Evolving the Cybersecurity Ecosystem

FACING THE CHALLENGES OF BAD GUYS, BIG DATA, CYBER INTELLIGENCE AND THE NEW THREAT OPERATIONS ENVIRONMENT

Don Bowers, CISSP
Chief Scientist/CTO
National Security Operation
Leidos

leidos

Abstract:

It is increasingly clear that defending networks and enterprises against cyber threats by using standard security operations modalities is (at best) marginally effective. In order to present a more effective operational cyber solution it is necessary to transition from these legacy security operations center (SOC) modalities toward an increasingly active (or better, proactive) cyber threat operations model.

Because this cyber threat operations model depends on presenting decision makers with near-real time, actionable intelligence and valid courses of action we must presuppose the effective collection, aggregation, analysis, correlation and visualization of numerous and disparate types and sources of data elements – all related to the mission of the network and its owner. Some of these data elements will be network-based (e.g., SNMP or packet captures) while others will be semi-structured (e.g., rss feeds, social media, vendor-based threat feeds, etc.) Therefore, the foundation of an efficient, cost-effective threat operations capability would be a system that can ingest these disparate data elements at extremely high speed, effectively filter and correlate them in near real time (i.e. < 50 ms.) The initial results (representing a filtered, reduced set of more pertinent data elements would then be subjected to further analysis. Only when the data elements have been correlated and a multiple analytical tools applied to them, would a picture of potential threats emerge and cause an alert to be produced. The emerging threat alert will be addressed by a multi-disciplined team of cyber professionals (security, analysts, incident response, compliance, etc.) who work together to effectively respond to an active or emerging threat.

Because much of the analytical capability employed against the raw data elements (and "enriched" by external 3rd party feeds) is based on detection of the antecedents of human behavior, the end results are inherently predictive in nature. This allows a collaborative effort among a cyber threat ops team to get ahead of a threat for the first time. Building this cyber intelligence "force multiplier" to assist cyber analysts in reducing time to detection and prediction of malicious actions is only the first step in the evolution toward the future of cyber operations. The next step is to rid ourselves of the pervasive, legacy mentality toward the way we detect network threats. In this case, we must begin (literally) to think "outside the box." Placing a sensor box on a network has arguably been at best marginally effective over time. As the cyber landscape has expanded to incorporate a cloud-based environment, network-based sensors now have very limited visibility to what happens "behind a hypervisor." A new approach involving a suite of self-organizing, lightweight, self networking (P2P) applications that function as "friendly bots" seems a better path. These autonomous agents would be controlled by and informed by but not limited by the back-end CTOC infrastructure as described above.

These new concepts represent a true fusion of cybersecurity, cyber-intelligence analysis, cyber incident response and cyber policy compliance with a "big data" analytic capability. The differentiator for this approach is a unique integration of multiple commercially available analytical technologies, Leidos-developed analytical tools and a high speed "big data" ingestion and storage architecture that when used together create a powerful cyberspace defensive capability. The net result is an intelligence-driven, risk based, end-to-end cyber solution.

Evolving the Cybersecurity Ecosystem

Facing the challenges of Bad Guys, Big Data, Cyber Intelligence and the new Threat Operations Environment

To understand the scope and complexity of what we can refer to as "the cyber cold war," one only need read a newspaper or turn on a 24 hour news channel. Every day new and more sophisticated cyber attacks are employed against government, industry, academia as well as average home computer users. The motives are as different as the attack methodologies. Some represent powerful nation-states with vast resources employed to hamper commerce or in some cases steal intellectual property. These adversaries are experienced, trained "cyber soldiers" who view their jobs in the same light as any other war fighter.

Others are organized crime elements whose intent is to steal or extort as much money as possible from businesses and private citizens using cyber attack vectors such as phishing emails and session hijacking. They are as or more effective that criminals who choose to practice more conventional methods of theft.

Rogue elements of various special interest groups represent perhaps the most dangerous threat because of their unpredictable nature. These could be radical religious groups such as al-Qaeda or so-called "social justice" groups such as "Anonymous" whose goals are to promote a specific religious, social or political agenda.

Defending against these adversaries has been a growing challenge for many years. Traditional IT security approaches have proven to be marginally effective at best. The "bastion mentality" of perimeter defense (represented by firewalls, intrusion detection/prevention and anti-virus technologies) provides limited value over time inasmuch as the technologies are forced to constantly update the foundation of their defense modality (the static defense 'signature') in order to keep ahead of evolving and ever more effective adversarial offensive capabilities.

In an effort to enhance the effectiveness of the bastion cyber tools, a security operations model was developed to aggregate feeds from network based cybersecurity appliances. These security operations centers (SOCs) house cybersecurity operations personnel that monitor the network based feeds and when alerted that a signature has detected a "security event" they will take a pre-determined action such as calling a supervisor or customer to inform them that they may have been attacked. An incident response team can be called upon to forensically determine if a real breach has occurred and attempt to assess the potential scope of this breach. Recovery from an attack or theft can then be planned. In most cases, any security policy-based compliance activities that an organization may employ are placed into the hands of information technology services or simple desktop support personnel. Coordination between these groups and the security operations is (in most cases) purely ad hoc. Figure 1, below illustrates the typical security operations architecture.

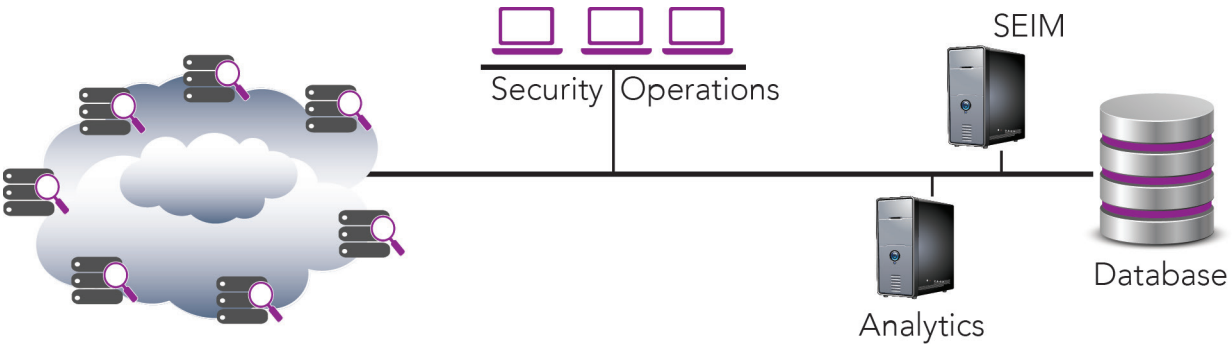

Figure 1: Prevailing Security operations Center (SOC) Architecture. This is the current state of the art for most organizations. On the left is a signature based sensor constellation that provides an alert to the SEIM. This is fine for alerting decision makers to the fact that the organization has been or (more rarely) is in the process of being compromised.

The architecture depicted in Figure 1 provides the typical "rear-view-mirror" picture of situational awareness. This prevalent architecture is very capable at telling decision makers what has happened but is totally lacking in ability to see beyond the event horizon to help forecast threats as they may emerge. Nor can it see to any great extent, beyond the network perimeter controlled from the network/security operations center.

Threat Operations

Both industry and government have begun to embrace the idea that our defenses as constructed today are just not working. In an effort to "fix' this issue it has been necessary to change the SOC model to a "TOC" or threat operations center model. This paradigm involves the addition of multiple teams with discreet functions (e.g., security operations, incident response, cyber analysis and intelligence, etc.) The interaction among these groups provides a much more holistic view into not only what is currently happening but with addition of the intelligence analysts the decision maker begins to have some view past the event horizon into events that pose potential threats to a network or enterprise.

Taking the Next Steps

In order to begin effectively taking the next steps in cyber evolution, there are two overarching requirements:

1. Increase the effectiveness of the cyber personnel (cyber warriors) who are monitoring the technical controls, sensors and defenses of networks. This must be done first in order to take advantage of the mountains of legacy cyber data elements stored in data warehouses all over the country.
2. Design and deploy the next generation ("NextGen") "sensor." In fact, "sensor" is not the right nomenclature for the capability necessary to effectively combat the existing and ever growing cyber threat. The remote constellation capability must be able to:
    a. Be deployed as either hardware or software,
    b. Communicate among multiple, disparate instances of the nodes themselves,

c. Provide not only sensing but active defense capabilities,

d. Self organize – collaborate in data sharing as well as active defense,

e. Be self protecting,

f. Be software defined,

g. Be informed by but limited by a threat-based communications, command, control and intelligence (C3I) platform as described in #1 above.

The evolution toward this cyber continuum that represents an end-to-end solution for covering each portion of the cyber spectrum is essential.

As the first step in the evolution of this new cyber ecosystem, Leidos and a group of its cyber consortium partners have started construction of the threat operations environment in the Leidos Columbia, Maryland facility. One of the fundamental underpinnings of this platform is its ability to dramatically increase the effective capability of cyber analysts. Creation of the back-end communications, command, control and intelligence (C3I) capability was undertaken first in order to begin increasing the capability for current cyber analysts and to support the security appliances already deployed by our company and its customer base. The code name for this effort is: "Raptor."

True Threat Analysis

A great deal of money and resources are now expended by government and industry in an attempt to accurately determine the origin and potential impact of cyber threats. This is a very labor intensive process and is accomplished by many trained, experienced cyber analysts looking at large volumes of data (usually collected by the tools previously described here) to "synthesize" or derive actionable information. This is a proven methodology but because of the mountains of data elements both real-time and stored that must be examined the process becomes too labor intensive for human analysis alone to be very effective. Today we call this a "big data problem."

A multidiscipline research group at the Leidos campus in McLean, VA and the associated cyber lab in Columbia, MD has developed an automated approach to dealing with this problem. The result of nearly four years of research and integration is an analytical platform with the capability (both technology and methodology) to collect data at high speeds from OUTSIDE the network, correlate this with current, INTERNAL data, subject this correlated feed to a suite of advanced cyber analytics to visualize real, actionable information.

The result of correlating the internal and external information is not only to provide operators and information providers visibility outside the network but using the internal information such as network health/status and security, asset and vulnerability information such as that obtained during continuous monitoring to provide a context which helps understand the potential impact of external events upon the internal elements of an enterprise.

With the application of a specific suite of advanced analytics which is based upon human behavior, it is possible to actually forecast specific threats as they emerge in the wild. When this is correlated against the internal data elements described above the result is a pre-zero day warning provided in the context of a specific network and its existing assets and vulnerabilities.

Functionality

The functionality of this analytical platform may be decomposed into five major sub-functions as follows:

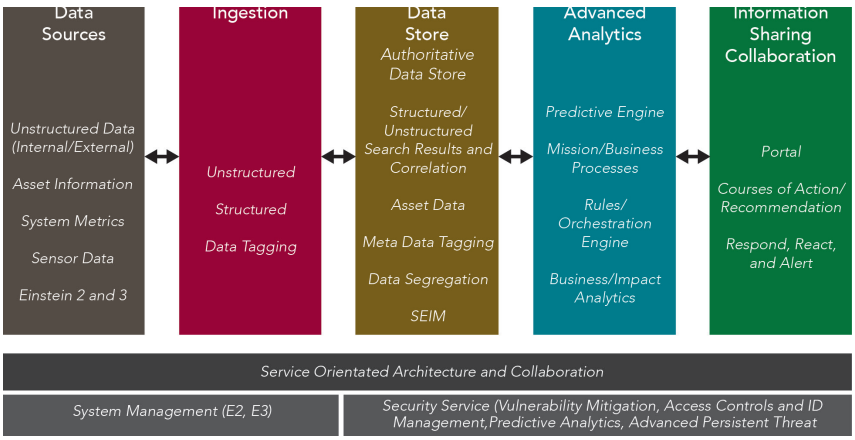

Figure 3: Solution Functionality: The five major functional areas comprising the threat analysis /situational awareness solution. The model and associated test bed were set up to represent a .gov network using current and our team's "next-gen" sensor technologies. In this case, sensor technology that closely resembles both the legacy Einstein 2 sensor constellation as well as the planned Einstein 3 sensor platform were incorporated. The latter can interact dynamically with the analytical platform to provide an active defense.

This representation was constructed to reflect how the solution might be utilized in the context of managing both live data feeds and stored, legacy data elements from existing Einstein v2 and future Einstein v3 DHS sensors.

Data is collected from both internal and external sources through multiple high-speed ingestion methodologies (dependent upon the form and type of data) and then stored in an authoritative data store that can be centralized or distributed. Correlation and analysis on the data elements is automatically synthesized into information which can then be provided to human analysts through visualization within a collaborative portal. The result is that an individual cyber intelligence analyst can be many times more productive (and accurate) in their synthesis of actionable information. This tool does not replace human analysts but instead acts as a "force multiplier" to help them be more productive.

Advanced Analytics

The current, legacy security event and information manager or "SEIM" technology represents the current best cybersecurity management practices broadly available today. These tools are designed to aggregate the data provided by current IDS/IPS and other security related tools. The SEIM performs simple correlation (associating single pairs of events) and alerts on these single-threaded, correlation events. This alert is based upon the matching of a detected event with a predefined pattern or "signature." Another alert condition can be produced if an event exceeds a heuristically developed baseline of network behavior.

The most important thing to keep in mind here is that in either case, the alert only takes place after an attack or other compromise has already occurred. This capability is important for efforts related to dealing with an ongoing attack or for forensic analysis and potential prosecution, but does not prevent or avoid the results of an attack.

As discussed earlier, what is necessary is a forward-looking, predictive capability. Again, the first step to extending cyber vision beyond the network boundary is the high speed capture and storage of data feeds that represent internal information from existing sensors (e.g., IDS/IPS, syslog, etc.) and the correlation of these against "unstructured" information (video, audio, jpegs, .rss feeds, social media, etc.) which are available both inside and outside the network. With appropriate technology (both hardware and software) it is possible to do pre-correlation at the ingestion point for these data elements. This means that advanced sensors within a network can perform some of the work for internal feeds and combine this with an aggregated feed from a SIEM. High speed collection of the unstructured data elements we are searching for can be accomplished through technologies such as Hadoop in connection with line speed data storage and tagging. Once the data has been captured and stored, a suite of advanced analytics can be applied to create actionable information.

There are two primary types of analytical engines used in the Leidos cyber situational architecture:

- Statistical analytics
- Automated, human behavioral analysis (Leidos patented, proprietary)

Statistical Analysis

In the context of this paper, statistics is one of the advanced analytical tools used to examine the large data sets associated with collecting in the cyber environment. This analytical capability is ideally suited to prediction and modeling of events related to the behavior of inanimate objects and systems. In this case, it can provide valuable projections into the potential impact of a specific attack and when used in conjunction with an expert system and an application that provides recommended courses of action (COA) the statistical analytics suite described here can provide a range of options to cybersecurity and network operations supervisors.

Human Behavior Analysis

The advent of Network Behavioral Analysis (NBA) technologies that analyze network traffic to discover unusual or anomalous activity, has given us a capability to recognize even internal misuse based upon the analysis of network traffic flows. This technology is not based upon "signatures" which must be updated regularly in order to detect an attack or other attempted compromise. Instead, it uses a heuristic algorithm, which "learns" the "normal" or baseline behavior of a network or enterprise and watches for any variances in this baseline. One potential problem with NBAs is that if a network is already compromised and this compromise has gone undetected prior to the installation of the device, the malicious activity could be characterized as "normal" and thus no alert would occur because the malicious behavior was incorporated into the "baseline" as "normal" behavior.

While NBAs are a step in the right direction (away from signature-based security systems), they still have the same essential shortcoming as ALL other legacy or existing network situational awareness tools: They can only monitor and understand the behavior exhibited by the devices on a network. More specifically, current state of the art SEIMs and correlation engines only look at a network and alert on things that (1) have already happened based upon known behavior of DEVICES or (2) establish a norm and identify significant deviations from that norm. However, in the case of the latter, "non-normal" does not equate to malicious intent, and typical anomaly detection approaches often result in inordinately high false positive rates.

The key analytical tool incorporated into the situational awareness solution is based not on network behavior

analysis but automated human behavior analysis. It is perhaps helpful to discuss the true benefit of human behavior analysis in these terms: "We are being attacked, yes but not by network devices. We are being attacked by humans using network devices." The key is to look for patterns of human behavior behind the attacks. To do this the Leidos R&D team used a combination of technology and methodology to automate human behavior analysis so that potential threats may be characterized and alerted upon even before they materialize as attacks. Because the Leidos technology examines the antecedents of behavior, this approach allows network operators and information providers such a US CERT (for the first time) to discover threats and events that can affect a network or group of networks, pre zero day.

Because of this transition from one of reaction to being proactive, network and enterprise cyber operators have time to prioritize mitigation activities because they know what could be on the horizon.

Putting the Picture Together

This predictive situational awareness capability was developed as an architecture into which multiple, disparate technologies for ingestion, storage and "big data" analytics may be used. With development efforts the reference architecture could easily accommodate almost any analytical tool of choice.

The figure below is a high level representation of the various components integrated to provide predictive situational awareness at end state. This particular architecture was developed with a DHS TIC and associated Einstein deployment scenario in mind. It incorporates both existing sensor technologies as well as a new Einstein V3 equipment.

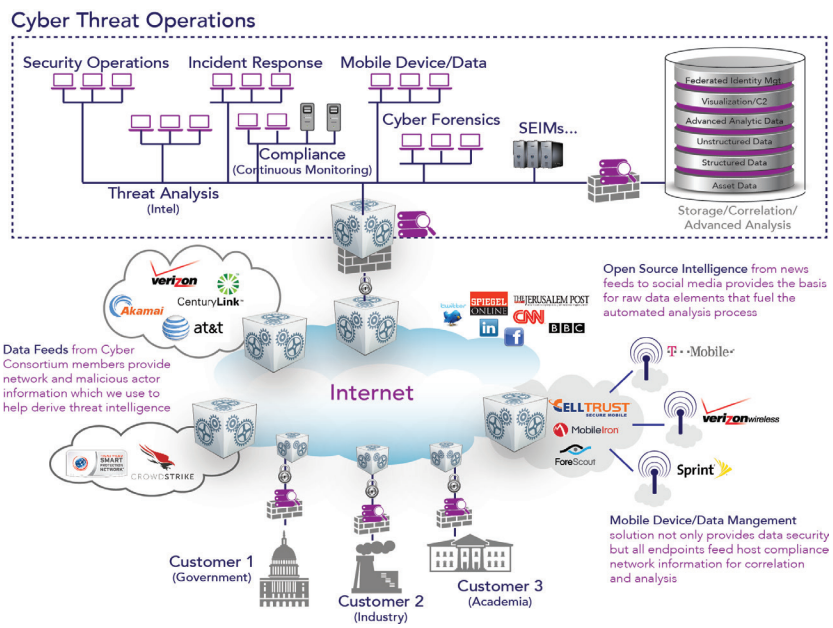

Figure 4: Raptor Predictive Cyber Situational Awareness Solution. This drawing illustrates the Leidos "Raptor" predictive situational awareness solution in the context of an existing government cyber environment. This reference architecture was designed with a DHS environment in mind. Structured data elements are ingested from sources external to the network and correlated, analyzed and visualized in order to provide a forward-looking threat operations picture to cyber decision makers. -

Creating a New Capability

Developing the C3I capability for true threat operations was only the first step in evolving beyond the failed network and signature based cyber defense approaches available to us today. We still need eyes and ears in the form of sensors for both structured and unstructured data elements. But sensors are not the answer. What is necessary is not "just another box on the network" but an entirely new approach.

Throwing Away the Book

From the start of the "NextGen constellation" effort, the first requirement has been: "scrap everything that looks like it is available today and let's design this as if we're starting from scratch." Then we began with some other basic requirements. The NextGen constellation must:

a. Be deployed as either hardware or software,

b. Communicate among multiple, disparate instances of the nodes themselves,

c. Provide not only sensing but active defense capabilities,

d. Self organize – collaborate in data sharing as well as active defense,

e. Be self protecting,

f. Be software defined,

g. Be informed - but not limited by a threat-based communications, command, control and intelligence (C3I) solution such as the Raptor described above.

Finally there are two overriding, hard and fast prohibitions:

a. The NextGen constellation must have NO reliance upon any legacy thinking PARTICULARLY signatures.

b. The capability must not be limited to any type of physical hardware. That is, it must be capable of functioning at either a network or host level, and may be virtualized. Whatever forms a node may take; it must still possess all the characteristics defined in the requirements above.

With these capabilities, the "sensor" now evolves from a legacy platform that lets the SOC know about something that has already happened into an Active Cyber Node (ACN) which not only informs but can actively work in conjunction with other deployed nodes to limit bandwidth, close or re-route data flows, manipulate data, or if the legal charter exists, provide a means for more aggressive defensive measures as appropriate.

This new, forward-deployed, node capability would be comprised of multiple types of agents:

- Controller – Deployed on multiple hosts, servers and networking equipment. This orchestrates the activities of some number of its "minion" agents. This agent also orchestrates the P2P network among the constellation nodes.
- Sonar – Performs the most sensor-like of all agents as it monitors the network for malicious activity on the networking infrastructure.

- Pshrink - Human behavior analysis updater. This agent coordinates with Controller for any updates on emerging threats. Enables the Sonar with new self-organizing maps (dynamic neural nets) to looks for emerging and existing malicious activity patterns.
- Sniper – This node embodies an "active defense" capability. This agent is not an "attacker" but is designed to work with the Controller and the Sonar to formulate a response to a malicious activity. This could take the form of:
  - Patching a system in conjunction with the "Compliance" portion of the Raptor.
  - Changing encryption keys – the Controller can work with cryptographic infrastructures like PKI and key repositories to change potentially compromised cryptographic keys. Obfuscates data from attackers while synchronizing the friendly infrastructure.
  - Changes individual or groups of packets to thwart malicious activity. Upon detection of an in-progress or pending exfiltration, perhaps we change the effected packet headers in the data.

Because the constellation nodes will constantly share information among themselves as well as the Raptor, they will be able to begin acting upon threats as they emerge in an automated way. This provides the ability to act many times faster that humans alone could possibly perform.

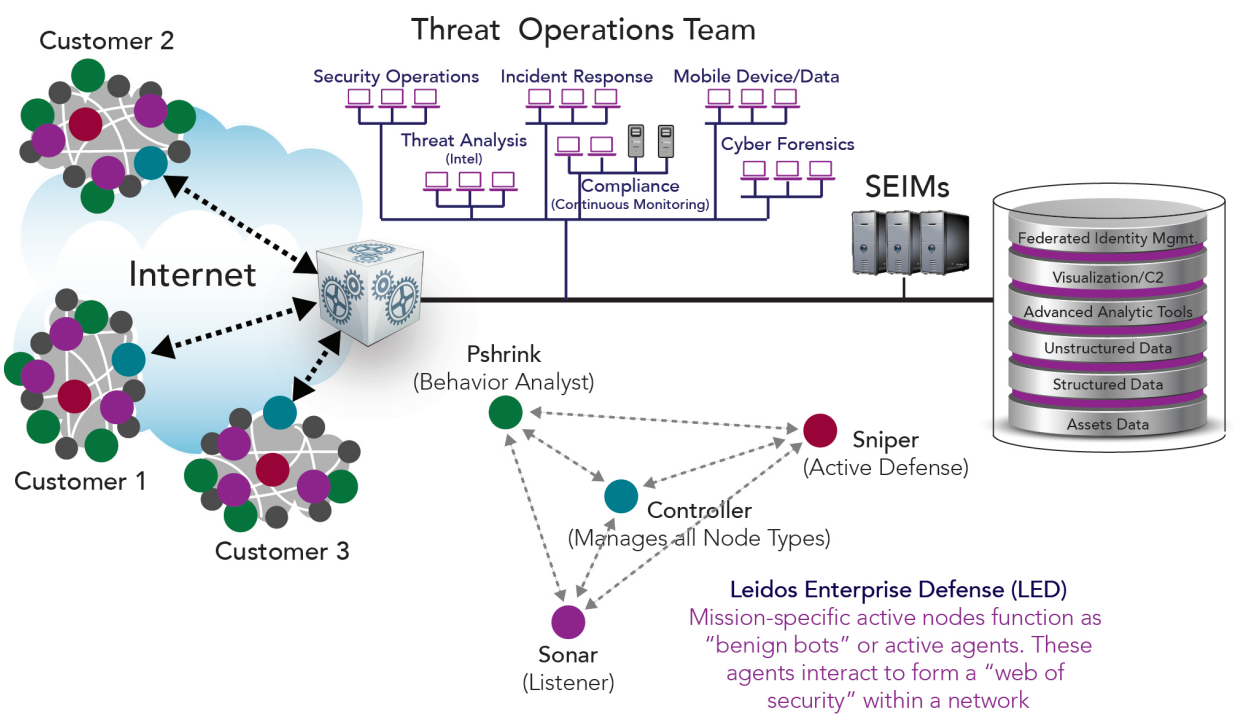

Figure 5: Next Generation Enterprise Defense. In this drawing, the various active nodes are are depicted in various colors as indicated above to emphasize their roles. Note the interaction among the nodes, each with its own functionality – all orchestrated by the controller. In this scenario, all data elements collected by the LED agents is forwarded back to the Raptor for analysis along with other network and external feeds to further enhance intelligence-driven, risk-based, predictive situational awareness for all customers. This technology is deployed in conjunction with legacy technology (IDS, Firewalls, DPI engines, etc.) as these are EOL they would be removed but no initial "fork lift" of customer investment is required.