empow

You have it in you.

Network Traffic Analytics

White Paper

Introduction

One of the main challenges security IT teams need to address is the advanced insider threat. Insider threat campaigns find ways to compromise internal hosts, hosts that are either remotely controlled by the attackers, used directly by attackers who were able to get user or administration credentials or simply both combined. This type of attack presents a major challenge for security IT teams because existing security solutions suffer from the following drawbacks:

Lack of Visibility

Many security solutions, such as NGFW's, IDS and IPS, anti-malware, reputation services, etc., were designed and deployed to detect and prevent network based attacks by monitoring traffic between internal hosts and external sites (north/south) and not internal organization traffic (east/west). As a result, the organization's east-west traffic (i.e. internal traffic), is left unmonitored.

Host-Based Security Solutions Can Be Compromised

Host-based security solutions such as anti-virus, Firewall and anti-malware, aim to bridge this gap and detect internal threats. However, one of the drawbacks of host based security solutions is that once the host is compromised, the attacker can control the security configuration, change policy, and deactivate the functionality of these solutions.

Zero-Day and Unusual Network Behaviors

Insider threat campaigns consist of behaviors that can be considered unusual but don't present any network policy violation or known network attack signatures. These behaviors will go unnoticed by solutions that don't include adaptive Network Behavior Analysis (NTA) capabilities.

NTA Solutions Typically Produce Too Many False Positives

NTA technologies, which have existed for at least 10 years, were expected to solve many of the existing issues. In reality network NTA technologies were not widely adopted because of the high level of false positives that typifies them.

This paper introduces empow's NTA module and the underlying technology which makes it unique in its accuracy, visibility and scale. It also describes how the service is integrated into empow's i-SIEM solution and works hand-in-hand with the embedded risk-chain analytics algorithm to effectivity address the aforementioned drawbacks, resulting in accurate detection of advanced attack campaigns.

1. Solutions such as NGFW's and IPS, must be deployed in line, so that all the traffic must bypass these devices. Covering all the internal traffic requires a lot of processing power, and therefore these devices will simply not scale to process all of the organization's traffic.

NTA Architecture

empow's NTA consists of a set of DPI software engines that are connected to tap devices or copy ports in the underlying network infrastructure, and monitor both east-west and north-south traffic flows. These engines process all the monitored connections and send digested information to empow's NTA service for further processing and analysis.

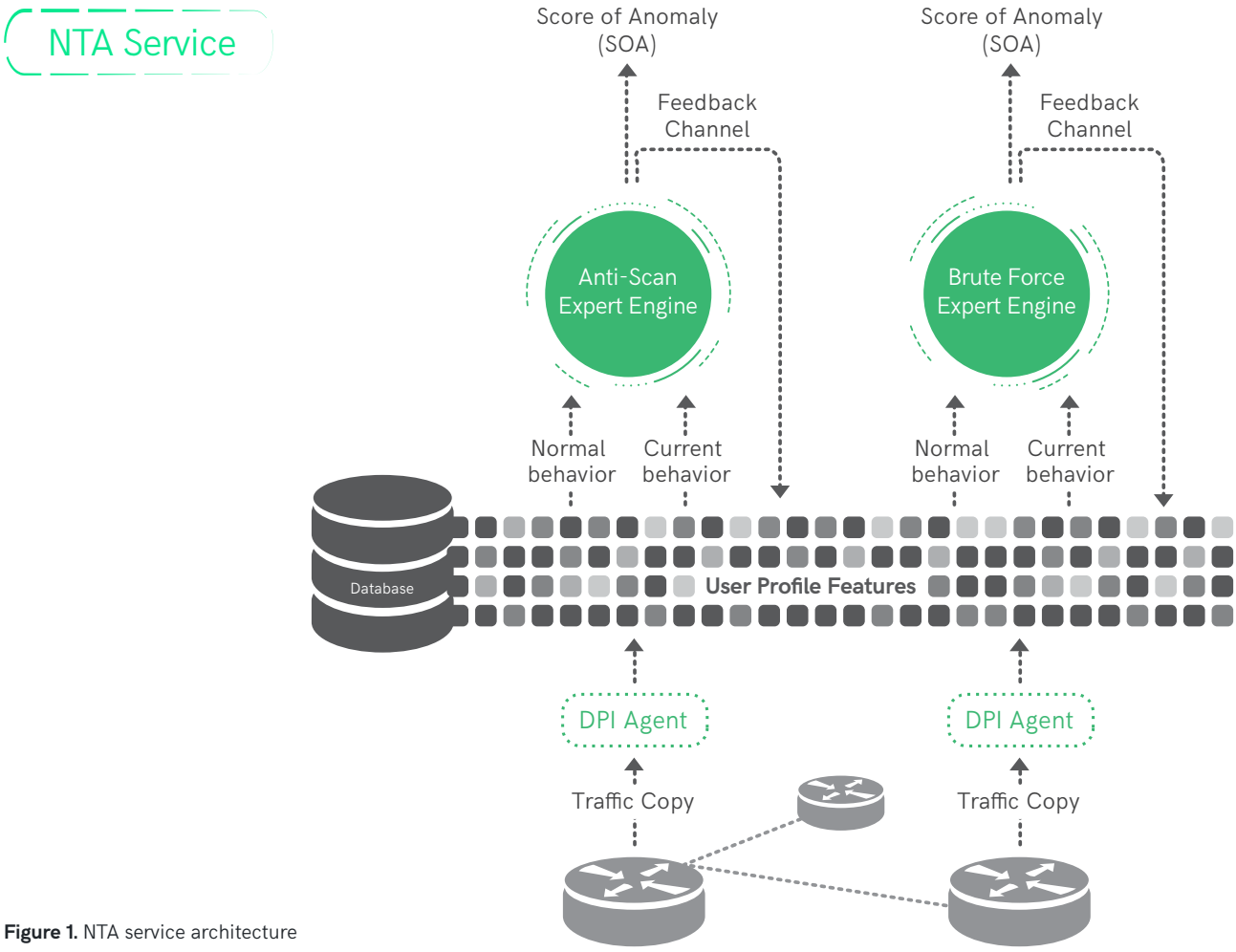


Figure 1. NTA service architecture

The NTA module stores connection information sent from the DPI engines in a user profile which has a hierarchical layered data structure. Thus, one can easily extract traffic behavior parameters per host, protocol and per application metadata in order to analyze and detect anomalies.

Analyzing the data stored in the user profile is done by a set of independent adaptive behavioral expert engines that were designed to emulate the human security expert's decision-making process. Based on network behavior characteristics extracted from the user profile, each expert engine is designed to identify traffic anomalies, which can be associated with a threat category. The engines generate a score of anomaly (SoA) that represents the level of threat that each host and each of its flows are associated with – a low SoA means no threat, while a high SoA represents a threat with high confidence.

DPI engine

In order to analyze network behavior, a set of DPI engines are deployed across the physical infrastructure. These engines receive a copy of all internal and external data traversing the network, and process it using an application level DPI engine that can detect network traffic characteristics.

These characteristics include: IP addresses, transmission protocols, L4 port number, as well as L7 protocols, applications, and the metadata associated with them. URLs, host names, operating systems and more are some examples of the L7 metadata. Each DPI engine maintains a data structure for all the connections monitored by the DPI probe itself. The data consists of IP addresses, IP protocols, ports, the state of the connection, statistical information such as inbound and outbound packets and bytes, connection duration and information regarding the upper layer protocols and applications detected by the DPI.

The DPI engine even holds application information such as type of application queries, protocol return codes, commands types, etc. Each DPI engine sends a periodic report containing up to date digested information for all tracked connections to the user profile within the NTA service.

The NTA module architecture enables multiple DPI engines to be deployed in parallel, working independently, supporting large scale networks (in terms of number of hosts, bandwidth and connections), supporting different network monitoring technologies. For instance, using RSPAN, or packet broker technologies, a single DPI or DPI cluster can be deployed to monitor the entire network. While using SPAN technology, one may deploy a DPI engine per endpoint switch. In case multiple DPI engines are deployed, it may be the case that some connections will be monitored by more than one engine. In such cases, the user profile (described in detail in the following section) is responsible for filtering such flow duplications.

While traditional L3-L4 flow analysis technologies may provide quite a lot of information regarding network behavior, the fact that empow's NTA is based on L7 application DPI, provides our customers with a much more sophisticated view into network behavior, and a more accurate analysis.

As an overall solution, our security platform reduces false positive alerts, increases threat detection which would normally stay under the radar of traditional L3-L4 analysis technologies.

In particular, while in the past each L7 protocol was used for one or very few applications, today it has become common for multiple different applications to utilize the same L7 protocol. The most obvious example is HTTP which is used by different applications such as Gmail, LinkedIn, Facebook, video / voice and many others. Each application presents a different behavior and therefore should be "baselined" separately in order to identify anomalies. In addition, L7 application layer analysis enables the DPI engine to provide complementary information regarding the state and the status of an application, such as the success or failure return code of a request (e.g. upon performing an FTP get command), the type of request (e.g. is the HTTP of a type GET or POST, DNS query type), etc.

User profile

empow's NTA host profile is a set of persistent data structures maintaining the baseline network behavior of the hosts monitored.

Baseline behavior patterns are characterized by a set of traffic characteristics such as the number of connections per minute, connection duration and size, packets per minute, traffic symmetry, connections success ratio, host and protocol distribution, etc. Each one of these characteristics can be used by one or more NTA expert engines, comparing the baseline behavior, behavior learned so far, against the current network traffic patterns. Deviation levels experienced between the current and the baseline behavior for different groups is used to detect security threats (more details on this in the expert engine section).

The baseline behavior patterns are learnt by monitoring connections collected by the set of DPI engines, while comparing both past and present values indicates if a shift in behavior exists. The combination of both sets of values are used constantly, calculating and updating the current expected behavior. In order to avoid malicious behavior becoming the baseline for a user profile, a feedback mechanism loop exists between expert engines (EE) and the user profile. Thus, once an EE detects abnormal behavior, the set of flow characteristics used by this EE to identify the anomaly is no longer used to alter the baseline, not until the EE indicates that the abnormal behavior no longer exists.

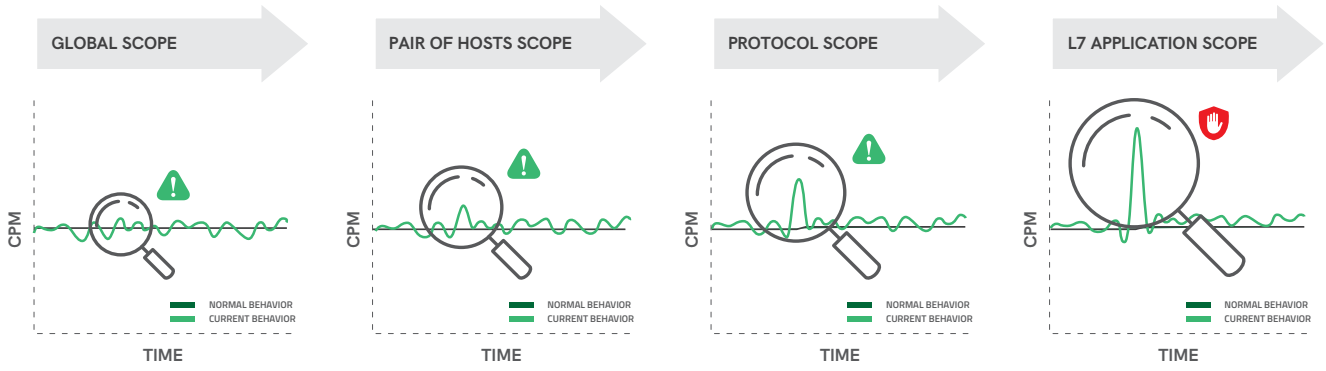


Figure2. While it is may be difficult to detect deviation in the normal connections per minute (CPM) feature in the host global scope, as the resolution is increased, the deviation becomes more and more clear.

Utilizing DPI capabilities to identify L7 protocols and applications, the user profile has a layered structure where the baseline behavior for each flow is maintained. The extent of layering can be extracted in different scopes, starting with the host global behavior (i.e. between the host and the rest of the network), through to the behavior between a pair of hosts, even down to specific protocol and application flows. This capability enables the expert engines to examine deviations between different scopes and thus reduces noises and false positives, while improving detection sensitivity. For instance, consider a host that has frequently used a specific set of applications over HTTP. Malicious activity using a different application over the same port may not be detected, since the protocol flow deviation for the valid HTTP traffic may be negligible when compared to the malicious traffic. On the other hand, considering per application baseline behavior and/or per destination baseline behavior, the deviation of the malicious activity becomes more significant and its chance of being detected increases.

The user profile is designed to support large scale organizations and networks consisting of tens of thousands of users, as well as supporting many anomaly expert engines of different types that extract and analyze multiple traffic behavior characteristics (defined as user profile features).

Expert engine

While comparing current network behavior, to create an average baseline, excessive amounts of false positive alerts can easily be witnessed, especially in cases where the network rates are unpredictable. External factors such as social events can cause unusual ripples in traffic and activity. In order to be able to accurately detect security threats based on network anomalies, empow's NTA module implements a series of multi-vector expert engine algorithms, where each expert engine is designed to emulate the human security expert in detecting a security threat category.

This kind of approach examines multiple traffic features at once, assigning each an anomaly weight. Once each feature has been assigned a weight, correlation is performed between the weights so that the decision made by the expert engine is balanced, taking into account both rate sensitivity as well as rate invariant features, minimizing the number of false positive alerts. Figure 3 illustrates the operation of the expert engine:

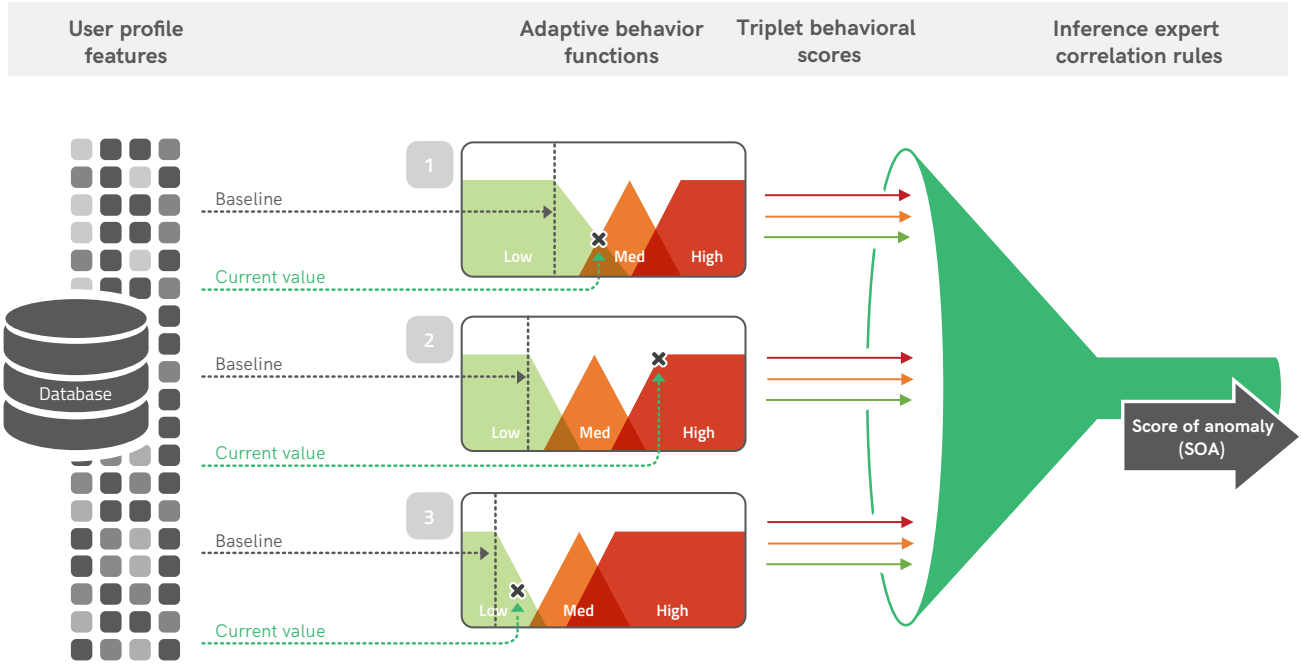


Figure3. Security expert engine

The NTA architecture consists of an adaptive behavior functions layer that receives both learned (baseline) and current feature values from the user profile, and transforms them into a unified set of behavioral parameters that the expert engine can understand and therefore process by its correlation rules, regardless of the type and units that represent each feature in the user profile.

Each feature's baseline and current value are processed by the relevant adaptive behavior functions, which are continuously shaped according to the learned baselines patterns of the hosts inside the network. These adaptive functions determine the level of membership of each feature value in three behavioral groups: low, medium and high anomaly membership groups. The level of membership is represented by a score, which results in an anomaly score triplet of each feature (high, medium and low score). The triplet score vectors, which represent the features anomaly levels, are processed by the expert engine's correlation rules, which determines the overall Score of Anomaly (SoA).

As mentioned before, depending on the expert engine "security role", each generated SOA can represent an anomaly associated with a group of hosts, a single host, and/or with a specific application flow.

Consider, for instance a Brute Force expert engine that is designed to detect brute force attacks, for example, user/pass cracking attempts and application vulnerability scans. A brute force attack can be characterized by a large number of relatively short connections, most of which fail. Thus, the engine takes into account three independent traffic features: the number of connections per second, the connection size (in terms of bytes), and the connections success ratio, that is the percentage of successful connections of a user accessing a specific application. Since the first feature (i.e. connections per second) is rate sensitive, relying on this feature alone would lead to a higher percentage of false positives. Thus, the engine also considers the size of the connections and the success rate, which are typically insensitive to network traffic rates, i.e., rate-invariant features. The expert rules of the EE correlate the anomaly weights of all three features, i.e. too many short connections with a relatively high failure ratio to a specific application, and generates an SoA that indicates if the user is engaging in a brute force attack or not.

Drop Zone Expert Engine

Guarding sensitive information, such as credit card numbers, personal medical files, etc., is one of the main challenges security officers face. This kind of data attracts many attackers, who are willing to perform very complex and persistent attacks in order to access the relevant data. Once an attacker has access to the server farm maintaining such data he still has to extract the data and send it to an untrusted environment. Typically, deployed security policies prevent these servers from having direct access to an unsecured environment, assuming that such a policy will prevent potential data leaks. In order to overcome this restriction, attackers try to release the data using intermediate hosts, called internal drop zones. Thus, instead of sending the data directly, the data is sent to one or more hosts that have both access to the sensitive data server as well as to an unsecured environment.

The Drop Zone expert engine is a unique engine which aims to detect drop-point or drop-zone activities in the network as described above. The engine detects the drop zone network entities as well as the source communicating with these entities in an abnormal manner, which represent a possible data leak.

The drop zone expert algorithm has three phases and it takes into account three traffic characteristic (features). In the first phase, a set of potential drop zone servers are identified by examining anomalies associated with the connection distribution, flows destined toward a particular server or groups of servers inside the organization (in terms of the number of hosts that access each server). In the second phase, for each potential drop zone server, the expert algorithm examines anomalies associated with L7 protocols or applications distribution. Identified anomaly protocols are marked as possible channels, that the malware utilizes, for dropping the data.

2. The indication of whether a connection succeeded or failed, is obtained by the DPI engines, capable of performing L7 parsing and analyzing per protocols and per applications metadata and transactions.

In the third phase, all internal hosts that access the potential drop zone servers with an abnormal upload activity (identified based on a traffic symmetry feature) through these malware channels are marked as hosts that drop the data.

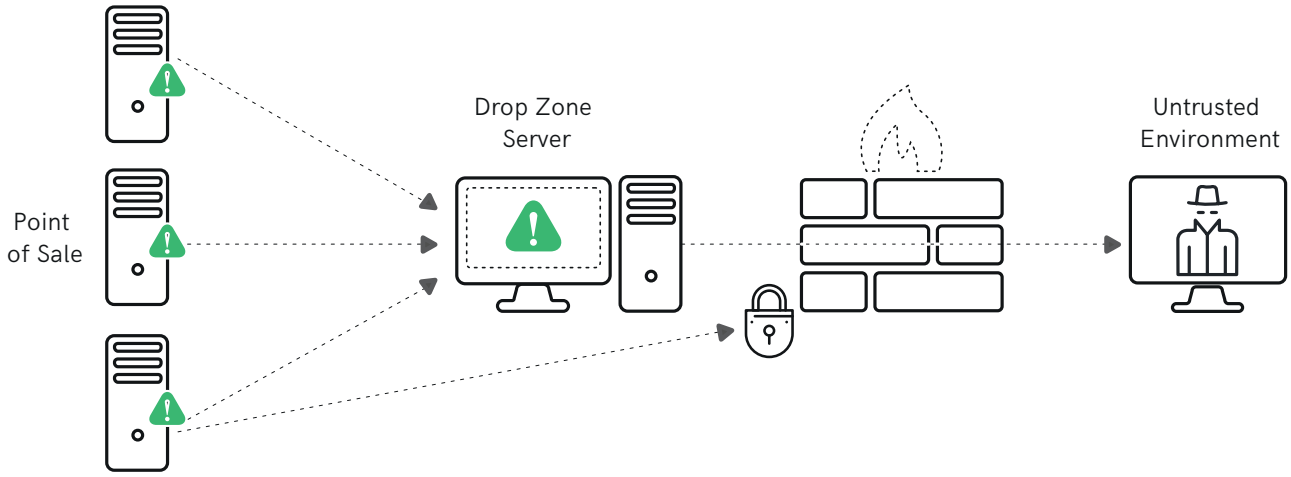


Figure4. Drop zone network activity

Anomaly Behavior Expert Engines

Each expert engine in the NTA module is programmed with different correlation rules, processing different sets of traffic features in order to identify different categories of threats. The following expert engines are supported:

Anti Scan

An expert engine that detects network pre-attack probes (network scanning) activities, including vertical, horizontal, IP sweeps, stealth scans & others

Vulnerability Scanning

An expert engine that detects potential vulnerability scanning over HTTP.

Password Guessing

An expert engine that detects manual attempts at authorizing brute-forcing through manual password guessing.

Malicious Mapping

An expert engine that detects malicious mapping of internal organization network assets, which can be an automated attack.

Brute Force

An expert engine that detects brute-force attacks including user/pass cracking attempts and application vulnerability scans.

Abnormal Data Transfer

An expert engine that detects suspicious data transfers that can be associated with data leak (download or upload of data) activities in the network.

New Flow Anomaly

An expert engine that detects hosts connecting to internal host using L7 protocol(s), which they haven't used in the past, indicating abnormal behavior.

Programming New Expert Engines

The way the NTA service was designed allows to "program" new adaptive expert engines that can identify entirely new types of anomaly behaviors in the network, fast!

The adaptive behavior functions normalize the traffic features into a form that human expert behavioral rules can process (the inference correlation expert rules). As a result, new expert engines can be created by empow's professional services team- an implementation process that takes days for empow, as opposed to a process that can take weeks or months by other NTA security vendors.

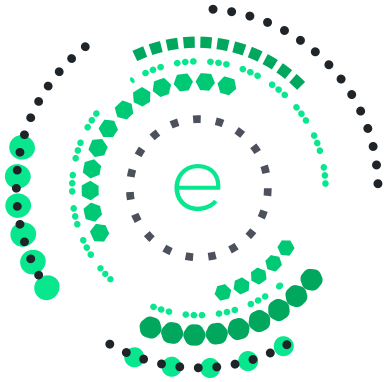
Integration with empow's i-SIEM

empow's intent-based SIEM, or i-SIEM, is the only next-generation SIEM that is actually able to decipher attacker intent.

The platform uses advanced AI and NLP algorithms to detect cyber attacks and automatically orchestrate adaptive investigation and mitigation actions in real time, without the need for human-written rules.

empow's strategic OEM partnership with Elastic makes empow's i- SIEM the optimal security solution for Elastic users. The i-SIEM seamlessly integrates with Elastic's data lake to provide a constantly updated loop of detection, investigation and response.

It's a whole new idea in cyber security -



Turning What You Have Into What You Need.

While one of the main advantages of our platform is the ability to manage security resources and correlate security events for multiple security products (e.g. IDS, reputation service, NTA, Anti-malware, Firewalls etc.), there is great value in using the i-SIEM with empow's NTA module.

In this case, and in addition to the traffic feature correlation inside the NTA's expert engine level, the platform automatically correlates between events detected by the different NTA expert engines. This allows the overall solution to detect more complex and staged attack campaigns. It also reduces the amount of false positive alerts that typify NTA systems which rely solely on a single NTA expert engine.

For instance, a typical security attack may consist of an intelligence gathering phase, in which the attacker scans the network, followed by propagation, resulting in a data leak. In this case, the i-SIEM's risk-chain discovery technology may request to detect scan activities in the network using the NTA's anti-scan expert. Once such an event has been detected, the brute force expert and the malicious commands expert will be used to detect malware propagation from the scanned or scanner host to other hosts. In the next stage, the risk-chain discovery module will direct the data leak experts to analyze if an abnormal data transfer or drop zone activities exist from and/or towards the possible infected hosts (those that are associated with possible malware propagation events).

The below illustration shows the progress of an advanced attack campaign and how the risk-chain discovery module assigns the relevant expert engines for each attack stage and then correlates their events in order to accurately identify the threats:

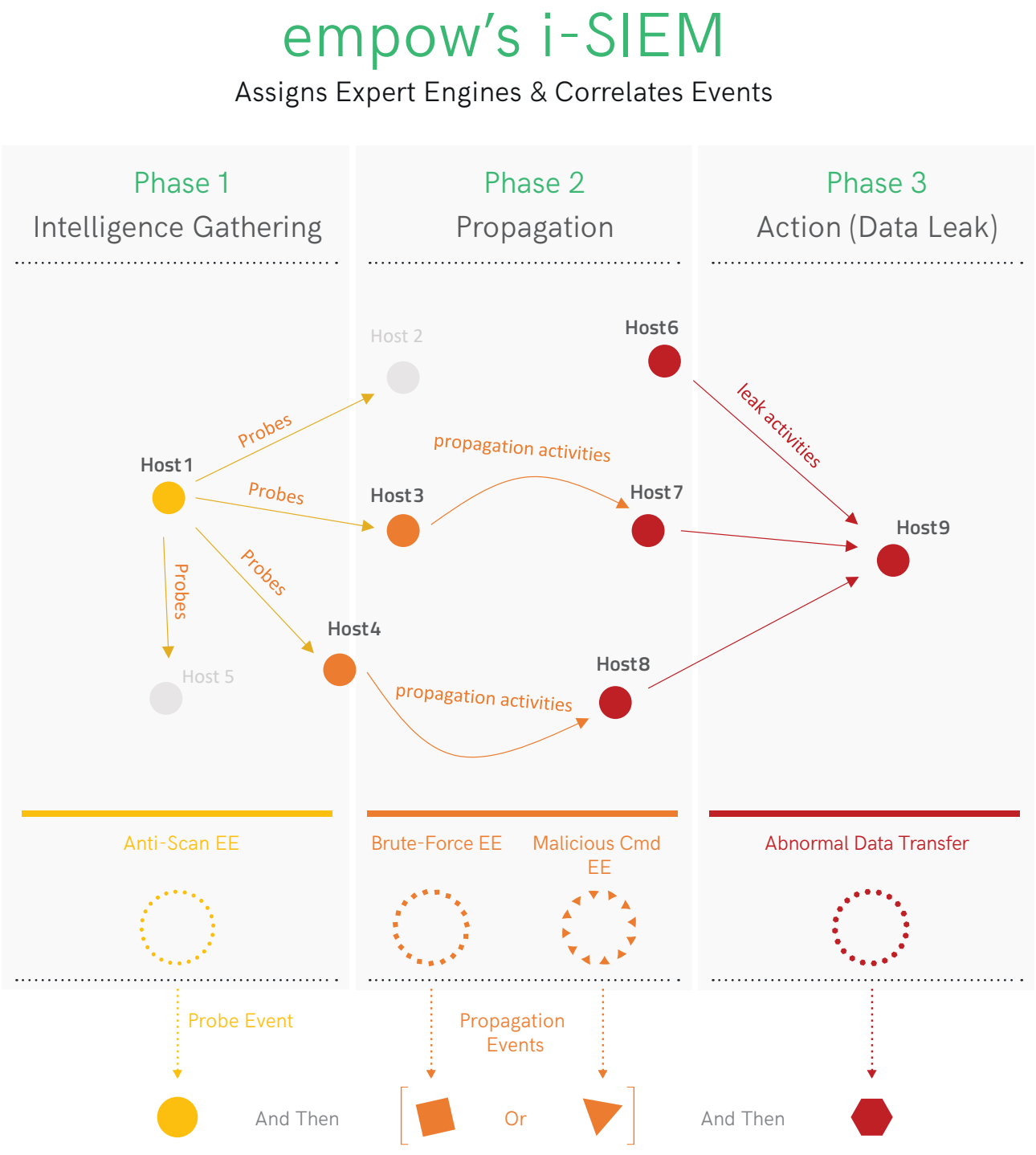
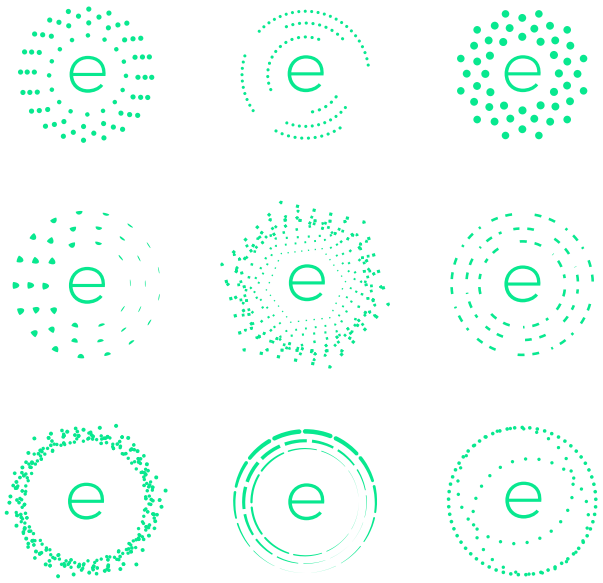


Figure 5. Automatic correlation between events detected by the different NTA expert engines

Summary & Main Values

empow's NTA module presents a new approach for the detection of traffic anomalies that are associated with "insider" attack campaigns. Attack activities such as intelligence gathering, brute force activities, manual and automatic malware propagation and data-leak are all covered by the service's security expert engines.

The NTA module's multi-vector expert engines, together with the ability to correlate all of the engines' outputs, results in a system that effectively identifies advanced persistent attack campaigns. The main value and advantages of the NTA service as part of empow's i-SIEM include:



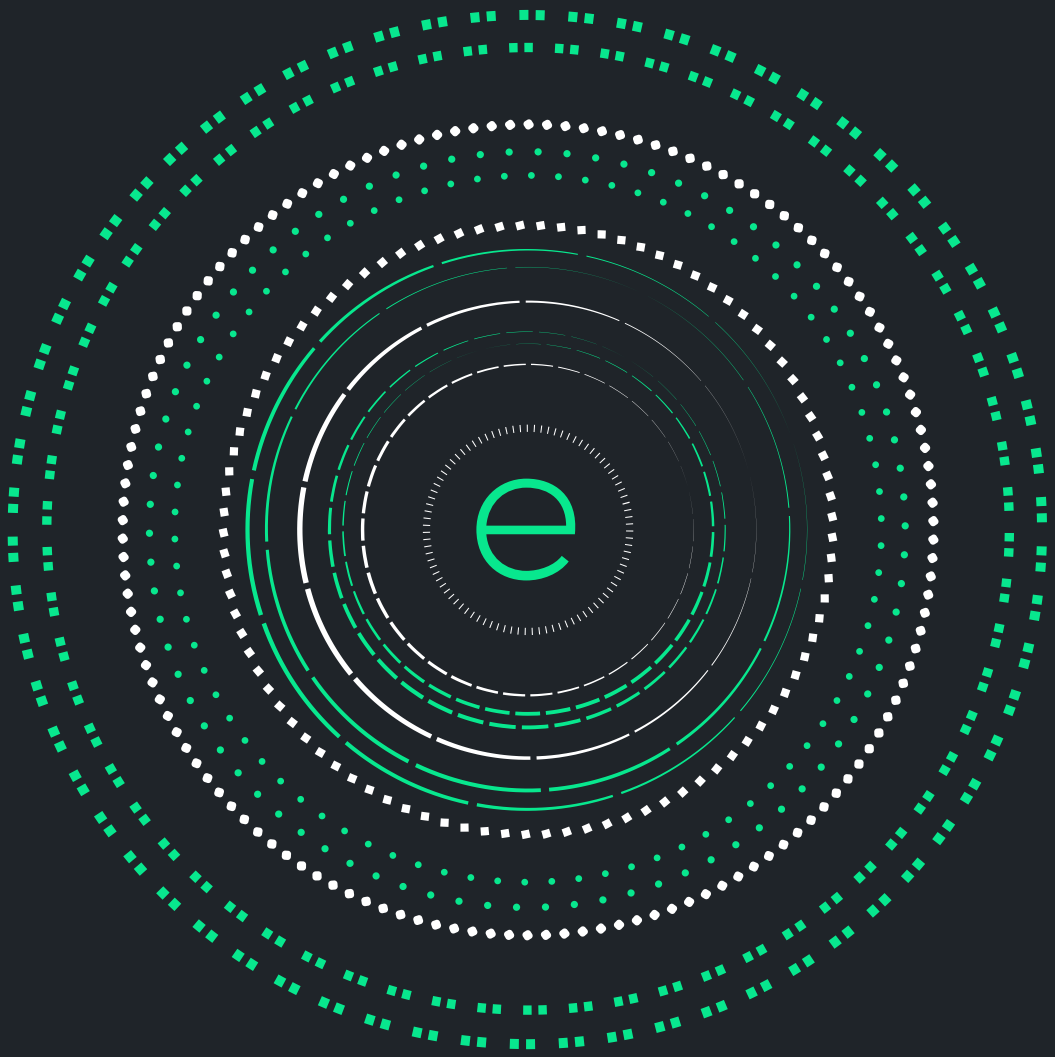
Identifies advanced attack campaigns in real-time.

Does not require any configuration and rules-set maintenance.

Introduces virtually zero false-positive decisions by:

- Expert engines that process both rate and rate-invariant traffic features.
- It is never "the same system twice" - adaptive down to the behavior of the host's application flow levels.
- Generates final context-based decisions by correlating the individual engines' outputs through empow's risk-chain discovery algorithms.

Real-time and accurate attack mitigation by integrating with the mitigation services embedded in empow's i-SIEM.



empow
You have it in you.

Tel: +1-877-647-4361
129 Newbury Street, 2nd Floor
Boston, MA 02116 USA

Tel: +972-3-519-5517
info@empow.co
Hayetzira 29, Ramat Gan, Israel 5252171

www.empow.co