mathematics generate cannot simply be categorized as 'yes' or 'no' but instead indicate degrees of certainty, reflecting the ambiguities that inevitably exist in dynamic data environments.

Ranking threat

The Enterprise Immune System accounts for ambiguities by distinguishing between the subtly differing levels of evidence that characterize network data. Instead of generating the simple binary outputs 'malicious' or 'benign', Darktrace's mathematical algorithms produce outputs marked with differing degrees of potential threat. This enables users of the system to rank alerts in a rigorous manner, and prioritize those which most urgently require action, while removing the problem of numerous false positives associated with a rule-based approach.

At its core, Darktrace mathematically characterizes what constitutes 'normal' behavior, based on the analysis of a large number of different measures of a device's network behavior, including:

- Server access
- Data volumes
- Timings of events
- Credential use
- Connection type, volume, and directionality
- Directionality of uploads/downloads
- File type
- Admin activity
- Resource and information requests

Clustering devices

In order to model what should be considered as normal for a device, its behavior is analyzed in the context of other similar devices on the network. Darktrace leverages the power of unsupervised machine learning to algorithmically identify significant groupings of devices, a task which is impossible to do manually on even modestly-sized networks.

To create a holistic image of the relationships within the network, Darktrace employs a number of different clustering methods, including matrix-based clustering, density-based clustering, and hierarchical clustering techniques. The resulting clusters are then used to inform the modeling of the normative behaviors of individual devices.

Network topology

A network is far more than the sum of its individual parts, with much of its meaning contained in the relationships among its different entities. Darktrace employs many mathematical methods to model the multiple facets of a network's topology, allowing it to track subtle changes in structure that are indicative of threats.

One approach is based on iterative matrix methods that reveal important connectivity structures within the network, in a similar way to advanced page-ranking algorithms. In tandem with these, Darktrace has developed innovative applications of models from the field of statistical physics, which allows the modeling of a network's 'energy landscape' to reveal anomalous substructures that could represent the first symptoms of compromise.

Network structure

A further important challenge in modeling the behaviors of a dynamically evolving network is the huge number of potential predictor variables. For the observation of packet traffic and host activity within an enterprise LAN or WAN, where both input and output can contain many inter-related features (protocols, source and destination machines, log changes, and rule triggers etc.), learning a sparse and consistent structured predictive function is crucial.

In this context, Darktrace employs a cutting-edge large-scale computational approach to understand sparse structure in models of network connectivity based on applying L1-regularization techniques (the lasso method). This allows the Enterprise Immune System to discover true associations between different elements of a network which can be cast as efficiently solvable convex optimization problems and yield parsimonious models.

Recursive Bayesian Estimation

To combine these multiple analyses of network behavior, generating a single comprehensive picture of the state of the devices that comprise a network, Darktrace leverages the power of Recursive Bayesian Estimation (RBE). Using RBE, Darktrace's mathematical models are able to constantly adapt to new information as it becomes available to the system. Continually recalculating threat levels in the light of new data, the Enterprise Immune System can discern significant patterns in data flows indicative of attacks, where conventional signature-based methods see only chaos.

Darktrace & Deep Learning

Darktrace also uses deep learning to enhance modeling processes. Deep learning is a subset of machine learning that uses the cascading interactions of layered mathematical processes – known as neural nets – to give intelligent systems a higher degree of insight. Multi-layered neural nets can improve the detection and remediation of certain threats, for example, in the identification of DNS anomalies, which are less effectively tracked by other machine learning methods. Darktrace's deep learning system assigns a score to all DNS data from a device, with the purpose of identifying suspicious activity even faster.

Darktrace also clusters devices into peer groups, based on its own understanding of how those devices behave, and uses supervised learning to uncover sequences of breaches, unusual patterns, or to detect aberrant activity at a higher, more holistic level. For example, the WannaCry ransomware was easily detected by Darktrace as it breaches a number of different 'pattern of life' models. Using supervised learning, Darktrace can replicate the process of a human interpreting various sets of breaches for a device or network over time and so present correlated alerts instead of a multitude.

Supervised learning is also used by Darktrace to understand more about the environment, without a human having to label it. By observing millions of different smartphones, for example, Darktrace gets faster and faster at identifying a new device as a 'smartphone', and even what type of smartphone it is.

Using deep and supervised techniques to complement its core unsupervised machine learning algorithms, Darktrace builds up unique, contextual knowledge about network activity and integrates the insights of our global deployments to improve threat detection.

Finally, Darktrace also uses deep learning techniques to automate repetitive and time-consuming tasks carried out during investigation workflows. By analyzing how seasoned cyber analysts interact with the Threat Visualizer, triage alerts, and leverage third-party sources, Darktrace is able to replicate those expert behaviors and automate certain analyst functions. This allows for increasingly efficient and simplified investigations for analysts of all maturity levels. It also gives security teams the crucial time they need to focus on higher-value strategic work, such as managing risk and focusing on broader improvements to the business.

Autonomous Response with Darktrace Antigena

Because Darktrace's machine learning is capable of understanding, at a granular level, the 'pattern of life', and therefore detecting specific deviations from normal activity, it is also uniquely capable of generating an appropriate autonomous response to an in-progress attack.

Empowering the machine to fight back autonomously for the first time, Darktrace Antigena works like antibodies within the immune system, neutralizing a threat by enforcing the known 'pattern of life' of a device or user.

Thanks to Darktrace's core unsupervised machine learning, this solution can also learn from itself, as well as learning passively from the data that it observes. For example, when Darktrace Antigena generates an autonomous response action, a feedback reinforcement loop is triggered. The resulting behaviors on the network are analyzed in turn to facilitate diagnosis and inform any further actions. Unlike guided reinforcement learning, this process is driven autonomously by the machine itself rather than a human operator.

Critically, Darktrace Antigena is built on unsupervised machine learning proven to detect only the most abnormal cyber events to a degree of accuracy that enables it to take precise action in response. Machine learning used in this way does not replace the human's function, but ultimately serves to enhance it. Antigena acts faster than a human, buying the operator precious time to catch up and take further measures if necessary.

> "Darktrace has completely changed our approach to cyber security. Autonomous response allows my team to spend its time and effort where it is really needed."
>
> **Campari**

Conclusion

Our generation is witnessing the machine learning revolution. We are seeing shifts in working practices brought about by the replacement of muscle with machine, the automation of repetitive tasks, and now the replacement of low value, thoughtful tasks with machines capable of handling big data and making vast calculations.

As networks have grown in scope and complexity, the opportunities for attackers to exploit the gaps have increased. Walls are no longer enough to protect the corporate network, and rules-based tools cannot keep up with all possible attack vectors. A constantly evolving cyber-attack landscape requires a step up in our detection capability, using machine learning to understand the environment, filter the noise and take action where threats are identified.

Utilizing probabilistic Bayesian mathematics developed by mathematicians from the University of Cambridge, Darktrace is a world leader in machine learning and artificial intelligence. The unique interplay of unsupervised machine learning, supervised machine learning, and deep learning, which powers the Enterprise Immune System has enabled Darktrace to become the world-leading AI company for cyber defense.

Under this new paradigm, we can build organizations their own immune systems that can autonomously catch and fight back against the cyber-threats that others miss, without any human input or bias about what 'bad' looks like.

Darktrace's technology has become a vital tool for security teams attempting to understand the scale of their network, observe levels of activity, and detect areas of potential weakness. These no longer need to be manually sought out, but are flagged by the automated system and ranked in terms of their significance.

Machine learning technology is the fundamental ally in the defense of systems from the hackers and insider threats of today, and in formulating response to unknown methods of cyber-attack. It is a momentous step change in cyber security.

About Darktrace

Darktrace is the world's leading cyber AI company and the creator of Autonomous Response technology. Its self-learning AI is modeled on the human immune system and used by over 3,000 organizations to protect against threats to the cloud, email, IoT, networks and industrial systems.

The company has over 900 employees and headquarters in San Francisco and Cambridge, UK.
Every 3 seconds, Darktrace AI fights back against a cyber-threat, preventing it from causing damage.

Contact Us

North America: +1 (415) 229 9100

Latin-America: +55 11 97242 2011

Europe: +44 (0) 1223 394 100

Asia-Pacific: +65 6804 5010

info@darktrace.com

darktrace.com