TIME TO RETHINK GUARANTEED DELIVERY

Guaranteed packet capture ensures that all data is collected in real time with zero packet loss, but this is no longer enough. To assure zero packet loss for the analysis application, the network appliance needs to be designed for guaranteed delivery. Guaranteed packet capture is the first step, but if the application is not designed correctly or the server configured properly, then packet loss can occur.

TIME TO RETHINK GUARANTEED DELIVERY

Any analysis is only as good as the data being analyzed. Any decision based on this analysis must therefore concern itself with the validity, precision, accuracy and completeness of the data that is analyzed. It is for this reason that OEM vendors of network monitoring and analysis appliances require guaranteed delivery of every packet as quickly as possible. Otherwise, a precise and accurate analysis is undermined.

Note that we refer to guaranteed delivery and not just guaranteed capture. This is an important distinction. Zero-loss packet capture is a necessary first step, but is not enough as the data also needs to be made available to the analysis application in software. While, on paper, this is straightforward, in practice, there can be surprising hurdles to achieving this that network appliance designers need to consider.

The vast majority of network monitoring and analysis appliances are based on standard servers. With the latest servers, tremendous processing power and memory is available for analysis applications. Nevertheless, without a proper understanding of modern server architectures, careful configuration and an appreciation for the issues that can arise when trying to perform demanding network monitoring and analysis, network appliance designers can experience performance issues, lost packets and, in the worst case, failing systems.

The following will take a closer look at these issues providing more insight on how network appliance vendors can avoid these issues and not face a customer support nightmare.

GUARANTEEING DELIVERY

There are many different types of network monitoring and analysis applications ranging from application performance monitoring to network emulation to network security. What these applications have in common is a need to analyze all the data in real time even at full load. In other words, they need to analyze every Ethernet frame and IP packet.

This is where intelligent network adapter vendors like Napatech help. We provide accelerators for network management and security applications that promise 100% packet capture under all network conditions and are widely used in different types of network appliance.

Nevertheless, 100% packet capture is not always enough. Especially, if the application or server on which the application resides is not capable of receiving all the packets captured. There are a number of reasons why this might be the case:

- The analysis application is single threaded and can only run on 1 CPU core, so when this CPU core reaches its maximum processing capability, packets are lost
- The driver for the network adapter is not optimized and consumes a single CPU core leading to packets lost
- The server is not configured properly so data is not being stored on memory caches local to the CPU core, but must be transferred across limited internal interfaces leading to congestion, delays and packet loss
- Multiple applications contend for the same CPU core cycles leading to a CPU core reaching maximum capacity and packets being lost
- Multiple network adapters or other peripherals contend for PCIe bandwidth leading to congestion and packet loss
- Insufficient cooling of network adapter hardware leading to overheating and network adapter failure (guaranteed packet loss).

It is therefore important that all of these issues are considered in the design of network appliances and in the choice of network adapters to use.

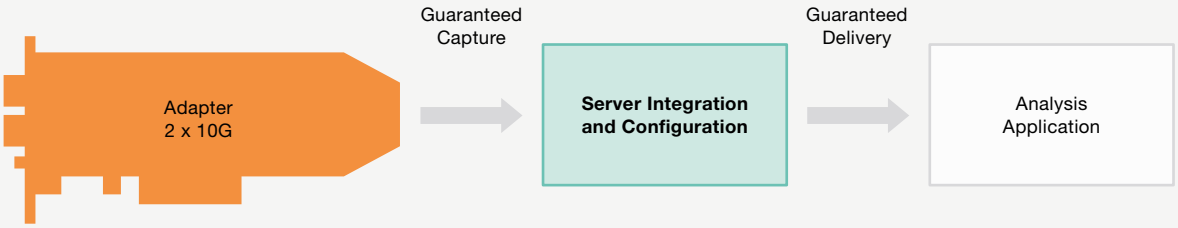


FIGURE 1
Intelligent network adapters can guarantee packet capture, but for guaranteed delivery to the application other factors need to be considered

SPREADING THE LOAD

Modern servers now come packed with multiple cores per CPU and multiple CPU sockets. However, there are still many analysis applications that are single-threaded (i.e. they are designed to run on only 1 CPU core). With network speeds increasing from 1 Gbps to 10 Gbps and 100 Gbps, it does not take long for a single core to reach its processing limit. Consider that if you want to analyze all the data in both directions on a 10 Gbps connection, then you have to be able to process up to 30 million packets per second.

Modern analysis applications are designed to be multi-threaded allowing the load to be spread over as many CPU cores as there are available. This not only solves your immediate problem, but also future problems as demands increase over time.

Of course, the network adapter must support multiple CPU cores in a balanced way and provide mechanisms to intelligently distribute traffic according to the type of data that is being analyzed. Load balancing is not always enough, especially if there is a leaning towards a specific type of traffic on the network. That is why Napatech provides 17 different hashing methods that can be dynamically enacted based on the type of data received.

ENSURING LOW CPU LOAD

One of the value propositions for Napatech accelerators is application off-load. In other words, the ability of the network adapter to take over responsibility for data handling tasks so that the application is free to use the available CPU cycles for more important tasks. These data handling features are implemented on the accelerator itself, thus consuming no CPU cycles. Examples are port data merge, frame slicing, de-duplication, frame classification, generation and inspection of checksums, IP fragment identification and tunneled traffic handling to name a few.

Directly transferring this data into server memory (and cache memory) also ensures that no unnecessary copying or CPU interrupts are generated, especially with bulk data segment transfers.

Without these features, too much needs to be performed in software by either the network adapter driver or the analysis application, leading to less CPU cycles for the actual analysis. If network adapters are not designed for network monitoring and analysis they can consume so many CPU cycles on a single CPU core that packet loss will occur.

This situation is typical of standard communication adapters or NICs. These are designed for communication from one end-point to another, where the focus is only on the Ethernet frames related to these endpoints. But, network monitoring and analysis requires analysis of every Ethernet frame, which is obviously not what these types of adapters are designed to do.

In addition, network analysis requires analysis of a merged stream of data from multiple ports. If you are receiving network data from a tap, this will provide two streams of data on two separate ports; one for the east-west traffic and the other for west-east traffic. This data needs to be merged for analysis of conversations between end-points. Standard NICs do not provide merging of data traffic on the adapter.

CONFIGURING THE SERVER

One of the common architectures of modern servers is the Non Uniform Memory Access (NUMA) architecture, where instead of one shared memory for all CPUs, each CPU has its own local memory. Configuring the server so that the right data is sent to the right memory is important to ensuring optimal performance.

In addition, each CPU has a layer 3 cache that is shared amongst the cores implemented on the CPU. Each CPU core has its own layer 2 and layer 1 cache. When a CPU core wishes to retrieve data, it will first consult its layer 1, layer 2 and then layer 3 cache. If it cannot find the data locally it reads from external memory, which could be either from memory or remote L3 cache if a remote CPU contains an unmodified copy of the memory.

There can be from 50 to 100 times difference in latency between accessing data locally in the layer 1 cache and retrieving it from memory. For this reason, modern implementations provide the intelligence for other CPUs to detect a memory read and redirect the request to their local cache if they know that the requested data is available in their cache.

So, if the server is not configured properly, data can be shared in multiple L3 caches giving both bad cache utilization and extra traffic over the Quick Path Interface (QPI). A more serious problem with shared cache lines arises when data is modified because that results in cache invalidation in the remote caches and the CPUs will have to re-fetch the data from memory. When they modify the data, the invalidation occurs again and the process is repeated. This is known as cache ping-pong. The most optimal solution is to avoid shared cache lines because it will limit the traffic on the QPI interface and will also provide better cache utilization because data is only placed in one cache.

Issues can also arise if multiple CPU cores are accessing the same remote cache to retrieve data. This can result in CPUs burning cycles evicting each others data and performing cache line replacement. This is also referred to as “cache eviction” or “cache thrashing”.

This combination of interface congestion, latency due to remote access and burnt CPU cucles due to cache thrashing can lead to packet loss.

AVOIDING CONTENTION

Standard servers are powerful platforms, but are also general purpose platforms, open to multiple applications and peripherals. Careful thought must be given to this when configuring the server for network appliance use.

Establishing the right CPU core affinity for each application is important to ensure that the analysis application has exclusive access to the CPU core processing power that it needs and will not need to have to contend with other applications that might be started due to automatic updates, computer scans or any of the many applications that are now commonplace and are often an afterthought.

Equally, when using multiple peripherals, such as RAID controllers, compression cards and others, it is important to manage these installations and assign correct CPU affinity to avoid contention for the same CPU cycles and no potential PCI Express contention.

When contention occurs, even if it is only for a few microseconds, this still means that several million packets can be delayed or lost before they can be analyzed.

ASSURING GUARANTEED DELIVERY

To absolutely guarantee delivery in a standard server environment where there can be issues that are outside the control of the network adapter, it is vital that the network adapter provides some local buffering capacity. This is typically for congestion situations or analysis application delays where packets cannot be transferred, but must be delayed and buffered. This is typically in the region of microseconds, but as mentioned before, this can still be a significant number of packets.

In theory, with Full PCIe bandwidth, fast processors, direct memory access and many high-capacity PCI Express lanes, this should not be necessary, but our experience with over 100 OEM vendors of network appliances indicate that if you want to guarantee delivery and never lose a packet, then on-board buffering is a must. This is not a typical feature of adapters not specifically designed for network monitoring and analysis.

CONSIDERING HEAT DISSIPATION

Intelligent network adapters are designed to off-load data handling tasks from the analysis application to the network adapter. In essence, the network adapter is performing more of the processing on the adapter than traditional standard NICs. More processing means more work, which in turn means more heat generated.

This will happen regardless of technology. ASICs traditionally exhibit lower power dissipation than FGPAs, which in turn have considerably lower power dissipation than NPUs, but there is a common heat dissipation challenge for all these technologies and that is the availability of airflow in the PCI Express slot.

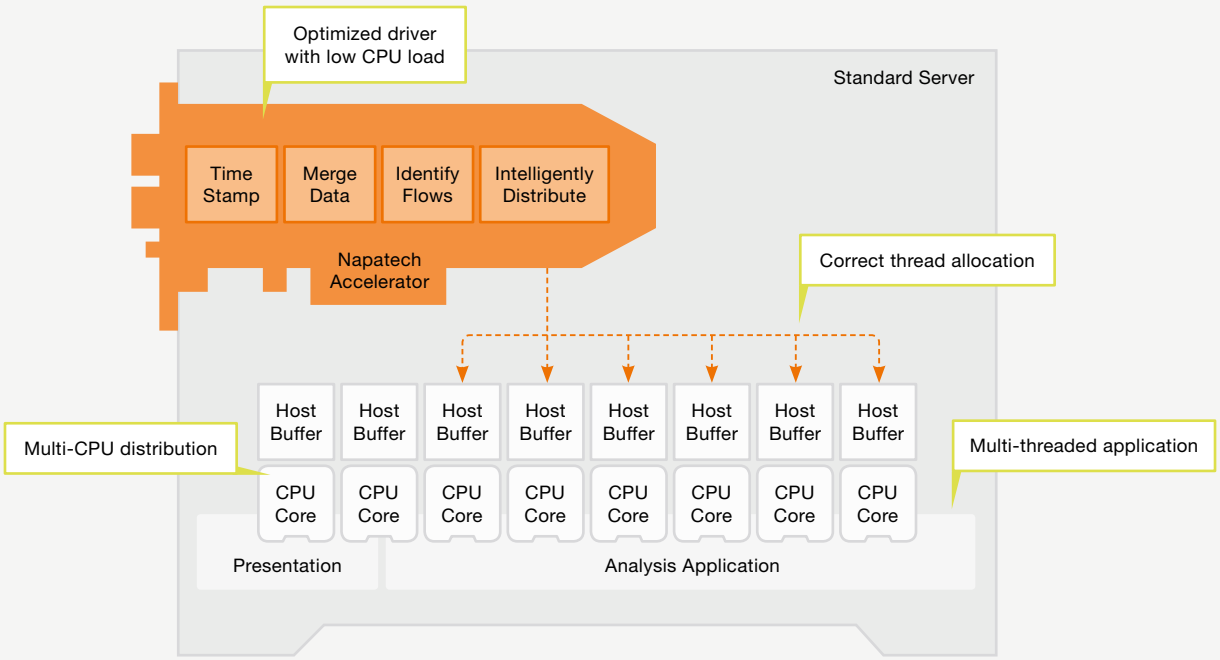


FIGURE 2
Assuring Guaranteed Delivery

In an effort to make servers greener and consume less power, modern servers try to minimize the amount of airflow required in the box. The power hungry CPU cores get priority, while it is often standard that very little or even no airflow is provided to the PCI Express slots. Without airflow, the heat generated by the network adapter cannot be removed. It will thus become hotter and hotter leading to degradation in performance and eventually failure. This effect has been seen for some network adapters almost immediately, whilst for others it can take several months before the impact of this issue is seen. The latter case is probably the worst, as this will lead to failures in the field as well as a support and customer relations nightmare.

At Napatech, we discovered this issue many years ago and have taken steps to ensure that our accelerators are designed to be self-cooling and can thus operate with extremely low airflow. It is possible to configure certain servers to provide more airflow and this should also be investigated, but to guarantee delivery, it is important to ensure that the network adapter is designed for efficient heat dissipation in low airflow environments from the beginning. Unfortunately, this is far from commonplace and at the time of writing, we are not aware of any other network adapter vendor capable of offering the same support.

RETHINKING NETWORK APPLIANCE DESIGN

As we move to even higher speeds and higher traffic loads, network appliance designers need to keep the issues of multi-threading and CPU configuration in mind and carefully consider the peripherals they use, especially which intelligent network adapter they choose to use.

The intelligent network adapter must provide the basics of 100% packet capture under all load conditions with low CPU load. But, it must also support hardware-based port data merging, intelligent multi-CPU core support, be NUMA aware and provide on-board buffering. Self-cooling capabilities are also important in ensuring that the network adapter can be widely used in any standard server and any environment.

For the designer, it is a question of thinking performance and reliability in from the beginning and ensuring not only that packets are captured, but that they are guaranteed to be delivered to analysis applications no matter the environment the network appliance will experience. This avoids costly support issues, disgruntled customers and the worst scenario of all, failing appliances in the field.

DISCOVER THE POWER OF NAPATECH

Napatech accelerators are designed to handle the maximum theoretical throughput of data for a given port speed. Napatech offers a range of accelerators supporting speeds from 10 Mbps to 100 Gbps. A single, common Application Programming Interface (API) allows application software to be developed once and used with a broad range of Napatech accelerators. This allows combinations of different accelerators with different port speeds to be installed in the same server. Additional features include:

- Napatech accelerators can identify, filter and distribute flows to up to 32 CPU cores
- Data merging functionality allows flows from different ports on different accelerators to be merged for analysis
- Data sharing functionality allows multiple applications to access the same data at the same time
- All of this can be performed with very low server CPU load

CHOOSE THE MARKET LEADER

Napatech is the market leading provider of accelerators for network management and security applications. Napatech provides global sales and support from local offices in all major continents, which is included in the price of the accelerator. This means that our highly experienced support resources are available for design and integration support, as well as field support without extra charge.

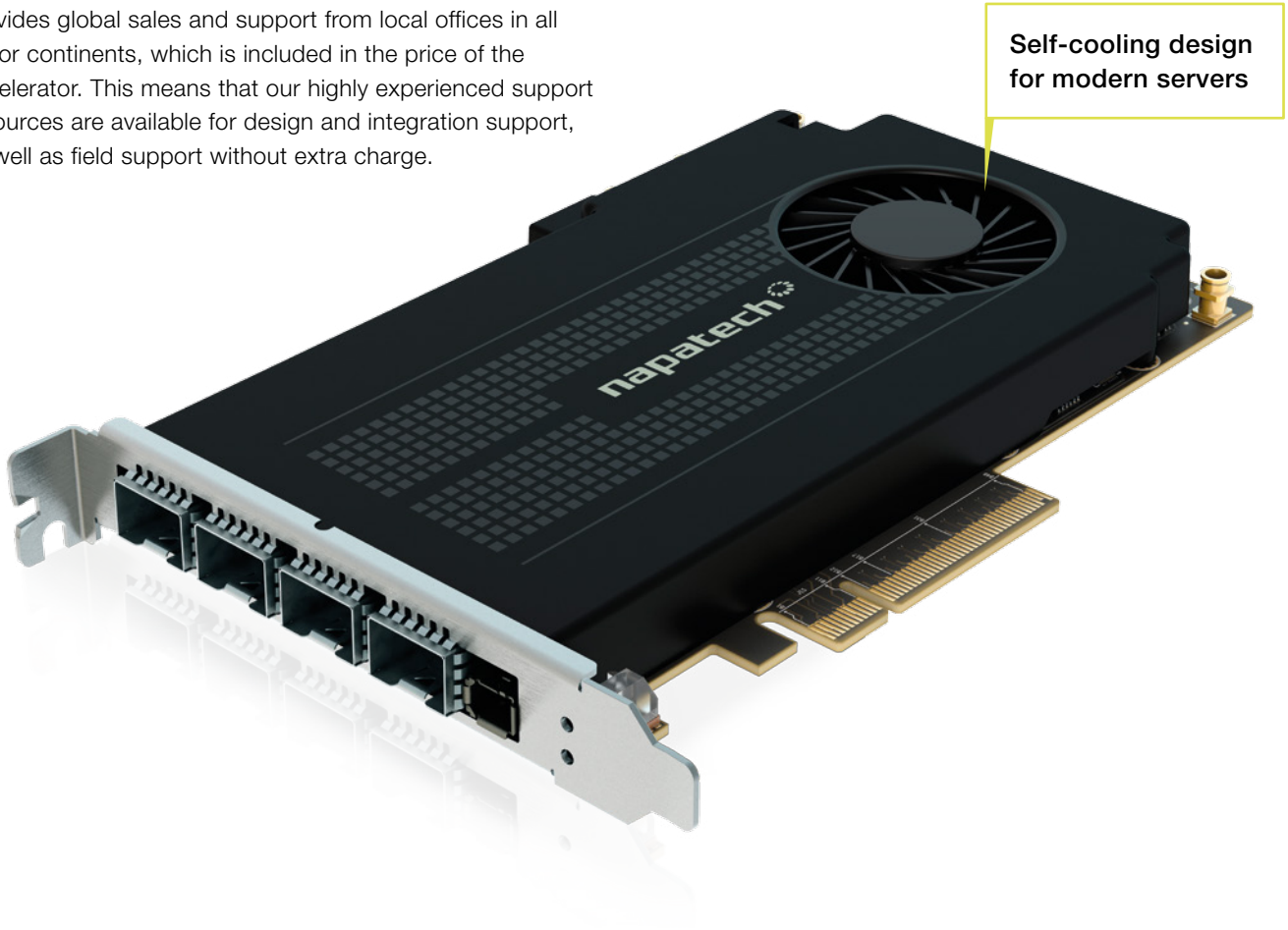
Napatech accelerators are manufactured to the highest standards by outsourced manufacturers in Switzerland and the USA supporting all major certifications including NEBS for telecom applications.

COMPANY PROFILE

Napatech is the world leader in accelerating network management and security applications. As data volume and complexity grow, the performance of these applications needs to stay ahead of the speed of networks in order to do their jobs. We make this possible, for even the most demanding financial, telecom, corporate and government networks.

Now and in the future, we enable our customers' applications to run faster than the networks they need to manage and protect.

Napatech. FASTER THAN THE FUTURE



EUROPE, MIDDLE EAST AND AFRICA
Napatech A/S
Copenhagen, Denmark

Tel. +45 4596 1500
ntemeasales@napatech.com
www.napatech.com

NORTH AMERICA
Napatech Inc.
Boston, Massachusetts
Mountain View, California
Washington D.C.

Tel. +1 888 318 8288
ntamericassales@napatech.com
www.napatech.com

SOUTH AMERICA
Napatech
São Paulo, Brazil

Tel. +55 11 2127 0782
ntsouthamericasales@napatech.com
www.napatech.com

APAC
Napatech Japan K.K.
Tokyo, Japan
Tel. +81 3 5326 3374

Napatech Korea
Seoul, South Korea
Tel. +82 2 6001 3545

ntapacsales@napatech.com
www.napatech.com