Incident Response Ups the Ante Against Evolving Threats

Machine learning and advanced analytics give Incident Response a fighting chance against a changing threat landscape

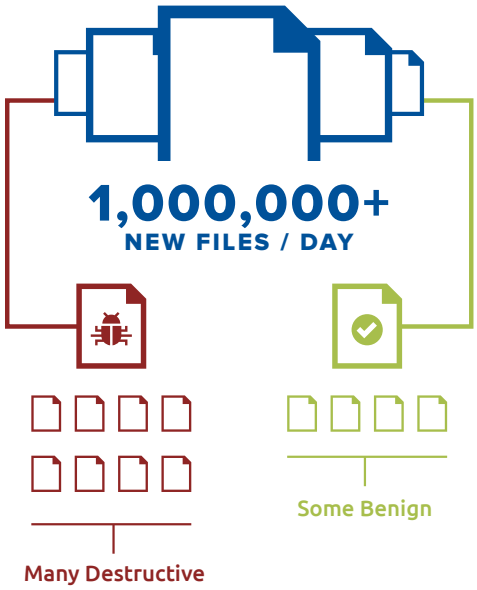
BLUVECTOR®

Contents

A Rising Tide of Threats

It's no secret that cyber threats to companies around the world are growing in number and intensity. Threat actors are more organized and the payloads more damaging. More than one million new files are seen a day, some of which are benign, **but more are destructive.**[1] An arsenal of security products doesn't help differentiate good from bad; in fact, it creates more confusion. Security products bombard the average organization with 17,000 malware alerts weekly, or almost 2,500 each day.[2] How many are real threats, and how many are false positives that the security analysts will need to chase?

No one can hope to keep up, especially given the wide variety of malware and the fact that much of it can lie dormant for long periods of time before it begins to do damage. In the infamous Sony hack of 2014, investigators say malware lived on the network for at least two months, identifying important files and planning for their exfiltration. Incident Response (IR) teams spend an inordinate amount of time on work that could and should be automated, while real threats can be wreaking havoc. Too many known threats escalate to the IR team to investigate, and time spent trying to identify the "needle in the haystack" mean that IR teams are in a losing battle.



IR Teams on the Front Lines

The first line of defense—the firewall, IDS/IPS, anti-malware and everything else arming the Security Operations Center (SOC)—is a necessary first start, but everything that gets past that line is typically escalated to the next level: an audit team or, in most larger organizations, an Incident Response team. The IR team's work is not limited to reacting to these incidents, rooting out breaches, containing them and managing the cleanup. This team also engages in a level of proactive efforts, driven largely by the increase in new, advanced and more damaging malware and ransomware (up 35% from 2014 to 2015[3]) that has affected up to 40% of all companies.[4]

---

[1] *Virginia Harrison and Jose Pagliery. Cyber Attack Hacks Security. CNN Tech, April 2016. Retrieved from http://money.cnn.com/2015/04/14/technology/security/cyber-attack-hacks-security/*
[2] *Ponemon 2016 Cost of Data Breach Study: Global Analysis. Ponemon Institute. 2016. Retrieved from https://securityintelligence.com/media/2016-cost-data-breach-study/*
[3] *Symantec Internet Security Threat Report 2016. Symantec Corporation, April 2016. Retrieved from https://www.symantec.com/content/dam/symantec/docs/reports/istr-21-2016-en.pdf*
[4] *Understanding the Depth of the Global Ransomware Problem. Osterman Research 2016. Retrieved from: https://go.malwarebytes.com/OstermanRansomwareSurvey.html*

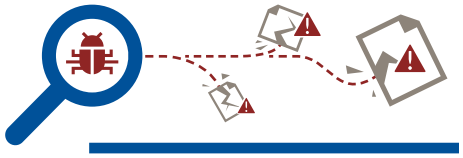
All this comes at a time when corporate networks are becoming increasingly more complex with more endpoints, including not only desktops and laptops, but IoT systems that are unmanageable, as well as standard systems with legacy or unique operating systems, such as HVAC or medical devices.

How do IR teams handle both reactive and proactive challenges in today's complex network? At a minimum, most will look for predictions: they review threat intelligence feeds, parse out what's relevant to them and either look for the threat in their own environment or take precautionary measures to prevent its entry. Some IR teams react to hunches, investigating what they've seen in forums or a social media topic that's gone viral, perhaps indicating increased, opportunistic spear phishing attempts. They routinely monitor Registry Keys, and use custom scripts and tools for an added layer of defense—troves of network logs are stored to do forensic investigations and look for trends among aggregated data. While the tactics of an incident response team will vary, the intent is always the same—get ahead of what's breaching the network to stop it before it does harm.

Confronting the Challenge of Tomorrow vs Yesterday

Let's face it, incident response is taking on too much. Working on hunches and relying on cleanup efforts to be sufficient is simply not enough, especially with the potential of ransomware or destructive malware that could take down entire systems and permanently erase file servers and brick devices. Stumbling upon the most damning of breaches in 100–200 days[5] or hearing about it when the FBI calls and informs you of a breach is not an acceptable outcome when the threats are more severe than ever. Investing in technology to detect breaches by their activity after they happen is a gamble of the severity of the threat.

The future of security, and the job of the IR team, is made even more complex due to the scarcity of skilled cybersecurity professionals. Cisco estimates there are more than 1 million unfilled cybersecurity positions around the globe. It is difficult and costly to hire analysts with the unique skills to investigate advanced threats that have bypassed your existing security defenses. Degree programs are still relatively rare, and many security experts have learned through certificate programs or on-the-job experience.



Stumbling upon the most damning of breaches in 100–200 days[5] or hearing about it when the FBI calls and informs you of a breach is **not** an acceptable outcome when the threats are more severe than ever.

[5] *M-Trends 2017. Mandiant, 2017. Retrieved from https://www.fireeye.com/current-threats/annual-threat-report/mtrends.html*

A recent analysis of US Bureau of Labor Statistics showed that more than 209,000 US cybersecurity jobs are unfilled, and the demand for positions is expected to grow by 753% through 2018.[6] With skilled and experienced analysts being actively recruited by multiple companies, it's important to ensure that your IR team has the support, and the right tools and automation to be successful without needing to rely on a rare, specially skilled cybersecurity unicorns.

Traditional Tools for a Modern War

The brunt of the work piling on incident response is due to the fact that traditional tools have known deficiencies that have been exploited by adversaries at alarming rates. The more adversaries are able to get through existing defenses, the more incident response teams are compelled to use creative techniques to uncover them.

Today's IR teams are a strong line of defense against breaches that are escalated from the SOC team, and they should be equipped with the most effective weapons in order to protect the organization before threat actors take hold and irreparable damage can be done. Traditional products like anti-virus, firewalls and intrusion detection systems are brittle, relying on constantly-updated signatures, rules or pattern matching to detect yesterday's threats. They are unsuitable for dealing with constantly modified and sometimes targeted malware and new attack forms, encryption methods, exploit kits and evasion techniques. To fill the gaps, we have turned to sandboxing, emulation and virtual detonation.

While promising, these tools can suffer from serious deficiencies:

- **They don't communicate well among themselves,** so data is scattered, and one security tool is unaware of what the others have found.
- **They can be evaded.** Common techniques include file obfuscation to allude signatures or delaying execution to wait-out the timeout period of a sandbox among several other techniques.
- **They have processing limitations.** For example, a sandbox cannot analyze everything, so it focuses on a subset of traffic and are slow to make a determination.
- **They don't provide context,** so it's difficult to make a reasoned determination as to what is really happening on the network—including where it originated and where it was destined.

The challenge that Incident response teams face is the targeted, advanced threat that is unknown, or exploits zero-day vulnerabilities. This is what keeps them up at night.

[6] *Arina Setalvad, "Demand to fill cybersecurity jobs booming' Peninsula Press, March 2015. Retrieved from http://peninsulapress.com/2015/03/31/cybersecurity-jobs-growth/*

Learn From Experience, Then Automate For Anyone

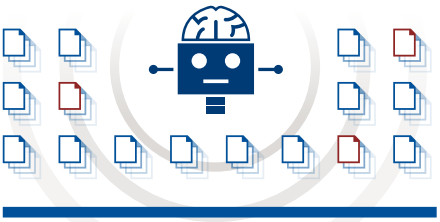
It's time to take a few lessons from how experienced cybersecurity professionals look for adversaries that have breached their defenses. With automation and intelligent data correlation it's possible to make efficient use of the tools, data and services already at their hands. It's also important to incorporate new technologies, where they make sense. With some new technologies, particularly with machine learning, analyses can be made similarly to how a reverse engineer or forensic analyst would perform her analysis but done so at incredible speed and scale.

With the right tools, it's possible to shift the mounting burden of uncovering and investigating a greater number of suspicious threats to less experienced Level I analysts. This then frees up higher-level analysts to do more purposeful investigations knowing that most of the hunches and combing through log files are covered by smart tools and automation. And the workload from these tools can easily be accomplished by less-experienced team members investigating suspicious threats that normally has taken several full-time employees resulting in a higher investigative cost per incident.

Machine Learning and Modern Defenses

Machine learning holds a lot of promise but what gets lost and is often confusing among the cybersecurity tools is that not all machine learning technologies are created equal; and what separates many is a difference in philosophy. For some, they believe there's reason to give up on catching a breach and instead focus efforts on post-breach activity by looking for anomalous behavior within their environment. On the contrary, others are committed to detecting the breach and giving incident response a fighting chance at getting ahead of the threat before harm is done.

BluVector uses supervised machine learning that analyzes files and software like how a reverse engineer would but in milliseconds across hundreds of pieces of content and millions of packets per second—a scale and volume that's not possible by hand. BluVector's machine learning, initially developed in the intelligence and defense community, has been finely trained by BluVector's malware experts and data scientists over the last decade with the aim of identifying threats that have never been seen before without any prior knowledge or intelligence of the threat.



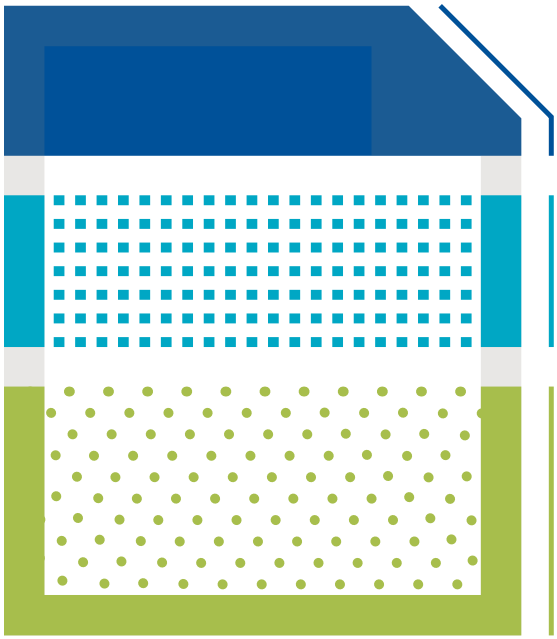
BluVector's supervised machine learning analyzes content in milliseconds across hundreds of pieces of content and millions of packets per second.

To do this, 30 file-type specific machine learning models analyze all layers of the content looking for features to determine if something is benign or malicious. This is done by comparing all layers of a file (figure 1).

Like a reverse engineer, the machine learning models go beyond a single characteristic to determine benign or malign—instead they look at look at all layers of the file searching for the presence and absence of a combination of characteristics they know to be typical in good or bad content. For a reverse engineer, the characteristics she looks for are based on what she has learned honing her craft, learning new telltale signs cultivated over years of experience. From one reverse engineer to another, experiences vary and subtle biases can occur.

BluVector's supervised machine learning models were trained on millions of pieces of content both benign and malicious and consider thousands of characteristics referred to as features. Features may include byte sets, tools malware authors use to obfuscate their malicious code, file size or entropy found within the metadata. The telltale signs or features the machine learning models look for have no bias, look at a complex combination of features both present and absent, and are based on the experience having trained on millions of files. The power of supervised machine learning is being able to achieve the equivalent of the art and science a reverse engineer wields but at speed, volume and accuracy not achievable by humans.

FIGURE 1: Example of a .exe file



Metadata layer: header information, file metadata, section metadata and more.

Code/data layer: More complex than byte sequence, this layer contains information such as IP addresses, API functions, etc.

Raw bytes layer: the lowest representation of a file, which includes a variety of information. BluVector's approach looks for very small byte sets and considers them in combination with the presence or absence of other features to create resilience against modifications or variants.

A Few Words About Machine Learning

Machine learning is the process of computer programs becoming more accurate at a task due to exposure to data, called training instances. However, several factors from the dataset, the chosen algorithm and features selected to train on provides vast variability both in accuracy and application.

For cybersecurity and for the purposes of this paper, we look at the two main categories that separates the approaches where machine learning is applied: unsupervised and supervised. To understand the differences in these approaches, let's review their most common use cases starting with their precursors.

The lowest level of sophistication is the traditional non-machine-learning solution: a signature-recognition system such as anti-virus (AV). An AV system looks for an explicit match to a known virus, and if it is detected, the system can take action to stop the threat. However, many anti-virus tools rely solely on lengthy byte sequences, and as a result are very brittle: the signatures (like rules or patterns) must be an exact match and a change in a single byte can break the signature. Since bad actors constantly make minor modifications to avoid detection, signatures need to be updated constantly. One benefit of signature-based approaches is that they have very low false positive rates on yesterday's known threats. This non-machine-learning approach was the status quo for years. Now, machine learning has come to the fore.

Understanding Unsupervised and Supervised Learning

Unsupervised

Unsupervised machine learning in its use within cybersecurity can easily be considered more sophisticated than traditional signature-recognition systems. When this method is used, there is no pre-labeling or sorting of the dataset. There are no "right" answers instead the algorithm is left to draw a reasonable conclusion for sorting the dataset by making reasonable determinations of how data should be clustered. This method is often used in anomaly detection—looking for anomalous network traffic or user behavior. In an initial setup, "normal" is defined by observing regular behavior for a period of time. Its goal is to find unusual patterns or behaviors by identifying what network or user behavior falls outside the "normal" behavior it has observed. Unsupervised machine learning in anomaly detection tends to be noisy since anomalous behavior is not necessarily malicious—resulting in false positives. In fact, most malicious software mimics normal behavior, leading to false negatives.

Supervised

Supervised machine learning, in contrast, is used for predicting the classification of unknown data input based on learned features by observing a pre-labeled dataset. As an example, supervised machine learning could be used to distinguish between cats and dogs based on the features or characteristics that it has learned from a dataset of pre-labeled images of both cats and dogs. When faced with an unknown image, it could then make a probabilistic determination to classify it as a cat or dog. The classification doesn't need to be binary but for our example this is sufficient.

Within cybersecurity it works by classifying features of content labeled as benign or malicious. It will look at a combination of attributes that are together statistically significant, while on their own may not indicate something bad. By considering both the presence and absence of features to determine the class of an unknown piece of content means a piece of software that doesn't look malicious may still be marked suspicious because it also doesn't look like benign software. Manually reviewing this suspicious content and considering outside factors such as threat intelligence, if it is confirmed malicious the machine learning model can be augmented by considering the features of this newly-discovered malicious software.

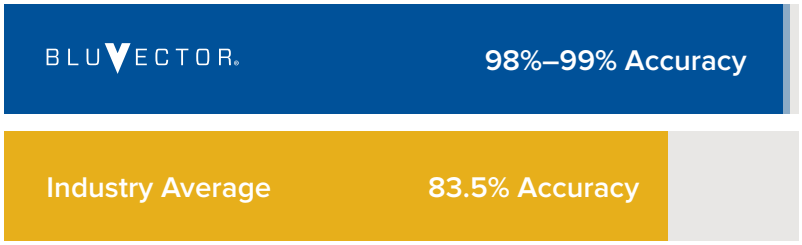
Supervised machine learning has the unique ability to perform static analysis to predict good and bad based on samples curated by experts—and augmented with additional samples collected over time—results in both incredible speed and accuracy at classifying content that is unknown. Adversaries today are relying on tricks and tactics to bypass security controls, deploying polymorphic malware and exploiting zero-day vulnerabilities. Relying on the content itself—the ground truth of the file that can never be altered—alleviates tactics like pausing the malware when it detects it's in a virtual machine and exceeds the wait-out period of a sandbox to pass through as safe. With the proliferation of ransomware and destructive malware, waiting to observe the behavior is a chance not worth taking. With the right supervised machine learning in cybersecurity, resilience, accuracy and speed can be achieved.

Accurately Detecting Unknown Threats

When searching for unknown or evolving threats, supervised machine learning can be a valuable tool for analyzing vastly more content with predictive accuracy that's not likely in traditional cyberdefense tools.

BluVector provides the IR team with an extremely accurate solution. Its supervised machine learning looks at thousands of characteristics per file to classify a file as more likely malicious or benign. The model can then be continuously enriched with added data sets or confirmed determinations of good or bad by an analyst of suspicious threats identified by BluVector. Based on the traffic that moves across the organization's own specific network, any of the machine learning models can be evolved locally on-premises for further accuracy with respect to the organization's specific environment and targeted threats.

Independent lab tests show that BluVector models each achieve 98%–99% accuracy[7] far above the industry average of 83.5%.[8]



BluVector aids the IR team in another important way: it provides the analyst with context. By collecting and correlating all the relevant context from the metadata, other detection tools, network logs, threat intelligence and other sources, BluVector ensures that all the information analysts need is at their fingertips, enabling them to make informed decisions faster.

Staying Ahead of Bad Actors

Any solution put in the hands of the IR team should automate technique and processes already working but with technology can be achieved at speed and scale with accuracy. For an IR team to get ahead of the curve and in front of advancing threats, a reduction in the time from detection to containment is critical. In today's world, response at digital speed is possible. This can be done through a network solution (appliance- or cloud-based) that scans all network traffic to analyze its contents. Quick detection calls for near-real-time analysis; overcoming the latency issues found with sandboxes and the need to update signatures or rules in other tools. The results could be made available to analysts immediately, in an easy-to-understand format, to enable faster time to resolution.

[7] *The 451 Report. BluVector malware detection and analytics console will work for data. April 2016. Retrieved from https://451research.com/report-short?entityId=88718*

[8] *BluVector Cyber Threat Detection and Hunting Platform. Miercom, April 2016. Retrieved from http://miercom.com/pdf/reports/20160205.pdf.*

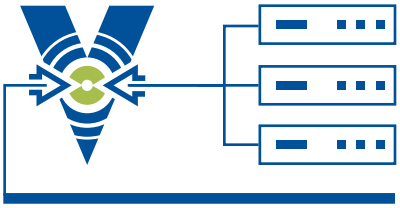
The BluVector solution is a network security monitoring and analytics platform available now as an appliance and soon as a cloud service that inspects all traffic at line speed. It analyzes 30 different file types, including image, compressed and archive formats. BluVector resides on the network and passively monitors all traffic, so it can detect potential security risks long before a file reaches the endpoint. Because it's passive, it does not disrupt normal business and it cannot be detected by anything malicious passing through.

Building a Case with Context

A malware analyst would expect to substantiate the case of a suspicious threat by looking at third-party threat intelligence based on any correlated information such as IP address or relevant metadata. She may even drill down and look at the metadata herself to see if anything stands out. AV signatures and custom rules or scripts that are triggered and by running it though a sandbox would help to definitively support the suspicion of a threat. A forensic analyst may comb through network logs for where the threat originated and where it was destined in addition to other activity that might have occurred—such as an additional file being downloaded—that may also be worth investigating.

BluVector's machine learning engine assigns a confidence score up to 100 of how likely it is malicious. Integrating with other tools and data sources, BluVector then provides this context around the suspicious threats using data already being collected in the environment, automating its collection and correlation for more insightful decision-making.

For example, it collects relevant network logs both before-and-after the event. For less suspicious threats, BluVector can automatically or manually send the event to a sandbox and pull the results back in. It is integrated with both external threat intelligence sources and other detection tools. If any threat intelligence feeds correlate with a threat, an analyst can review the information to learn more, and if a threat matches a signature or rule, BluVector can automatically route it directly to a Security Information and Event Management (SIEM) tool for blocking. This level of integration helps IR analysts with fast, accurate and complete information to connect the dots and work within existing processes.



BluVector's integration helps IR analysts with fast, accurate and complete information to connect the dots and work within existing processes.

Streamlining and Simplifying Analyst Workflows

If it's difficult to find seasoned security analysts, it's even harder to keep them from getting bogged down in manually-intensive, slow work steps. The right tools for the IR team would allow senior analysts to focus on high-value tasks, advanced work now becomes more routine allowing inexperienced analysts to pick up a larger portion of the work without incurring additional risk. Given the scarcity of seasoned security analysts, advanced tools can help ensure that the ones currently onboard can work as efficiently as possible. In addition, it's very helpful if the tool allows the IR team to connect the dots, correlating information from a variety of tools to provide full context and remove doubt.

BluVector satisfies these requirements by design. BluVector brings a 5:1 efficiency gain, dramatically increasing the incidents analyzed and reducing the time from detection to containment. By allowing analysts to uncover new, never-before-seen threats, the IR team can cover more ground and improve overall organizational security. In addition, its automation empowers Level 1 analysts to respond faster and have advanced visibility into threats. BluVector enables Level 1 analysts to increase the scope of their work, investigating more suspicious threats and gaining additional value from threat intelligence and other tools thanks to the automation and data correlation provided by BluVector. This frees up higher-level analysts to do more purposeful investigations with their time.

BluVector Increases IR Team Effectiveness—and Your Organization's Security

The cyberthreat is growing far faster than traditional tools or mere humans can cope. Automation is the answer, and machine learning is key to its success. Because machine learning solutions can take many different forms, it is worthwhile ensuring that the solution for malware detection and analytics employs an approach that is accurate, effective and field-proven, and is tuned to your specific environment. The supervised machine learning approach as implemented by BluVector is designed to ensure accuracy at detecting unknown, advanced threats—even those targeting your specific environment. This network-based security monitoring and analytics platform operates at line speed, producing fast time-to-results, and its accuracy far surpasses any other approach on the market.

With the ability to evolve the machine learning models based on an organization's own environment, the BluVector platform produces highly tuned results, and at the same time makes itself much less vulnerable to evasion. Even if a bad actor could obtain the system, the results obtained from factory settings would not be the same as those found in any given organization where learning has tailored the results to the specific environment.

BluVector makes the IR team's work smoother and more effective. Low false positive and false negative rates mean that the team can focus on what is really important to the organization. By combining information from a variety of security tools, BluVector can send alerts to a SIEM to integrate with existing processes and provide full context in an analyst-friendly console.

With BluVector, it's finally possible for IR teams to connect the dots, focus on new and emerging threats that are bypassing other defenses, and achieve unprecedented efficiency to stop them before they do harm by taking action within minutes, not months.

BluVector helps IR teams stop emerging threats by taking action within **minutes, not months**.

BLUVECTOR®

BluVector helps security teams respond to malicious threats up to 80% faster than current approaches. As a leader in Network Security Monitoring & Analytics, BluVector applies supervised machine learning and automation so security teams can detect and respond to advanced security threats at digital speed. For more information visit: **www.bluvector.io**

www.bluvector.io • info@bluvector.io • 571.565.2100