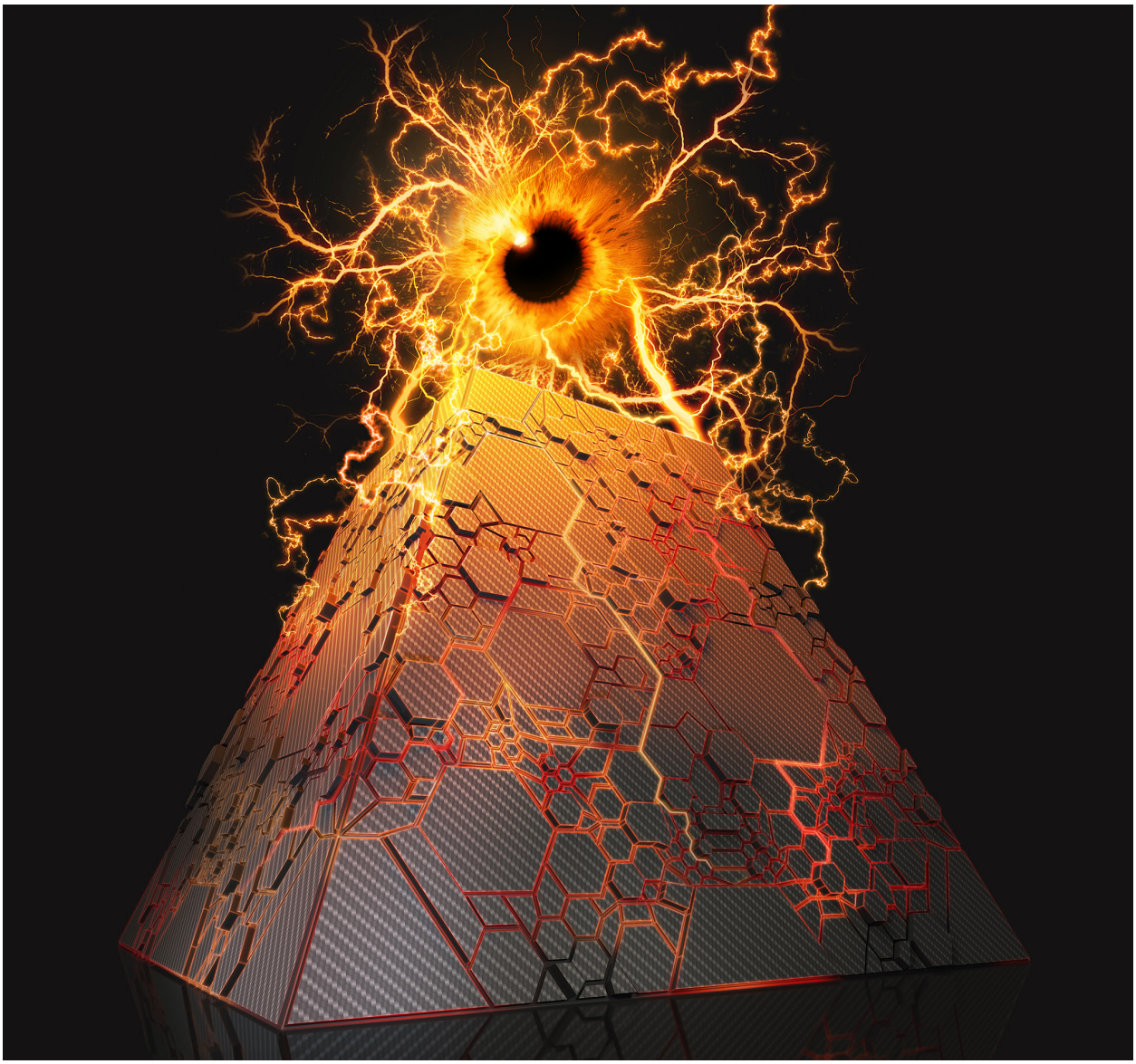


THE VISION OMEGA

A proposed 100 gigabit packet capture appliance using a time division packet steering strategy to sustain full duplex line rate capture to disk.

July 13, 2016

Andrew Watters, CEO
Rællic Systems
(415) 261-8527
director@raellic.com

CAGE: 782B5
DUNS: 079548638

RÆLLIC SYSTEMS

EXECUTIVE SUMMARY

Allied governments and large corporations may require 100G packet capture appliances to monitor, analyze, and record traffic on their 100G networks. The Vision Omega would meet this requirement and also permit traffic replay of activity preceding a major event, such as a cyber intrusion or act of terror. The system uses commercial off the shelf hardware and software, with some modifications unique to this demanding environment.

The system occupies one and a half or two and a half racks, depending on the configuration. The base model uses 42U of cluster-based storage plus two 4U control units that split each direction of a passively tapped 100G connection into twenty 10G connections, which are connected to each unit in the cluster. In the base configuration, each cluster device captures approximately 1/40th of the traffic on the connection on a round robin basis, with capture to disk at 625 MB/s per cluster unit. This is sufficient to provide sustained full duplex line rate capture to disk on a fully saturated 100G connection using only spinning hard drives (SSD's would not last long in this environment). The upgraded configuration uses an additional control unit and one cluster for each direction of traffic, in which case each cluster unit captures 1/80th of the traffic. Each configuration permits long-term storage of interesting traffic in a high performance database (available separately).

The length of available traffic replay on a fully saturated 100G connection with one storage cluster is 11 hours. With two storage clusters, each capturing one direction of traffic, replay on a fully saturated 100G connection is doubled to 22 hours. Additional storage clusters may be added to increase traffic replay, potentially without limit, but the price for the system increases significantly with each added cluster. It should be noted that this is a worst case scenario. A more likely scenario is a 1/3rd or 1/4th-saturated 100G connection, in which case traffic replay increases to between 30 and 88 hours or more, depending on configuration.

The capture files are standard pcap files, which are aggregated using high resolution (4 ns) synchronized timestamps and analyzed with any commercially available traffic analysis tool, such as OmniPeek or Wireshark. Long-term storage is accomplished simply by iterating over the binary pcap files, converting them to ASCII on the fly, and storing relevant sections of text in a database.

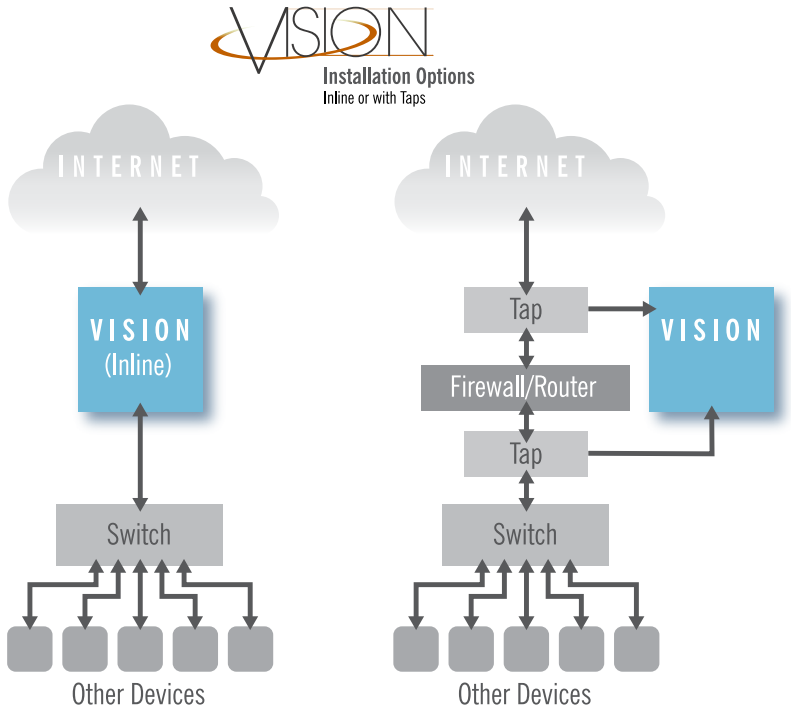
The product may duplicate or partially duplicate systems currently operated by the National Security Agency, in particular XKEYSCORE. However, this is unknown due to the extraordinary secrecy surrounding such systems. In any event, an export determination is pending (case no. CJ 0408-16).

Formed in 2012, Rællic Systems consists of one principal/owner, several contractors, and some associates and advisors. The

company is based in San Mateo, California, which is at the Northern end of Silicon Valley.

BACKGROUND

The Vision is a gigabit-class packet capture appliance, which I offer to anyone who wishes to capture traffic on one or more segments of a gigabit-level network. The device can be used inline or with passive copper taps:



The recommended setup is to tap the pre-firewall connection and the post-firewall connection simultaneously so the appliance captures attacks that make it through the firewall. The Vision comes with several different capture utilities, including the standard tcpdump, a little-known utility called gulp, the nTop suite, and Snort. Each of these applications has costs and benefits. For example, tcpdump has no built-in daemon mode, which would be ideal for The Vision, while Snort has appeared substantially more likely to result in packet loss due to the processing of IDS/IPS rules while capturing to disk. The Windows version of the gigabit-level appliance, which dual-boots Linux, is available for those who desire a more friendly setup and are willing to sacrifice some performance.

In any event, with this combination of capabilities, The Vision appears to be the most cost-effective gigabit-class capture appliance available to the public, at around $12,000 for the entry level system including installation and configuration. This figure can double or triple if the customer desires a faster or more secure system.

The first advertisement for The Vision will run in the August 2016 issue of Signal magazine.

THE VISION OMEGA

While looking at options for the 10G/40G version of the appliance, I thought I might as well skip to 100G because many of the challenges are the same. For example, capturing both directions of a saturated 10G connection requires sustained write to disk of 2.5 GB/s. To get that level of performance I would have to use a SSD array or have a RAID 0 of at least ten conventional hard drives, or a large RAM disk from which capture files are rotated out regularly. The SSD approach is undesirable because the service life of non-volatile RAM in a sustained write environment is exponentially less than spinning drives. The RAID 0 approach is undesirable because of the space requirements for ten conventional drives and the potential for data integrity problems when I have no hot spares. The RAM disk approach is undesirable because it only acts as a storage buffer to deal with bursts above what the hard drives can handle. With that preface, I turn to the strategy I came up with to solve these problems on both the 10G/40G and 100G versions of the appliance.

TIME DIVISION PACKET STEERING

Accolade Technology makes the FPGA-based NICs that are planned for this project. The Accolade NICs support built-in packet steering based on rules/filters. That is, the NIC determines what to do with a particular packet without burdening the host system. Options include (1) passing the packet to the host, (2) re-transmitting the packet out of another port on the NIC while bypassing the host, (3) dropping the packet, and other actions.

I propose a strategy in which each NIC passes packets to the host based on time division, rather than rules or filters, so that the host can send the received packets out of another NIC on the host during the host's limited time window receiving on the 100G stream. In this fashion, the stream of 100G traffic is divided among each cluster unit into small enough pieces that each cluster unit can capture to disk at a reasonable rate. I call this strategy time division packet steering because one second is divided into twenty sections of 50 ms each. This technique contrasts with a different strategy I am looking at that I call wavelength division packet steering: a prism splits an optical signal into color bands, each going to a separate capture device.

The Vision Omega requires one of two options in order to effectively implement time division packet steering. Both of the options are possible with the Accolade NICs and can be implemented using their software development kit. The first option is that every 20th packet that arrives on the control unit's 100G NIC is passed to one of the twenty 10G interfaces in the control unit on a round robin basis. The second option is that all packets arriving within a span of 1/20th of a second

(50 ms) are passed to one of the twenty 10G interfaces, again on a round robin basis. It could be as straightforward as setting a rule on the NIC, or it could require some development. Either way, the cost to implement this feature using Accolade's SDK is negligible compared to the total system cost.

It is also possible that some packets will be prioritized or segregated, or re-transmitted, based on rules. For example, the Accolade NICs can steer packets based on IP address range. The client could steer packets from the FBI (153.31.113.*) or other government agencies to a separate machine or system if desired and designate that machine for longer traffic replay. In addition, while passive monitoring is the primary method, with some modifications that are under consideration at Accolade it would be possible to set up The Vision Omega inline as a gigantic router that also captures all traffic-- a larger scale version of the gigabit-level appliance, which offers this capability.

TCPDUMP WRITE STRATEGY

The primary tools I expect to use are a modified version of tcpdump and the standard version of tcpslice. Modifications to tcpdump would include interface labels, a robust daemon mode (not merely using scripts to run tcpdump in the background), and single instance multi-interface capture. The standard version of tcpslice supports merging tcpdump binary pcap files by default, so that is really fortunate. I expect to set the tcpdump buffer to 2 GB and write to disk when the buffer fills or after 50 ms have passed, whichever is earlier. This way the write to disk is guaranteed to occur during the "off" portion of the capture while the other cluster units are capturing, and the unit has ample time to write to disk before going "on" again a full second later. The buffer should be flushed for each 50 ms of capture to ensure it does not fill while the machine is "on" the capture. A utility called gulp, which is similar to tcpdump but has some advantages, may also be useful.

DATABASE WRITE STRATEGY

There is no included database because writing to a database at the same time as capturing to disk would severely impact performance of the capture. This has the disadvantage of limiting the amount of traffic replay in proportion to the number of cluster units available, because the oldest files would be overwritten as disk space limits or specified capture size limits were reached. However, searchable longer-term storage is highly desirable. I have two alternate solutions to that problem. First, I could simply copy the capture files to another storage system by reading them from the capture system just before they get overwritten. Second, a more efficient solution I have come up with is (1) iterate over the binary capture files using a fast command line utility such as tshark to convert the capture data to ASCII, (2) pipe the output to other command line utilities or scripts such as grep or sed, and (3) insert sections of the

resulting ASCII stream into an extreme performance database running on another cluster. The goal is to permit longer-term storage and analysis of interesting traffic, which is accomplished by having the cluster units read, not write, during the capture. Oracle indicated in response to an inquiry that its fastest database, the Exadata, may be suitable for this application.

OPERATING SYSTEM

Linux seems to be the ideal choice for the control units, especially with the use of FPGA-based NICs to offload packet processing from the hosts. I would prefer to use PitBull from General Dynamics due to its numerous security features. In addition, I understand PitBull is certified to handle classified information, and I want to get The Vision Omega certified under the NSA's "Commercial Solutions for Classified" program. One key challenge would be the cost of PitBull, which would be prohibitive if used on all the cluster units. I could get PitBull for the two control units and use Red Hat Enterprise Linux for the cluster units, if the customer were prepared to spend that kind of money on the OS.

HARDWARE

I plan to use single processor 1U servers for the cluster units and two 4U dual-processor servers for the control units. Accolade makes the NICs I want to use. Garland Technology makes the taps I want to use. I have a systems integrator that I have previously worked with already lined up to actually build the system at very competitive rates.

The customer would be responsible for supplying electric power and cooling. As currently designed, the base system is 50 rack units in size, i.e., just under one and one quarter racks. Each additional storage cluster is a full rack.

COST

The Vision (the gigabit-level appliance) cost breakdown follows:

$	The Vision 1U capture device
$	Installation and configuration
$	Total (PitBull adds $

The Vision is priced at $12,000 for the base model, which is approximately half of competing systems and represents a small profit margin. The PitBull model is currently priced at $32,000.

As for The Vision Omega, I have quotes or estimates for each component, and the following figures show what I expect to spend on the base system.

$ The Vision Omega storage cluster (42 1U servers)
$ The Vision Omega control units (two 4U servers)
$ Multiple FPGA-based NICs from Accolade Technology
$ Standard NICs from Intel (estimated)
$ Passive fiber tap from Garland Technology (estimated)
$ Consulting and configuration services (estimated)
$ Cables and miscellaneous equipment

$ Total, not including PitBull (adds about $

In this region, I would need to charge at least $300,000 for the base configuration to make it worth my while, and that is bargain basement because I have low overhead. It would probably cost the government multiple millions of dollars to deploy this system from a conventional defense contractor.

The upgraded system cost breakdown follows:

$ The Vision Omega storage cluster (2)
$ The Vision Omega control units (3)
$ FPGA-based NICs from Accolade
$ Standard NICs from Intel
$ Passive fiber tap from Garland (estimated)
$ Consulting and configuration (estimated)
$ Cables and miscellaneous equipment

$ Total, not including PitBull (adds about $

The upgraded configuration requires a third control unit to split each direction of 100G traffic into two halves, each of which goes to ports on the other two control units. In other words, each port on each of the two control units receives 1/4th of the stream, 1/20th of which goes to each cluster unit. In this region, I would need to charge at least $550,000 for the upgraded system.

The database system would be tailored to meet the customer's requirements for long-term storage, and would probably cost 50% or 75% of what the storage cluster here costs.

POTENTIAL CUSTOMERS

The ideal customer is the U.S. government, an allied government, or a large enterprise. These customers have a need for the system as well the means to use it well. Frankly, I understand from the Snowden files that the U.S. government doesn't really need this system, so I would like to focus on allied governments.

EXPORTABILITY

The system does not use any proprietary hardware, but it may be determined to fall into several categories of the U.S. Munitions List because it is specially designed for intelligence gathering and may partially duplicate systems used by the U.S. government.

CONCLUSION

The ability to capture any and all network traffic at this level is an extraordinary, potentially explosive capability not previously available to the public. The customer should be fully vetted so that the system does not end up in nations with oppressive regimes as an instrument of control. As a starting point, I propose exporting The Vision Omega to NATO countries.

AUTHOR

Andrew Watters

COMPANY INFORMATION

Rællic Systems
[redacted]
(415) 261-8527
director@raellic.com

CAGE: 782B5
DUNS: 079548638

Andrew Watters
CEO

[redacted]
Programmer

[redacted]
IT Consultant

[redacted]
Research Scientist

[redacted]
Sales

VERSION HISTORY

1.4 -- July 13, 2016
Technical revisions.

1.3 -- July 12, 2016
Added longer-term storage/database strategy.

1.2 -- July 12, 2016
Updated cost breakdown for upgraded system.

1.1 -- July 11, 2016
Updated cost estimates and revised pertinent sections of the brief following feedback from Accolade.

1.0 -- July 9, 2016
Initial version.