Progress® Flowmon®

Transforming Your Network Operations with Enriched Flow Data

WHITEPAPER

Executive Summary

The digital transformation of corporations worldwide has led to all-pervading connectivity and high-speed data exchange with massive productivity growths. Network infrastructure plays such a fundamental role that ensuring its reliability and security is absolutely vital for every organisation today.

Yet networks are becoming increasingly abstract, understood in such metaphors as a 'public cloud' where we do not even think in terms of switches or routers anymore, let alone how to configure them by hand. Vendors of network equipment are making networking more seamless, and consequently, the expertise of engineers is slowly shifting from knowing how to use a screwdriver to how to secure a network with tens of thousands of IoT devices. SDN, NFV and virtualization goes hand-in-hand with this trend - enabling corporations to focus their resources on the core business instead of managing the network. This leads to a bandwidth explosion in corporate networks, which makes the traditional approaches of network monitoring appear to be a primitive tool. We at Progress are re-thinking the old concepts and endeavouring to tear down the walls that limit progress.

This whitepaper is about our belief that merging flow and packet level visibility into one versatile solution is the technology that will help you scale to future performance and capacity needs. The solution also retains detailed information about network traffic and presents the outputs in a straightforward and understandable way. You are about to discover how flow data, often perceived as a billing and top statistics tool, can, when mastered, fully replace full packet capture and analysis and provide never-seen-before and future-ready scalability.

The history of network monitoring: Packet analysis

Packet analysis looks inside communications to analyse its content. There is no aggregation, compression or trimming involved, and data is stored in its original size. Consequently, this method has extremely demanding performance and disc capacity requirements.

Just imagine the capturing of a network with 250 Mbps traffic on average. This equals a data load with over 31 MB per second, 1.8 GB per minute, 108 GB per hour and 2.6 TB a day. In the case of 10Gbps networks we reach hardly believable numbers - it would be more than 100 TB of stored data per day.

However, large volumes of data are not the only drawback. The principal limitation of packet analysis is encrypted traffic. Without the encryption key, we cannot understand the content of any transferred data, and often not even uncover the transfer protocol or application. Nevertheless, volumes of encrypted traffic are constantly growing.

Continual, full-scale traffic recording (full packet capture) demands the right technical equipment, especially high-speed storage arrays with adequate capacity. Such an approach to network monitoring is very expensive, suitable only for critical infrastructure and networks with a specific purpose. It must be underlined that storing such data may not be the only problem. Once the data is stored, any troubleshooting involves extensive mining of information that requires considerable experience and skillset. For the majority of network incidents outside of the scope of expensive continuous full packet capture, corporations rely on another approach.It is called on-demand packet capture. When employing this approach, we capture packets only when needed – typically when we deal with system compatibility issues – for instance, when discovering missing or damaged packets. On-demand packet capture is a simple method, affordable to every network administrator, but it does have its pros and cons. The limitation of this approach is that the administrator has to decide in advance which traffic should be stored. Consequently, there is no option to reach the traffic archive to get the right information for analysis when an incident occurs. It requires a network administrator to go to a physical location (e.g. server room) with their notebook, connect the notebook to a mirror port or TAP and to carry out the network traffic recording. Problems may arise when the location is far away, and also in optical network interfaces and 10Gbps infrastructures – limitations that could hardly be overcome with a notebook.

Although the need for capturing packets is not going away, the demand, however, is certainly shrinking. There is a clear scalability problem, especially because of the ever growing number of devices connected to a network or the number of applications and services delivered from cloud that require higher bandwidths. Solutions for full-scale traffic recording and analysis are very resource consuming and, consequently, expensive. Also, there are technology limits in a high-speed networking environment and restricted possibilities of use when traffic is encrypted.

The future of network monitoring: Enriched flow data

When it comes to network traffic monitoring, troubleshooting or threat detection, network engineers would rarely think there were two options at their disposal. The first one is full packet capture and analysis providing complete network visibility. The other is flow data.

Flow data is an abstraction of the network traffic itself. Flow statistics are created as an aggregation of the network traffic; using the source IP address, destination IP address, source port, destination port and protocol number as attributes that identify the individual flow records. The content of the communication is not stored, and the achievable aggregation rate is about 500:1. With the information listed above, we are able to analyse traffic structure, identify end-stations transferring large amounts of data or to troubleshoot network issues and wrong configurations. In other words, we can handle 80% of network incidents, as Gartner has reported since 2013.

It is evident that flow data does not contain enough information for some tasks. By contrast, as a result of packet analysis, the IT department is usually overloaded with barely manageable volumes of detailed data. When we combine both perspectives and extend traditional flow data with information from the application layer, we can get the appropriate detail, providing an insight into data communication, flexible reporting and effective troubleshooting of operational issues and the automatic detection of security incidents. This approach is called enriched flow data, leveraging flexibility of the ***IPFIX protocol***. In our experience, we are now able to handle 95% of network incidents with the most scalable, cost-efficient and easy-to-use solution based on flow data.

The best known implementation of this technology is Cisco's NBAR2 (Next Generation Network-Based Application Recognition). Flow data monitoring is combined with continual packet analysis that extends the traffic statistics with the name of an application or application protocol. Based on this information, modern ***flow collectors*** enable traffic reporting and analysis.

One of the most widespread communication protocols is HTTP, or its encrypted version HTTPS. Today it is used to provide access to websites, but this is not its sole function. The protocol is also the basis of communication between the components of business systems, or applications working with sensitive data (e.g. electronic banking). By identifying this transfer protocol, we can extend flow data statistics by fundamental HTTP request attributes – a hostname or URL information. Thanks to SNI (Server Name Indication), we can get hostname information even when the HTTPS protocol is used. Similarly, we can get other information from HTTP communication; for example, the operating system and its version, the identification of a browser and its version or a device type in the case of mobile phones. And this is only one example of many protocols for which we can use L7 information without the necessity of manual data mining.

Nonetheless, flow data can be enriched by something perhaps even more powerful in the modern world - Network Performance Monitoring (NPM). NPM metrics can significantly help with network performance troubleshooting. Using the Server-Response-Time and

Round-Trip-Time metrics, we can distinguish between delays in the network infrastructure (e.g. a malfunctioning access point), from delays in the server (e.g. insufficient HW resources). This kind of information is crucial for fast network troubleshooting. The delay and jitter metrics should interest us especially when we use VoIP calls or video conferences, as they can indicate bad audio and video quality. When we are talking about transferring large volumes of data, we are interested mainly in the number of TCP retransmissions, which can indicate problems on the physical layer (e.g. interference, faulty port) and a lower bit rate, or out-of-order packets, which can signpost failures in communication links.

And when this level of information is still insufficient, ***Progress® Flowmon®*** allows triggering of on-demand full packet capture. This can be done manually or automatically on a detected event. When this happens, the capturing filter is determined autonomously by the system, narrowing the captured data volume to an absolute minimum, retaining only the relevant part of the traffic. This can obviously be done remotely in any part of the network and at speeds of 10Gbps; unachievable with Wireshark or WinPCAP.

The benefits of ***flow data monitoring*** and application layer analysis are clear. We get more detailed information about data communication, better capabilities for traffic analysis. At the same time we retain the excellent compression rate of network traffic statistics vs. original traffic volume to scale to multiple hundred-Gigabit networks. Also, the system aggregates the most important information, so it can be delivered at a glance, avoiding the necessity of manual data mining with packet analysis. This leads to a tremendous reduction of Mean-Time-To-Resolve while lowering the skillset required to use the solution. With Flowmon it is always possible to carry out full-scale traffic recording if necessary using the same platform.

Why choose flow over packet? Business and technical benefits explained

In the earlier part of this whitepaper we identified differences between continuous packet capture and flow data in the context of using these technologies to successfully monitor network traffic. Let us summarize the benefits of enriched flow data over packet capture:

- Budget requirements on packet analysis technologies rarely allow monitoring of the entire network traffic. So, it is deployed only to monitor top critical systems as opposed to flow, where covering the entire corporate network traffic, including data centres and cloud is a standard scheme.

- Troubleshooting in general is not conducted in real time. In corporations, it often takes days until a reported incident is reviewed by a network administrator. With only a limited data retention period, retrospective analysis would be impossible. Flow monitoring can easily provide retention of weeks or months, so you can easily prioritise your work and focus on retrospective analysis whenever you have finished more important tasks.

- Seamless deployment, integration with existing network equipment, compatibility with a wide range of flow sources, fast training of administrators are among many reasons it is so easy to introduce flow technology to your network and get an instant value.

- The level of detail provided by packet analysis makes it suitable for deep-dive forensic analysis of persistent problems. Companies turn to Flowmon to minimize the time needed for root cause analysis and gain more time for root cause remediation using human readable dashboards, context aware presentations and drill-down capabilities.

- Too detailed granularity of packet analysis means higher costs, lower scalability and a much higher required skillset to use. However, only a small percentage of the captured data is relevant. Enriched flow data on the other hand keeps the most interesting and important information so that 95% of network incidents can be resolved with it. Additionally, Flowmon enables on-demand full packet capture for the rest of cases.

- Packet analysis had been built with unlimited visibility in mind. It is well suited for time consuming forensics of persistent problems. To restore business-as-usual promptly, corporations are turning to Flowmon to be provided with analytical workflows and automation to help streamline resolution.

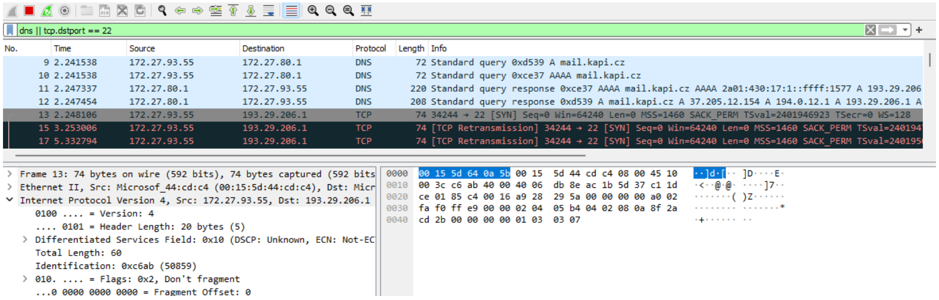
- With more and more traffic being encrypted, packet analysis becomes useless. While exporting Flow data from encrypted traffic, Flowmon focuses on non-encrypted IP headers that will help to resolve 80% of incidents. Additionally, it uses different techniques to extract information from the application layer otherwise hidden to an inexperienced eye.

- Public cloud providers do not allow tapping into their network to enable full packet analysis. However, both cloud providers and virtual hypervisors often export some form of flow data compatible with Flowmon, enabling seamless deployment of quality network monitoring.

Use cases

Troubleshooting using Packet Capture

I have my packet analyzer with continuous capture in place. So, let us hope that my rolling buffer still keeps the data I need. Thankfully, we can download the PCAP containing the traffic of IP address 193.29.206.1 and open the traffic in Wireshark.

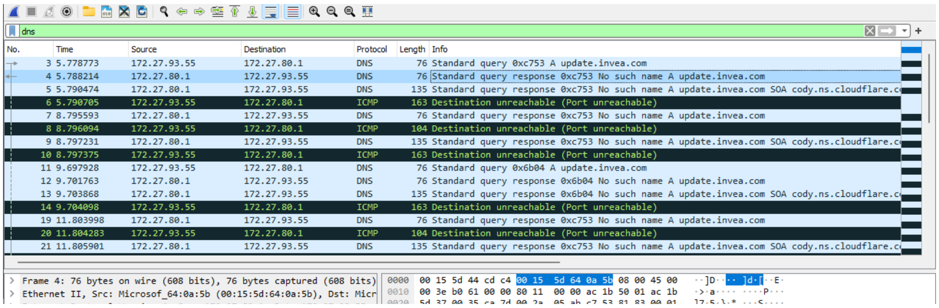


```
dns || tcp.dstport == 22

No.  Time      Source         Destination    Protocol  Length Info
   9 2.241538  172.27.93.55   172.27.80.1    DNS         72 Standard query 0xd539 A mail.kapi.cz
  10 2.241538  172.27.93.55   172.27.80.1    DNS         72 Standard query 0xce37 AAAA mail.kapi.cz
  11 2.247337  172.27.80.1    172.27.93.55   DNS        220 Standard query response 0xce37 AAAA mail.kapi.cz AAAA 2a01:430:17:1::ffff:1577 A 193.29.206
  12 2.247454  172.27.80.1    172.27.93.55   DNS        208 Standard query response 0xd539 A mail.kapi.cz A 37.205.12.154 A 194.0.12.1 A 193.29.206.1 A
  13 2.248106  172.27.93.55   193.29.206.1   TCP         74 34244 → 22 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM TSval=2401946923 TSecr=0 WS=128
  15 3.253006  172.27.93.55   193.29.206.1   TCP         74 [TCP Retransmission] 34244 → 22 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM TSval=240194
  17 5.332794  172.27.93.55   193.29.206.1   TCP         74 [TCP Retransmission] 34244 → 22 [SYN] Seq=0 Win=64240 Len=0 MSS=1460 SACK_PERM TSval=240195

> Frame 13: 74 bytes on wire (592 bits), 74 bytes captured (592 bits
> Ethernet II, Src: Microsof_44:cd:c4 (00:15:5d:44:cd:c4), Dst: Micr
v Internet Protocol Version 4, Src: 172.27.93.55, Dst: 193.29.206.1
    0100 .... = Version: 4
    .... 0101 = Header Length: 20 bytes (5)
  > Differentiated Services Field: 0x10 (DSCP: Unknown, ECN: Not-EC
    Total Length: 60
    Identification: 0xc6ab (50859)
  > 010. .... = Flags: 0x2, Don't fragment
    ...0 0000 0000 0000 = Fragment Offset: 0

0000  00 15 5d 64 0a 5b 00 15  5d 44 cd c4 08 00 45 10   ··]d·[·· ]D····E·
0010  00 3c c6 ab 40 00 40 06  db 8e ac 1b 5d 37 c1 1d   ·<··@·@· ····]7··
0020  ce 01 85 c4 00 16 a9 28  29 5a 00 00 00 00 a0 02   ·······( )Z······
0030  fa f0 ff e9 00 00 02 04  05 b4 04 02 08 0a 8f 2a   ········ ·······*
0040  cd 2b 00 00 00 00 01 03  03 07                     ·+······ ··
```

Figure 1: Captured network traffic analysis in Wireshark.

In the figure above we can see the communication between user and mail.kapi.cz. Domain mail.kapi.cz is resolved correctly to IP address 193.29.206.1, but after the DNS response was received, a user tried to establish a TCP session with no response from the external IP address. We have to check our firewall settings if this communication is allowed.

The second issue is related to a non-existent domain queried by the user machine. We can see that update.invea.com does not exist, which probably implies a wrong configuration of the user and not a network related issue.

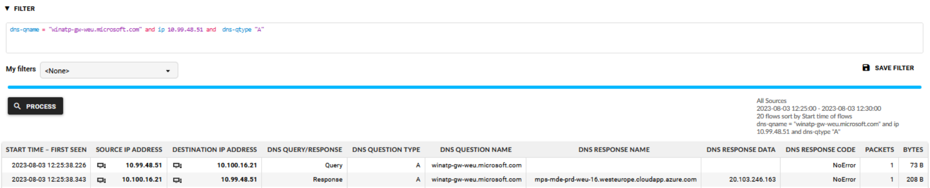


```
dns

No.  Time       Source         Destination    Protocol  Length Info
   3 5.778773   172.27.93.55   172.27.80.1    DNS         76 Standard query 0xc753 A update.invea.com
   4 5.788214   172.27.80.1    172.27.93.55   DNS         76 Standard query response 0xc753 No such name A update.invea.com
   5 5.790474   172.27.80.1    172.27.93.55   DNS        135 Standard query response 0xc753 No such name A update.invea.com SOA cody.ns.cloudflare.c
   6 5.790705   172.27.93.55   172.27.80.1    ICMP       163 Destination unreachable (Port unreachable)
   7 8.795593   172.27.80.1    172.27.93.55   DNS         76 Standard query response 0xc753 No such name A update.invea.com
   8 8.796094   172.27.93.55   172.27.80.1    ICMP       104 Destination unreachable (Port unreachable)
   9 8.797231   172.27.80.1    172.27.93.55   DNS        135 Standard query response 0xc753 No such name A update.invea.com SOA cody.ns.cloudflare.c
  10 8.797375   172.27.93.55   172.27.80.1    ICMP       163 Destination unreachable (Port unreachable)
  11 9.697928   172.27.93.55   172.27.80.1    DNS         76 Standard query 0x6b04 A update.invea.com
  12 9.701763   172.27.80.1    172.27.93.55   DNS         76 Standard query response 0x6b04 No such name A update.invea.com
  13 9.703868   172.27.80.1    172.27.93.55   DNS        135 Standard query response 0x6b04 No such name A update.invea.com SOA cody.ns.cloudflare.c
  14 9.704098   172.27.93.55   172.27.80.1    ICMP       163 Destination unreachable (Port unreachable)
  19 11.803998  172.27.80.1    172.27.93.55   DNS         76 Standard query response 0xc753 No such name A update.invea.com
  20 11.804283  172.27.93.55   172.27.80.1    ICMP       104 Destination unreachable (Port unreachable)
  21 11.805901  172.27.80.1    172.27.93.55   DNS        135 Standard query response 0xc753 No such name A update.invea.com SOA cody.ns.cloudflare.c

> Frame 4: 76 bytes on wire (608 bits), 76 bytes captured (608 bits)
> Ethernet II, Src: Microsof_64:0a:5b (00:15:5d:64:0a:5b), Dst: Micr

0000  00 15 5d 44 cd c4 00 15  5d 64 0a 5b 08 00 45 00   ··]D···· ]d·[··E·
0010  00 3e b0 61 00 00 80 11  00 00 ac 1b 50 01 ac 1b   ·>·a···· ····P···
```

Figure 2: Host queries to resolve non-existent domain.

Troubleshooting using Enriched Flow Data

I have Flowmon Probes in place monitoring network and Flowmon Collector with weeks of history of un-sampled and non-aggregated traffic statistics. So, I can directly ask about the DNS question to domain winatp-gw-weu-microsoft.com.



```
dns-qname = "winatp-gw-weu.microsoft.com" and ip 10.99.48.51 and  dns-qtype "A"
```

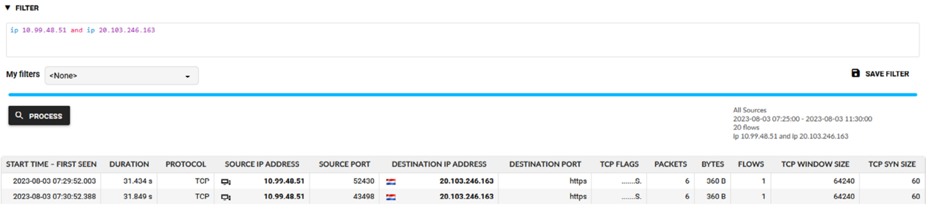
My filters <None>

All Sources
2023-08-03 12:25:00 - 2023-08-03 12:30:00
20 flows sort by Start time of flows
dns-qname = "winatp-gw-weu.microsoft.com" and ip 10.99.48.51 and dns-qtype "A"

START TIME - FIRST SEEN	SOURCE IP ADDRESS	DESTINATION IP ADDRESS	DNS QUERY/RESPONSE	DNS QUESTION TYPE	DNS QUESTION NAME	DNS RESPONSE NAME	DNS RESPONSE DATA	DNS RESPONSE CODE	PACKETS	BYTES
2023-08-03 12:25:38.226	10.99.48.51	10.100.16.21	Query	A	winatp-gw-weu.microsoft.com			NoError	1	73 B
2023-08-03 12:25:38.343	10.100.16.21	10.99.48.51	Response	A	winatp-gw-weu.microsoft.com	mps-mde-prd-weu-16.westeurope.cloudapp.azure.com	20.103.246.163	NoError	1	208 B

Figure 3: Filtering in DNS traffic – flow statistics enriched with DNS information.

I can see the IP address provided by the DNS server as a response and check for the traffic going to that IP address easily. It is possible because flows from the Probe corresponding to DNS are enriched with the most important L7 information from the DNS protocol.



```
ip 10.99.48.51 and ip 20.103.246.163
```

My filters <None>

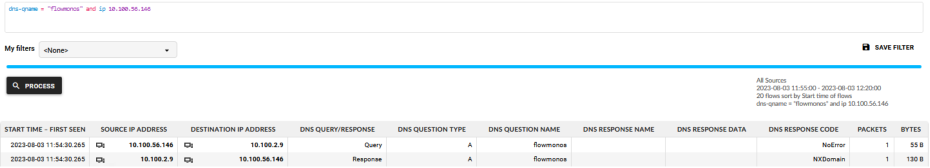
All Sources
2023-08-03 07:25:00 - 2023-08-03 11:30:00
20 flows
ip 10.99.48.51 and ip 20.103.246.163

START TIME - FIRST SEEN	DURATION	PROTOCOL	SOURCE IP ADDRESS	SOURCE PORT	DESTINATION IP ADDRESS	DESTINATION PORT	TCP FLAGS	PACKETS	BYTES	FLOWS	TCP WINDOW SIZE	TCP SYN SIZE
2023-08-03 07:29:52.003	31.434 s	TCP	10.99.48.51	52430	20.103.246.163	https	.......S.	6	360 B	1	64240	60
2023-08-03 07:30:52.388	31.849 s	TCP	10.99.48.51	43498	20.103.246.163	https	.......S.	6	360 B	1	64240	60

Figure 4: Filtering user's communication with external service. Host is not receiving any response.

I see that only SYN packets are being transmitted to the network with no response from an external IP address, which implies the need to check firewall rules. Besides common L3/L4 information, I have visibility into TCP specific items such as default TCP segment (window) size, which can help me to troubleshoot the TCP session.

The non-existing domain is in my flow evidence as well. The domain flowmonos does not exist, so the DNS server replies „NXDomain".



```
dns-qname = "flowmonos" and ip 10.100.56.146
```

My filters <None>

All Sources
2023-08-03 11:55:00 - 2023-08-03 12:20:00
20 flows sort by Start time of flows
dns-qname = "flowmonos" and ip 10.100.56.146

START TIME - FIRST SEEN	SOURCE IP ADDRESS	DESTINATION IP ADDRESS	DNS QUERY/RESPONSE	DNS QUESTION TYPE	DNS QUESTION NAME	DNS RESPONSE NAME	DNS RESPONSE DATA	DNS RESPONSE CODE	PACKETS	BYTES
2023-08-03 11:54:30.265	10.100.56.146	10.100.2.9	Query	A	flowmonos			NoError	1	55 B
2023-08-03 11:54:30.265	10.100.2.9	10.100.56.146	Response	A	flowmonos			NXDomain	1	130 B

Figure 5: Filtering in DNS traffic – searching for NXDomain DNS Response Code

Testimonial: Root cause investigation workflow

Let us imagine a department of Level 3+ engineers in a 50 thousand people bank. These engineers focus on root cause analysis of network incidents that no other team before could resolve. For example, to figure out why a VPN connection between the customer and the bank, both on different continents, had outages. It is very common that these engineers spend many hours digging in Wireshark through petabytes of data generated by hundreds of different systems in a complicated and heterogeneous environment. And this is exactly a situation of one Flowmon customer who asked us to deliver an alternative and more effective solution for dealing with operational incidents.

It is important to say that our customer built a whole platform to help them with network operational tasks. It was based on a commercial tool for continuous packet capture, a customized open source software for flow monitoring, and an SNMP based tool showing data transfer heat maps in real time.

Soon it became clear that maintaining, supporting and upgrading the custom solution to fit everyday operations was too expensive and time-consuming. So, they sought NetFlow/IPFIX technology to replace the original solution. The bank's IT department looked for a solution to completely replace the original one. Although budget was not an issue, the choice was not as easy as it first appeared. Testing different vendors, sometimes the problem was data aggregation, sometimes no virtualization, but always it was the slowness in terms of the time needed to provide measured results.

Their dream solution should:

- provide not only context-aware top-level dashboards, but one that allows manual drill down to any flow.
- not aggregate stored data and to keep the raw flows for as long as the storage lasts.
- not necessarily be dependent on its own sensors, since they cannot afford to be locked into a single technology due to their heterogeneous environment.
- be virtualized, so management and migration is as flexible as possible.
- combine flow monitoring with on-demand full packet capture.

- provide, most importantly, outputs of the measured statistics of the flow data faster than the platform they built themselves ten years ago from an open source tool.

Then the team came across Flowmon, which fully fitted their needs. Since the proof-of-concept project, Flowmon has become fundamental in their tool set. It has become the root cause investigation workflow itself. The engineers now start with the dashboard, go into top-level statistics, deeper onto levels of ***NetFlow*** and then only focus on a small part of the traffic where they can run full packet capture.

> **“Flowmon gives KBC a great overview of the dataflow metrics in the network so that the network health can be easily assessed. In case of an issue, the tool allows very fast and efficient troubleshooting by visualizing the traffic that is causing the problem.”**
>
> **Marc Deamen**
> Senior Systems Engineer, KBC

Conclusion

The dynamics and diversity of today's networks challenge the prevailing network monitoring approach. Facing an increase in network speeds, visibility gaps caused by migration to cloud, IoT and software defined networking, packet capture solutions struggle to bring expected results promptly and at a reasonable price.

Packet capture solutions were designed at the time when the dynamics of today's network environments would have been hard to imagine. Nowadays, they work well in specific use cases, but they cannot cope with the flexibility, scalability and ease of use of flow data in most of the everyday use cases that network engineers face.

We have demonstrated a use case where flow data with extended visibility is equally powerful to full packet capture and packet analysis. On the other hand, it is fair to say that even with extended flow level visibility you still might face issues where analysis of PCAPs is unavoidable.

We in Progress® Flowmon® believe that merging flow and packet level visibility into one versatile solution is the technology that will help you scale to future performance and capacity needs. So, let's do continuous flow monitoring and packet capture when needed. At the end of the day, you will most probably need to analyze PCAPs less than you expect. Go and see how it can help to your organisation.

Request your FREE trial of Flowmon for 30-days

About Progress

Dedicated to propelling business forward in a technology-driven world, Progress (NASDAQ: PRGS) helps businesses drive faster cycles of innovation, fuel momentum and accelerate their path to success. As the trusted provider of the best products to develop, deploy and manage high-impact applications, Progress enables customers to build the applications and experiences they need, deploy where and how they want and manage it all safely and securely. Hundreds of thousands of enterprises, including 1,700 software companies and 3.5 million developers, depend on Progress to achieve their goals—with confidence. Learn more at www.progress.com

/progresssw
/progresssw
/progresssw
/progress-software
/progress_sw_