ManageEngine
Powering IT ahead

High-Precision-And-Recall Network Anomaly Detection using Continuous Stream Processing

Some advanced strategies for building effective and reliable Network Behavior Analysis systems

Chandramouli Srinivasan
Technical Architect
ManageEngine
chandramoulis@manageengine.com

Table of Contents

Executive Summary

With respect to cyber crimes, in the recent times, the risks faced by enterprises have multiplied. There's been a clear shift from hacking for fame and thrill to a focus on data stealing, identity forging and extortion attacks, as well as controlling a large pool of internet resources for achieving those ends. Continuous evolution of intrusion techniques has made the task of ensuring network security increasingly difficult in spite of becoming all the more critical.

Building comprehensive network security infrastructure involves at least three very important dimensions. They are (i) Firewall Systems, (ii) Intrusion Detection/Prevention (IDS/IPS) Systems and (iii) Network Behavior Analysis (NBA) Systems also known as Network Behavior Anomaly Detection (NBAD) Systems. While all three of them have their own unique strengths and weaknesses, they complement each other to form a holistic network security strategy. However, the first two are widely prevalent and perceived as essential components, the third is not so. This leaves the network vulnerable to several zero-day attacks, unknown worms, internal threats, etc., as well as letting them lag behind in terms of overall traffic visibility, access policy decisions, security posture assessment and a reasonably sure confirmation of network security.

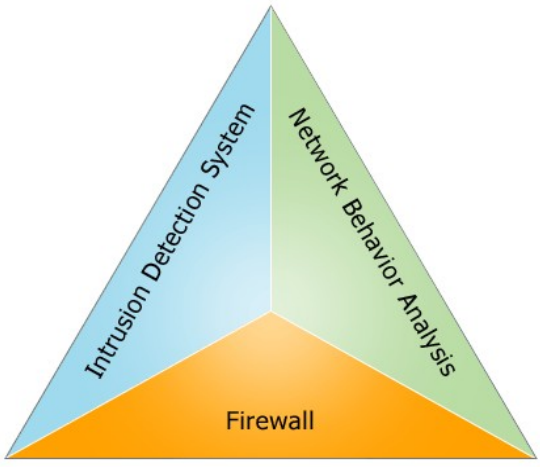
Understandably, the prevailing reluctance or uncertainty towards deploying present day NBA systems is not without reasons. Some of these are alert proliferation, alert repetition, high false-positives, and a general inability to capture low foot-print intrusion activities. In other words, serious lacunae in terms of precision & recall issues impact the reliability of these alerts. However these can be overcome largely by infusing some of the latest technology developments in the field of high-throughput-low-latency continuous stream processing and high-precision-and-recall complex event correlation, and above all; some effective profiling and advanced single-pass data mining algorithms.

ManageEngine NetFlow Analyzer's Advanced Security Analytics Module (ASAM), a network flow based NBA tool for security analytics, helps detect & classify zero-day network intrusions real-time, using the state-of-the-art *Continuous Stream Mining Engine*™ technology. ASAM offers actionable intelligence to detect a broad spectrum of external and internal security threats as well as continuous overall assessment of network security.

Introduction

With the network becoming the cornerstone of today's enterprise infrastructure, the utmost need of the hour (or the era) is to secure the same. In an attempt to make the applications and data more user-friendly and accessible, they have become highly vulnerable to threats. The dawn of network era has made life easier for the users as well as hackers. Enterprises cannot afford to have server downtime, data theft, Denial of Service (DoS) attacks, to name a few perils of network breach. Enterprises have come to rely on the networks to such an extent that the productivity of enterprises has come to depend on the uptime of their network. Though the number of benefits outweighs the harm caused, the magnitude of the harm caused is very different. The magnitude and the aftermath of an attack depend on the time of attack, the target of the attack and the time taken to restore the systems.

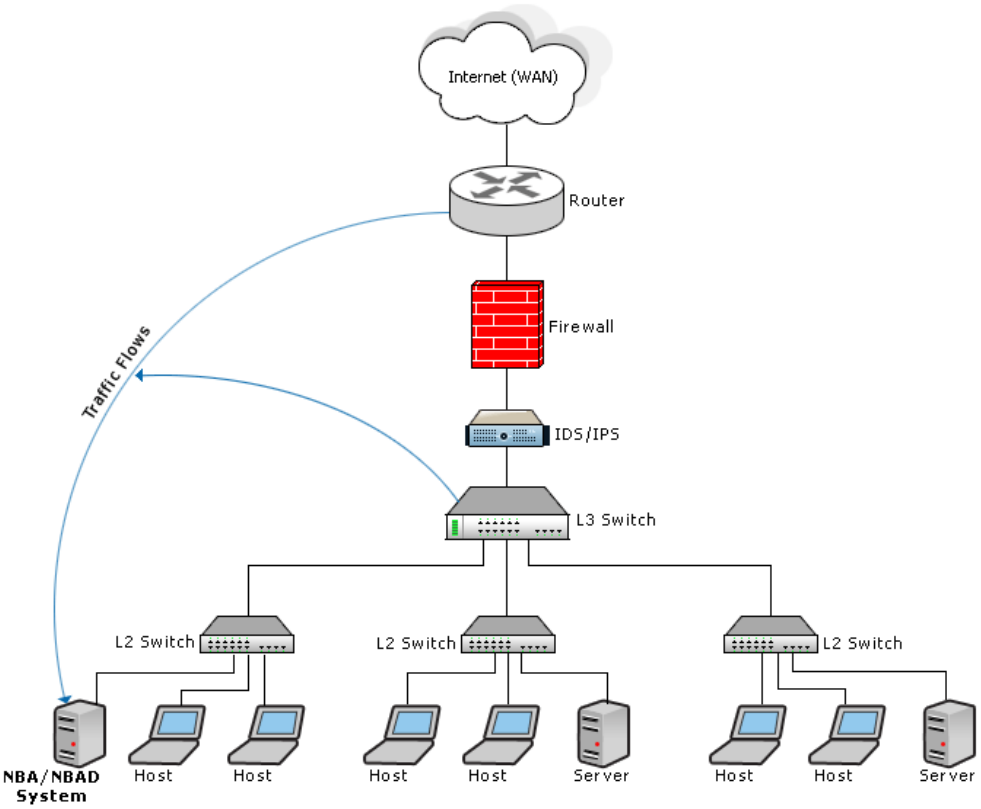
These threats can be broadly classified as threats from within (internal) or outside (external) the enterprise. The external threats are attacks from outside the enterprise and can be waded off with a good perimeter security system. On the other hand, there is also a great deal of threat that can be caused by the user within the enterprise. Misuse of permission or hacking from within the organization to steal data can be one such case. There are also possibilities of a device (eg. laptop), taken out, getting attacked and the threat getting carried over inside the organization. Once inside the enterprise, neighboring systems are infected leading to severe damages to the enterprise network. With evolving threats, the race for better threat detection system is on. The below diagram gives an overview of the comprehensive network security:



Comprehensive Enterprise Network Security

Threat Detection Systems (TDS)

The threat detection systems are many, but almost all of them fall short of achieving the true purpose of the system. Ideally, security systems have to evolve as and when the threats evolve, but in the real world not all systems are such. The ability of the TDS to understand the threat and evolve is the differentiating factor between the good, bad and ugly of the TDS world. While some of the TDS monitor the network periphery - shallow and broad - the others look narrow and deep, thus losing out on the holistic picture of the network. The other problem these systems face is to "play it safe", they churn out too many false-positives, thus diluting the real problem. Rather than helping the network manager to quickly obtain an overall assessment of network security, they end up making the task harder for the network manager. Let us look at a network, where the various TDS are placed and then delve into the each of these threat d detection systems in detail.



Typical Network Security System Deployments

Of the three threat detection systems, the most prevalent security systems in the current enterprise are the firewalls and the Intrusion Detection Systems (IDS). The below given table compares all the available threat detection systems:

Network Security System	Typical Input Data	Typical Methodology
Firewall System	Packet Header (up to Layer 3 & and selective Layer 4)	• Access Policy Enforcement • Simple Interaction Patterns
IDS / IPS	Packet Header & Payload (Deep Packet Inspection)	• Detailed Signature Matching • Simple Interaction Patterns
NBA / NBAD System	Traffic Flows (derived mainly from packet headers)	• Advanced Interaction Patterns & Sessionization • Statistical Analysis • Access & Traffic Policy Monitoring

Firewall - The first level of defense

Firewall is the primary component of network security and provides protection at the perimeter level. It can be compared to a combination of security guards, initial baggage screening and metal detector checks performed at the airport terminal. Firewall's main task is to ensure access policy control and it does not provide extensive threat detection capabilities due to the large amount of traffic handled. Once a threat gets past the firewall, it is largely free to harm the network.

Firewall's weaknesses in detecting sophisticated zero-day attacks, port-protocol hopping & tunneling applications, payload based signature threats, etc. need to be mitigated by various firewall helpers such as IDS/IPS and NBA systems. However, once a threat is identified and analyzed by any of these firewall helpers, network administrators can use the information to upgrade the firewall policy.

Another drawback of firewall is its lack of visibility to internal traffic. Though firewalls are an important part of a network security infrastructure, they cannot be solely trusted to provide comprehensive network security.

Intrusion detection/prevention systems (IDS/IPS) – The second level of defense

IDS/IPS is a second line of defense meant to complement firewalls by bringing in deep packet inspection based signature detection capabilities. It can be compared to a combination of frisking, detailed body imaging and luggage scanning checks performed, for a closer inspection, at the airport terminal.

IDS being primarily signature based has its own drawbacks. First, scanning each and every packet for several hundreds of signatures is very resource intensive and so has to be selective. Another significant pitfall of the IDS is its inability to identify the Zero-day threats in spite of the frequent signature updates.

Network Behavior Analysis (NBA) - The new security hawk

NBA is primarily a holistic decision support system meant for providing network traffic and security analysis. It complements the other two threat detection systems and can be compared to the overhead multi-point camera observation and vigilance at an airport terminal. NBA is capable of detecting several zero-day attacks and intrusions based on typical interaction patterns, as well as offers an overall network security assessment.

Some of the common drawbacks of NBA systems are high false-positives, alert proliferation, lack of actionable event information, and a general weakness in detecting low foot-print or slippery network activities. However, a good actionable NBA system can boost the overall traffic visibility, access policy decisions, security posture assessment and provide a reasonably sure confirmation of network security. Such a system can thereby impact the overall network security perception significantly.

Summary of prime merits and demerits of the available network security system:

Network Security System	Prime Merits	Prime Demerits
Firewall System	• Ideal inline real-time protection & access control • Great traffic visibility at the perimeter • Good for packet header level signature matching	• Blind to internal traffic • Vulnerable to zero-day intrusions • Vulnerable to port-protocol hopping & tunneling applications • Vulnerable to well-known port tunneling applications • Decentralized policy management
IDS / IPS	• Inline real-time intrusion detection and prevention • Great traffic visibility at nodal traffic points • Great for full packet signature matching and inspection	• Limited to signature based intrusion detection • Dependent on frequent signature upgrades • Vulnerable to zero-day intrusions • Vulnerable to pattern based intrusions like scans, DDoS, botnet, etc. • Decentralized signature patch management • Forced to inspect traffic selectively as it's highly resource intensive
NBA / NBAD System	• Centralized agentless data collection, analysis and management • Good visibility into both external and internal traffic • Great zero-day intrusion detection capabilities • Ideal for pattern based event correlation and advanced data mining • Optimal for holistic security assessment coupled with traffic analysis	• Blind to packet payload content and signatures • Vulnerable to port-protocol hopping & tunneling applications • Prone to high false-positives, alert proliferation & repetition • General inability to capture low foot-print intrusion activities • Inflexible data grouping and monitoring points for analysis

Understanding the problems of the present day NBA Systems

Issues such as lack of visibility to packet payload and vulnerability to port-protocol hopping & tunneling applications of the NBA System are because of the missing data in the traffic flows, at least in the earlier flow formats. In other words, these are not the core competencies of the NBA System, just like the other two security systems too have their own set of problems inherent to them. However, the rest of the precision-and-recall issues such as higher false-positives, alert proliferation, insensitivity to low foot-print intrusions, offline data processing, etc. are of greater concern as they impact larger issues such as usability, actionability and reliability of the NBA System as a whole. It's important to note that these issues are not so inherent to NBA approach as such, and are more implementational in nature. Before moving on to the solution part, here is a listing of some of the relevant problems, causing these precision-and-recall issues, prevalent in the present day NBA systems.

Offline SQL query based and/or multi-pass data processing

Storing flows in the database and processing for actionable events using a sequence of queries is inflexible, resource intensive, has high latency and is suitable only for simple pattern correlation procedures. This also means more administrative effort for managing the NBA System.

Inflexible Resource Segmentation & Grouping

Different types of events have different reference points for monitoring. For example Host Scans are better caught by Source IP or Network address, whereas, DDoS & Port Scans by Destination IP address. Similarly, distributed anomalous flows are better caught by Router IP or Input Interface, whereas, P2P applications by port or port group, etc. So uniform data grouping for all type of event is inappropriate and flawed

Lack of actionable information collation

Using database queries for computing events can limit the collated information to the query functions of the underlying database. Collating information such as unique lists of connections, IP addresses, ports etc. or other Link Analysis based information like Network Diameter are difficult with SQLs.

Arbitrarily Fixed Thresholds

Fixing the threshold limit of, say, Host Scans as 5 or 10 hosts is arbitrary, and keeps varying with time & environment. This can lead to either too many alerts or too few depending on the value and situation. Manually tuning thresholds and limiting the number of events to the optimal is difficult

Lack of actionable event classification heuristics

The information that needs to be collated for various types of events is largely similar. However, classifying an event for host scan, port scan, P2P or DDoS has to be based on different threshold limits

on different metrics. Simply reporting SYN Violations or 'ICMP Unreacheable' is not as actionable as classifying events as scan, P2P, DDoS, etc.

Insensitive to repetitive or prolonged data patterns

While it's important to continuously track repetitive or prolonged events, reporting them multiple times is annoying. The idea is to not overwhelm the user with the same information but constantly keep a track of them and report in a controlled phased manner.

Insensitive to temporally persistent data patterns

Normal threshold violations find it difficult to capture stealthy but periodic intrusive activities such as 'botnets control channels' and other malwares. Here too, distinguishing the valid (white-listed) periodic activities from the unexpected ones is equally important.

Insensitive to spatially distributed data patterns

Marginal but scattered intrusive activities such as slow scanning worms, P2P and other malware are difficult to capture by normal threshold violations. Again, distinguishing the valid (white-listed) distributed activities from the unexpected ones is equally important.

Some Advanced Strategies for building high-precision-and-recall NBA systems

Most of the above mentioned problems can be overcome largely by infusing some of the latest technology developments in the field of high-throughput-low-latency continuous stream processing and high-precision-and-recall complex event correlation, and above all; some effective profiling and advanced single-pass data mining algorithms. The principal idea is to bring all these advances together over a single platform, in a systematic layered manner, to build high performance event processing systems. Given below is a brief about data processing strategies that will help in building effective and reliable Anomaly Detection Systems.

Fast bulk-lookup rules matching

Rapid classification of input data is a crucial first step and is a CPU-intensive problem. However large number of rules can be effectively grouped by a few criteria fields they are based on. Rules Engines can leverage this aspect to perform both fast and bulk-lookup of rules matching the data.

Continuous event stream processing

Incremental in-memory data processing to detect significant event patterns in real-time. This involves a structured systematic data transformation approach similar to the database query processing with multiple queries being computed simultaneously over the given data.

Complex Event Correlation

Putting together disparate pieces of relevant smaller level events to construct bigger level events, which are of greater concern. This approach enhances actionability of the alerts and eliminates false-positives significantly.

Multi-granular context-sensitive resource modeling

The reference point of a network event depends on the type of the event. It could be Source Host, Destination Host, Application Port or Port Group, Router or Interface, End Point, Network Connection, and so on. Capability to aggregate and correlate events simultaneously across these different resource types, and apply appropriate thresholds and classification heuristics is essential.

Temporal clustering & advanced sessionization

Segmenting data into temporally sequenced set of events based on various types of windowing constructs like sliding or jumping windows, active or inactive timeouts, termination flags, request-response lag timeouts, maximum time span or event size, and other limiting criteria.

Event classification heuristics

Heuristics that can evaluate events belonging to different problem streams, and further classify them into actionable problems. In other words, a second level of data pattern is matched post event generation, based on various criteria and thresholds.

Automatic threshold adaptation

Manually fixing a minimum threshold limit is easier for network administrators based on their valuable experience of the network environment and overall activity. However arriving at an optimal threshold limit for tracking the most important events is a difficult task and should be automated. The system should be capable of continuously adapting the threshold limit so that only a fixed percentile of events is reported irrespective of activity level in the network.

Event deduplication

Repeated event reporting for the same entity and problem is annoying. Once a certain number of events are reported the system could hibernate event generation activity for a certain time period and then resume it. However the system should keep track of the specific network activity passively even during the hibernation period so that overall statistics collection, continuous threshold adaptation, etc. could go on.

Special functions for data collation

Aggregation functions for collating unique list of values, Transitive Closure functions for Link Analysis, etc are very handy in enhancing the usability of the events. It's important to note that building user defined functions in the database for similar operations is much more cumbersome.

Problem specific data mining algorithms

Rule based event generation coupled with complex event correlation can capture a variety of network activities. But capturing certain tricky activities such as stealthy malwares & botnets, p2p applications, fast worms, etc need some problem specific data mining and profiling algorithms. It's important for the system to offer a systematic API rich framework for plugging in any such custom routines.

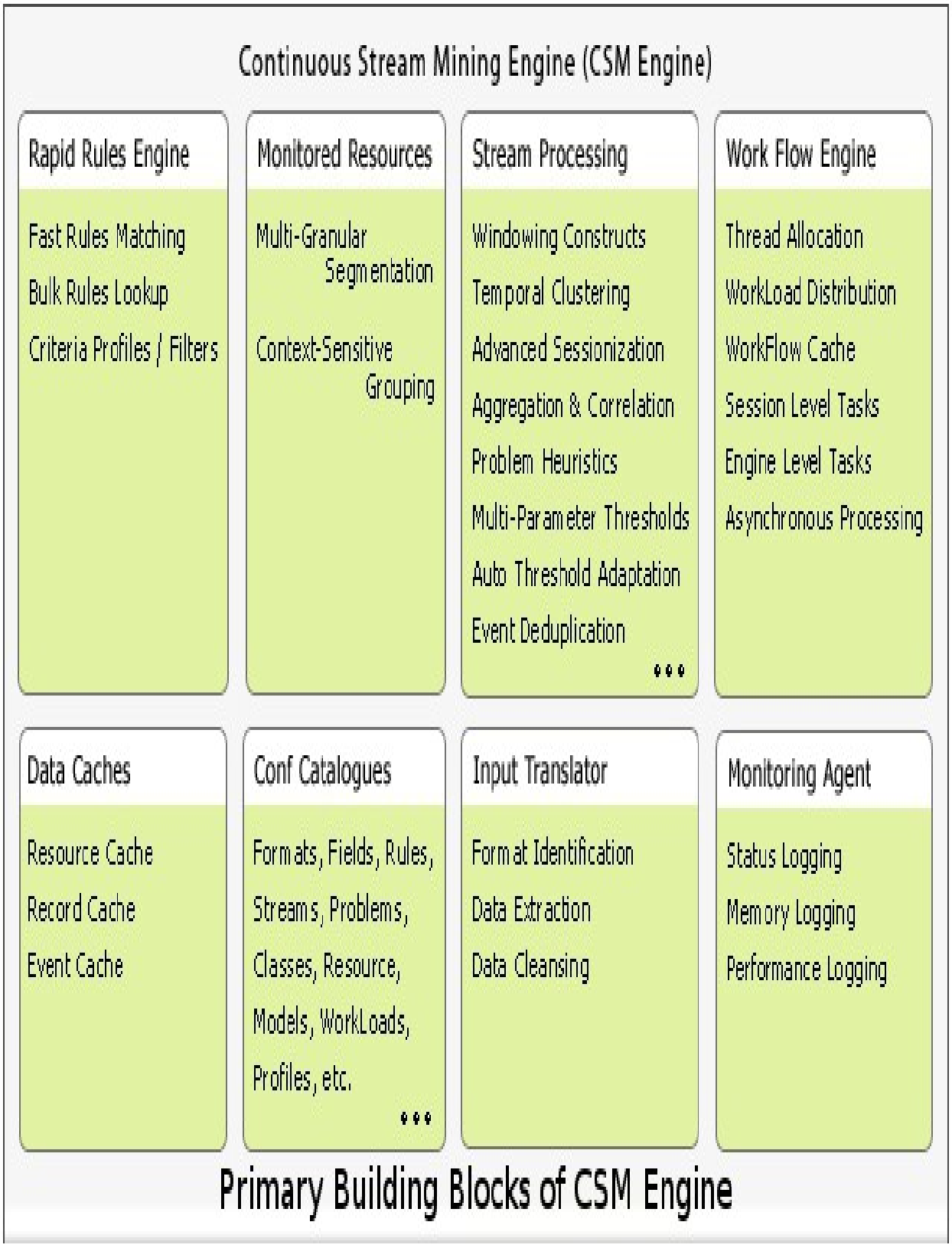
ManageEngine NetFlow Analyzer's Advanced Security Analytics Module (ASAM)

ASAM is a network flow based NBA tool for security analytics. It helps detect & classify zero-day network intrusions real-time, using the state-of-the-art *Continuous Stream Mining Engine*™ technology. ASAM has taken into account the above mentioned strategies and provides better, easy-to-understand information on your network security. ASAM offers actionable intelligence to detect a broad spectrum of external and internal security threats as well as continuous overall assessment of network security. ASAM, offered as a simple add-on module of NetFlow Analyzer, leverages the underlying platform's agent-less centralized data collection and forensic analysis capabilities, to offer greater value. NetFlow Analyzer is a robust, scalable and a proven platform offering bandwidth monitoring and unified traffic analytics.

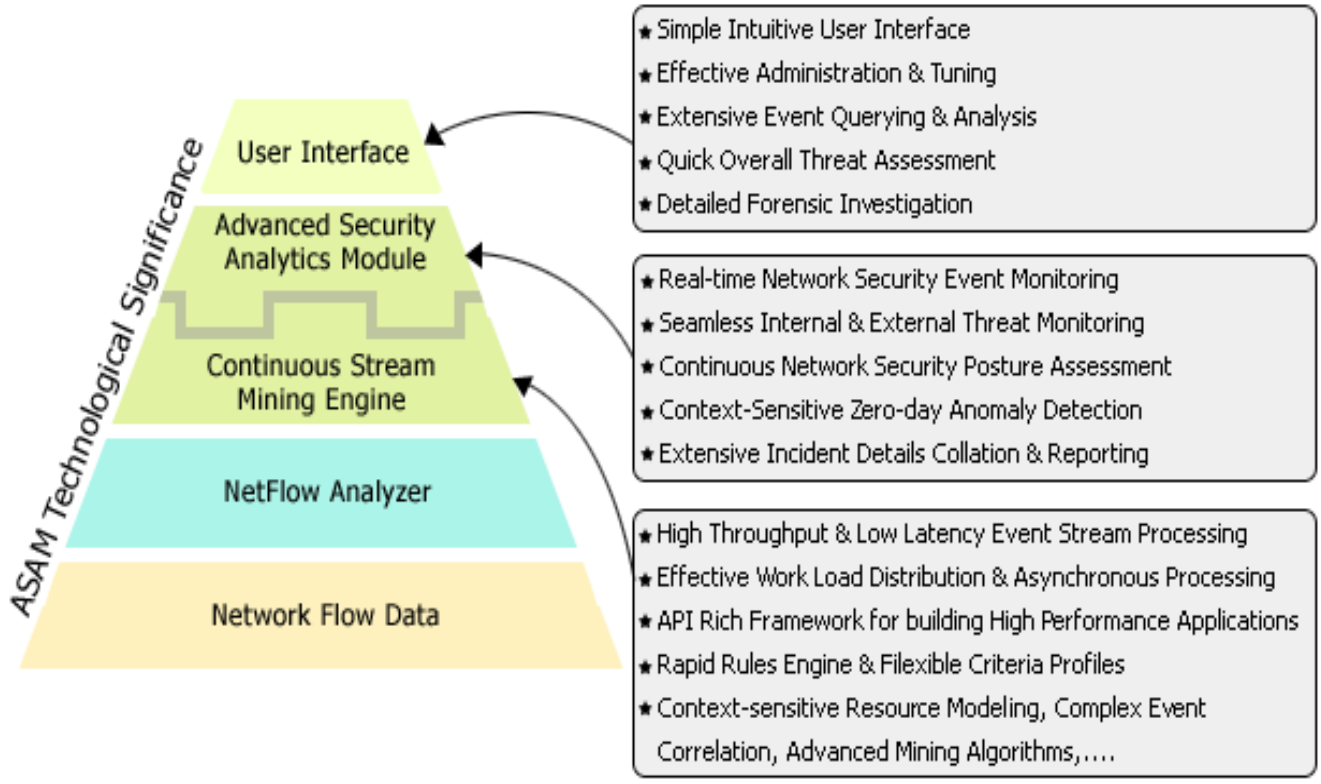
Continuous Stream Mining Engine™ (CSME)

CSME is a Java based Complex Event Processing (CEP) Engine for real-time complex pattern matching & event correlation across multiple events, based on some effective strategies involving fast bulk-lookup rules matching, multigranular context-sensitive resource modeling, temporal clustering & advanced sessionization, automatic threshold adaptation, de-noising & de-duplication, and heuristics based event stream classification. Primarily it employs a Rapid Rules Engine, a variety of data structures for indexing & caching, partitioning & windowing constructs, contracts/interfaces and generic base implementations for data aggregation and event correlation. It offers a configurable and extensively customizable, API rich framework for building high performance Event Stream Processing (ESP) applications.

Architecture of ASAM coupled with CSM engine:



ASAM Technology Benefits:



About NetFlow Analyzer

NetFlow Analyzer is a, web based (no hardware probes), bandwidth monitoring, network forensics and traffic analysis tool that has been optimizing thousands of networks across varied industries for peak performance and helping them to optimize their bandwidth usage. NetFlow Analyzer is a NetFlow / sFlow / JFLow (and more) collector, analyzer and reporting engine integrated together. With close to 4000 enterprises using NetFlow Analyzer for an in-depth visibility into their network traffic and its patterns, NetFlow Analyzer continues to earn trust of more users by giving business knowledge of real-time network behavior and how traffic impacts the network's overall health.

About ManageEngine

ManageEngine is the leading provider of low-cost enterprise IT management software and the only one making the 90-10 promise – to provide 90 percent of the capabilities offered by the Big 4 at just 10 percent of the price. The ManageEngine suite offers enterprise IT management solutions including Network Management, HelpDesk & ITIL, Bandwidth Monitoring, Application Management, Desktop Management, Security Management, Password Management, Active Directory reporting, and a Managed Services platform. ManageEngine products are easy to install, setup and use and offer extensive support, consultation, and training. More than 40,000 organizations from different verticals, industries, and sizes use ManageEngine to take care of their IT management needs cost effectively. ManageEngine is a division of ZOHO Corporation. For more information, please visit www.manageengine.com.

References

- Complex Event Processing (CEP) - http://en.wikipedia.org/wiki/Complex_event_processing
- Event Stream Processing (ESP) - http://en.wikipedia.org/wiki/Event_stream_processing
- Precision and recall - http://en.wikipedia.org/wiki/Precision_and_recall

Request Demo

Product Download

You can reach the author at **chandramoulis@manageengine.com**