Reducing IT Support Costs Through Historical Application Monitoring

Overview

The proliferation of public and private voice and data networks has created massive challenges for software developers, equipment vendors, and businesses. Providing an electronic infrastructure that serves content reliably, securely and with acceptable levels of performance and quality requires considerable technical coordination and unification of the diverse networked infrastructure.

Within the Enterprise, network and application services must be constantly monitored to support twenty four hour work days. Further, this monitoring must be compared against service targets or service level agreements that establish formal benchmarks for these services. Many products exist that can provide 'current-state' service monitoring. Some of these products can provide a summary of network service performance and reliability over a few weeks of time script driven static reports. Open source products such as Multi Router Traffic Gopher (MRTG) offer a glimpse into what is possible. Application performance remains a more clouded subject.

Due to the diversity of application types and the increasing demands placed on them by a post-Internet ecosystem, they must support diverse and distributed end user communities. Further, the expectation that these application services will be highly available and secure creates a tremendous need to monitor their effectiveness. Having the capability to understand application service levels such as performance, reliability and security in real-time and in historical time across the infrastructure is the focus of this paper.

Today's Environment

IT organizations that install and maintain infrastructure services and related applications have difficulty understanding the historical reliability and performance of services. When problems occur, finger pointing between various IT groups happen when trouble spots cannot be isolated and the root cause of problems identified. In many cases, unnecessary capital expenditures such as upgrades are performed in a blind attempt to correct perceived shortcomings within the network, or application infrastructure.

Accurate infrastructure asset management must involve more than topology, inventories and help desk services. Eliminating guesswork surrounding service level bottlenecks is only possible if one has access to highly detailed historical data that accurately depicts current and historical conditions. Sustaining a traceable history of reliability and performance is required to properly manage these assets.

Today's IT Infrastructure support teams must be able to deliver a more accurate long-term picture a broad array of networks, and application services are expected to perform according to targets, specifications and agreements. Ideally, both remote and central sites must be monitored, providing IT with the ability to capture, aggregate and access service metrics in a common way across the entire infrastructure end-to-end, top-to-bottom. However, in most cases, infrastructure monitoring *nirvana* has not been achieved.

Reducing IT Support Costs Through Historical Application Monitoring

How Do We Optimize Our Services and Investments?

Questions are being asked by every IT Manager, Director, Vice President, and their Board Members surrounding the return on investment for expenditures. Once systems and solutions are purchased and implemented, the real cost of ownership begins. As much as eighty (80%) percent of the total costs of ownership of IT & Network infrastructure occur after the initial capital investment. Questions about the return of these investments include:

- How do we maximize return on capital investments?
- Are the assets deployed performing as advertised or as designed over time?
- How do our IT resources make informed business decisions concerning asset management?
- What are our service level goals and how do we achieve them?
- What is the real cost of downtime?
- Can we deliver content in a secure fashion that has not been compromised, or tampered with?
- Can we support our manufacturing and operating service centers 24x7x365?
- Can we deliver application quality of service?
- Do we know how our core applications and systems are running over time?

The answers to these questions vary widely depending on several factors including:

- Technology
- Business
- People

This paper will focus on effective technology and business models that facilitate optimal IT service delivery and asset utilization.

Service Models: Internal & Outsourced Services

Internally groomed ***IT services*** such as private networks and custom applications require constant monitoring and support to ensure reliable delivery and future predictable performance. Application development environments such as Open Source, J2EE and .Net provide alternatives for IT organizations to develop and support applications. Integration and implementation of these services cannot be efficiently accomplished without some form of service monitoring.

Many enterprise corporations are purchasing outsourced services from external service providers. These **external *IT services*** offerings are available as bundled and unbundled services such as converged voice and data services, storage, and security offerings such as virtual private networks. These managed service offerings are supported with guarantees of service and are in some cases, backed with monitoring and reporting for proof of service. Internal monitoring is still necessary. Ensuring that you are getting what you pay for by monitoring managed services becomes a critical need to ensure that your vendors are delivering what is promised.

Proof of Service Delivery

Taking accurate measurements of service levels and then comparing them against a baseline or target is an essential part of determining adherence to service targets. Understanding your service becomes a quantitative and qualitative exercise. The deployment of formal documents, specifications and targets are best described as one or more of the following:

- Service Level Agreements
- Service Level Specification
- Service Level Monitoring

Service validation, be it quantitative or qualitative is an essential part of proving that services are, and were delivered.

Quality of Service (QoS) – another service attribute, can be measured against hardened agreements that are part of Service Level Agreement contracts mentioned above. QoS in a general sense is the experience an end user can articulate when utilizing infrastructure services. More specifically, it may be voice jitter or service latency. Measuring response times, latencies and other key performance indictors over time provides a visible trend depicting service delivery.

Monitoring service providers must be accomplished for delivery of contracted or managed services such as:

- Network Bandwidth
- Application Latency
- VPN Connectivity

Reducing IT Support Costs Through Historical Application Monitoring

Many companies have unnecessarily spent money for services they paid for but never used or that were never verified to have performed as advertised.

Today’s Solutions

- **Equipment Providers** have addressed device level management of their products, and to a certain extent, provide some forms of service management. These point solutions are typically concerned with service definitions and topology illustrations only in real-time. They do not address life cycle asset optimization or provide any historical service level baselines.

- **Open Source** platforms provide point solutions and must be supported in-house.

- **Independent Software Providers and Open Source Platforms** have provided software products that address specific niches within the infrastructure management field. A few software providers have pioneered heterogeneous service management. Products such as Hewlett Packard’s OpenView™ are found in many IT organizations around the world.

What is Missing?

Simply stated, it is a lack of available historical data that depicts service levels over time. The scope and frequencies of data samples that must be collected to provide meaningful profiles of activities that span the Enterprise are enormous. Widely utilized technologies used for the storage of this data include brand name relational databases. These products have provided tremendous support with respect to monitoring service attributes and relative current state conditions. In reality, logging extended time periods of service metric data places tremendous strains on these systems. This is in part due to the fact that they were not designed for long term historical archival of data. Also, because the performance, reliability and security of networks and application services are growing exponentially, known limitations become exacerbated. Having a detailed data preserved in its original granularity is much like having an accurate fossil record. With an accurate record, there are no missing pieces that hide the true nature of the beast.

Technical Considerations Storage requirements for collecting more data must be dealt with in an economical way. Response requirements for retrieving months of data in real-time can and do overwhelm relational databases. The amount time required to retrieve a scalar array that represents thousands or tens of thousands of individual data points can be several minutes or longer. Time-series archival products can reduce this task to a few seconds.

Back End Storage. The implementation of a time series archive that works ‘behind the scenes’ to address data aggregation and presentation compliments existing topology, fault and service management systems. Implementing this technology will assist with:

- Capacity Planning Needs
- SLA & QoS Verification
- On-demand Computing Initiatives

Remote Service Monitoring and Logging

This is essential within the distributed enterprise. Acquiring data from remote and local sources for central storage in a continuous fashion is another key aspect of service monitoring. In order to determine when, where and why problems appear within the enterprise infrastructure, data collection from all vulnerable and sensitive points must be vigorously and accurately monitored. These remote sources include:

- Remote authentication servers/ports
- Gateways
- Firewalls
- Server log files
- Power and control system health
- LAN/WAN pinch points
- Voice over IP services

This remote monitoring can be performed by utilizing commodity priced appliances that locally poll and trap events of interest. Once captured, these appliances act as a temporary storage location until verified replication of data has occurred. Providing automatic fail-over support helps ensure that no data is lost.

Diversity of data types available for collection also add tremendous value by allowing support resources to visualize problems through correlations of seemingly unrelated events. The fact that environmental control systems can be monitored and logged in the same archive as network and application service data allows central IT organizations to better determine the true root cause of service failure or degradations.

Intelligent Consolidation of remotely collected service data provides a common set of data for IT support personnel. In cases where infrastructure performance, reliability and security data is missing, or is simply not readily accessible, IT must guess at conditions that if properly logged, would reveal the nature of the problem at hand. The continuous centralization of remotely captured service metrics from multi-site, multi-technology data sources is not unique. What is different about intelligent time-series archival is that data can be preserved in its original fidelity without requiring massive storage devices to absorb redundant or meaningless data. The intelligence of collecting events of interest is the effective application of remote and central policies. These policies are applied as attributes to each unique source point or tag of data. They serve as intelligent filters that effectively weed out unwanted data, preserving the true signatures of events. Examples include:

- Faults
- Congestion
- Poor Quality
- Intrusions or Hacks

User Driven & Programmatic Access to Time-Series Service Metrics

These events are displayed in graphic detail to help reveal the true root cause of unplanned events that disrupt and compromise services. Providing add-hoc trending of time-series data allows users to understand relative performance and reliability of services.

Integration or exchange of this information across various service platforms is also essential. The ability to expose intelligent events as time-series sequences is a key capability of any service validation system.

Gaining programmatic access to time-series sequences through API/SDK calls or XML/SOAP web services affords a rich set of options for developers seeking to capitalize on the high availability of service metric data. Examples include linking to other applications that support help desk operations, security policy management and provisioning/activation platforms.

On-Demand Service Performance & Reliability

With the advent of self activating or customer driven service activation, on-demand services must be monitored in new and innovative ways. Illustrating service metrics is typically done using transient real-time values and highly summarized static reports. Within a true on-demand environment, this approach is wrong.

Customer facing and internally positioned Portals that unify dynamic data across the enterprise must show the characteristics of service startup, operation and eventual termination. Without complete flexibility over the manner and fashion by which service metrics are logged, determining the relative performance and reliability during different phases of a service is difficult. On-Demand service that offers service improvements based on both fixed and dynamic requests for higher priority services are marketed by name such as:

- Gold
- Silver
- Bronze

The ability of a customer to request streaming video and interactively adjust the quality of service is becoming a reality. Having the capability to archive service metrics based on changing service level specifications or profiles is essential. Marking transitions in service attributes in real-time while data is being collected requires a sophisticated ‘batch’ or product specification subsystem to archive transitions in the Service Level Specifications (SLS).

Security Monitoring

Today's solutions may show details depicting network traffic such as packet headers for a few hours. Log files containing this data are typically manually consolidated into large numbers of text files that are not linked or indexed. For the few packages that manage to store structured records, the time ranges available are very short. Proper monitoring includes proper logging and display of key events that illustrate threats in a clear and meaningful way for IT Security analysts. Batch processing of historical log files as well as real-time packet capture are complimentary components and necessary to support the analysis of firewalls, packet filters, proxy servers and other security devices found on the front lines of any protected environment.

Each customer has different security concerns. Unique site configurations, fluctuating user accounts, and mission-critical applications interact to provide a rich set of details that describe traffic patterns and other important details about resource utilization. Each month, millions of suspicious events can occur in large service networks. Sorting through this extremely high number of events on a day-to-day basis represents a daunting challenge for many enterprises. This haystack of events will usually reveal a few surprises. IT organizations must uncover increasingly disguised events that elude the first line of defense.

Firewalls and router-based packet filtering generate thousands log files in a single day that contain information describing which sources, destinations and protocols were observed traversing networks. Firewalls cannot provide protection against dynamic and agile hackers who figure out how to bypass these devices. Post-attack forensics must include the ability to play back the sequence of events leading up to attacks or abuses of resources. Relying on limited historical sequence of events prevents security analysts from rapidly triangulating threats. Questions such as

- How many failed server login attempts over the last month occurred?
- What was the time frequency of scans or passive hacking attempts?
- Who or what was affected (and how often) by successful penetration of DMZ zones?

Determining these answers by perusing millions of records across extended time periods in a distributed environment is a very difficult task without a centralized storage and retrieval management system. Time-series storage coupled with intelligent retrieval and data mining tools can provide the ability to look back in real-time to ascertain the answers to these questions.

Application Monitoring & Quality of Service

Return on Investment (ROI) and Total Cost of Ownership (TCO) are two terms that are used to explain why application monitoring is important. Perhaps the biggest reason is the ability of organizations to function all of the time, everywhere in a dynamic and agile fashion. Common measurements of application service quality include:

- Response time
- Throughput
- Utilization by resource
- Expected behaviors

These can help you determine how well you have provisioned services, planned for variable utilization patterns and enacted quality of service policies to support mission critical applications. In many cases, control layer solutions from equipment vendors are able to mark packets with higher priorities, ensuring that critical business applications execute as designed. The challenge is perhaps more acute within the mobile workforce. Extranet and Internet connections are especially vulnerable to lapses in application quality of service. Even secured private networks should constantly monitor the execution times of mission critical services.

Simulating transactions across the service infrastructure verifies that device level configurations are correct and that the entire infrastructure is available to execute end user requests. The results of these simulations must be logged and compared against benchmarks, or service targets to ensure that applications are meeting business demands. Appliances that help regulate resources for high priority application execution must be monitored as well. Without a timing chain to synchronize these devices, simple configurations such as access control lists or more complex on-demand computing control planes can never be optimized.

Technology in Continuous Processes

Process manufacturing operations are in their simplest form, a continuous series of events. Examples include power generation, oil & gas production, and pulp production. IT services are also continuous processes. Monitoring, data archiving and display of key events in continuous process environments requires specialized software to handle the load.

Time-based data archiving is a key software technology that provides efficient storage and recall of performance and reliability service metrics. The key steps in this process include:

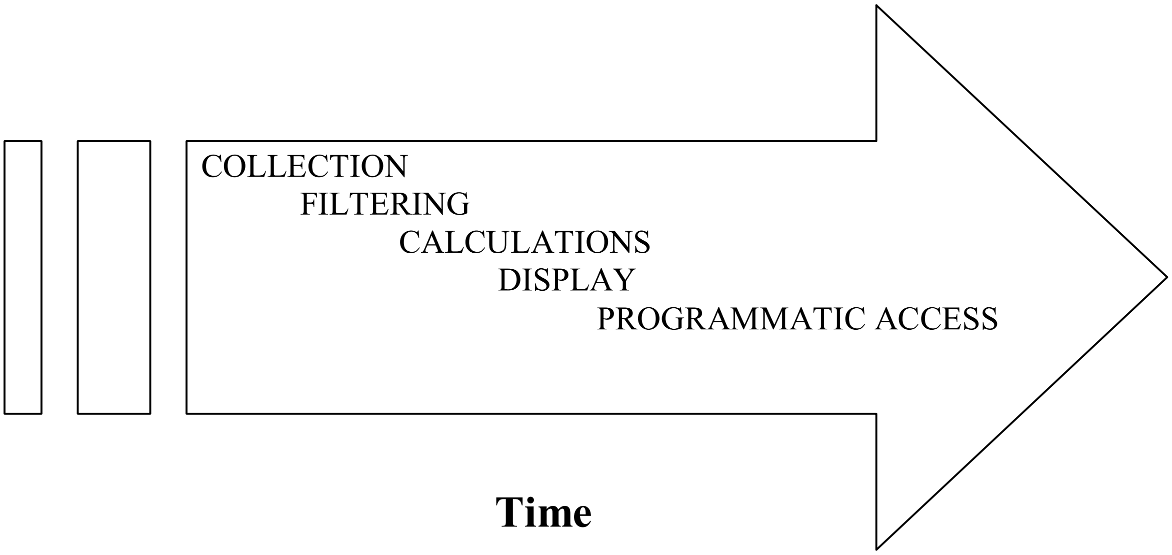
- **Collecting** application service response metrics and associating a time-stamp is the first step towards understanding fundamental application quality of service.

- **Filtering** of white noise from raw data is another step in the transformation of unconsolidated events to intelligent logging.

- **Calculations** that provide key performance indicators or consolidated numbers for billing/utilization applications is also a key time-based sequence of events that must be provided.

- **Displays** and programmatic access of time-series service metrics complete cycle of events that define the proper use of software in IT service environments.

See diagram below.



Mission Critical applications can only be effective if the appropriate types of information that depict the overall service infrastructure are collected and aggregated. Examples of these data include:

1. **Server/Operating Systems Key Parameters:** CPU, Memory, Page File or Swap Space
2. **Network Interface/Device Metrics:** Bandwidth Utilization
3. **Internet Protocol (IP) Packet Headers:** Source and Destination IP/Port Addresses
4. **Voice Quality:** Latency, Jitter
5. **Response Times**: ICMP Pings, TCP Connection Times

Gaining Insights

Reducing IT Support Costs Through Historical Application Monitoring

Capturing, filtering, calculating service data with effective preservation and reproduction of high-fidelity events of interest provide an accurate archive of performance and reliability over time. Visual integration of time-align service metric trends reveal the root cause of network saturation and application latencies over extended time periods. Continuous real-time trending coupled with historical access to raw data, and derived events revealed unique patters of application resource competition.

Who Needs All This Data?

IT resources that require access to detailed service include:

- Capacity Planners who struggle with retrieving enough relative information about application performance to make informed decisions about future capacities
- Security & System Analysts are faced with a myriad of fragmented node and server log files. They must manually peruse this disjointed tapestry of seemingly unrelated information.
- Network (NOC) Engineers must fight to keep services alive and performing 24x7.
- Application Analysts must have application utilization statistics determining application design flaws, excessive or unauthorized application utilization.
- CTO/CIO must understand the overall efficiency and throughput of the IT infrastructure through a highly available and visible monitoring and development platform

A New Breed

A new breed of IT infrastructure support tools are now in demand. These systems must be able to address:

- Real-time & historical trending of performance, reliability over extended time periods (more than 90 days)
- On-line and off-line reporting of resource utilization patterns down to the packet layer
- Central logging of distributed IT asset performance, reliability data
- Consolidation of fragmented firewall and server log files into a central database
- Rapid creation of display & reports that draw on a rich mixture of infrastructure data to display the details of events leading up to faults, system failures and underperforming assets.
- Rapid scalability across diverse environments

- Central storage of months and years of data
- Remote collection with buffering
- Multi-platform, multi-technology support
- Integration capabilities via API/SDK and emerging standards such as XML/SOAP
- Little or no impact on production services or infrastructures
- Interoperability with existing/legacy systems.

Platform Paradigm Shift

In summary, in order to effectively manage the availability, reliability and security of IT Services, a software platform is necessary to provide the following capabilities:

- Absorb high volumes of time-sensitive, time-series data
- Reproduce this data on-demand in accurate and meaningful ways
- Unify diverse data sources
- Support legacy investments
- Provide a best of class infrastructure architecture

The implementation of a product that unifies diverse infrastructure data sources and provides quantitative and qualitative application service level measurements across extended time periods, provides new opportunities to derive higher value and predictability from assets and applications. Understanding who, what, when and where application services are invoked and how they performed or did not perform over time, is instrumental in reducing support costs and improving service delivery.

OSIsoft, Inc.
www.osisoft.com