incognito®

WHITE PAPER

IPDR vs. DPI: The Battle for Big Data

An Incognito White Paper
January 2016

IPDR vs. DPI: The Battle for Big Data

Contents

Summary

Gaining accurate network and subscriber data has had numerous benefits for service providers hoping to improve internal processes and enhance their customers' quality of experience. As a result, Big Data collection platforms have quickly become one of the most important aspects within broadband service provider operations. In a recent Dell report on Big Data, 96% of organizations surveyed had an existing Big Data initiative, or planned to start one in the near term[1].

In the same Dell report, 89% of respondents with a Big Data initiative in progress reported significant improvements in their company's decision making[2]. These results make one thing clear: in the digital information age, broadband service providers must collect and analyze network and subscriber usage data to stay successful.

Some of the benefits that service providers gain from gathering accurate network usage intelligence include:

- Preventing network traffic congestion by enhancing insight into peak service consumption periods
- Enforcing fair-access policies to reduce unfair usage and over consumption of network resources
- Providing insightful data for product launches and marketing campaigns
- Creating new revenue streams enabled by usage-based billing models
- Accurately forecasting future capacity requirements to maintain high subscriber quality of experience

The advantages are numerous and clear, but how operators can best go about extracting data from their networks to achieve these benefits has become a topic of some debate.

IPDR vs. DPI

There are two mainstream approaches for network and subscriber data collection, Internet Protocol Detail Record (IPDR), and Deep Packet Inspection (DPI). While both of these methods have supporters across major markets, and both can achieve the benefits highlighted above to some

1 http://www.dell.com/learn/us/en/uscorp1/press-releases/2014-04-28-dell-software-big-data-midmarket-survey

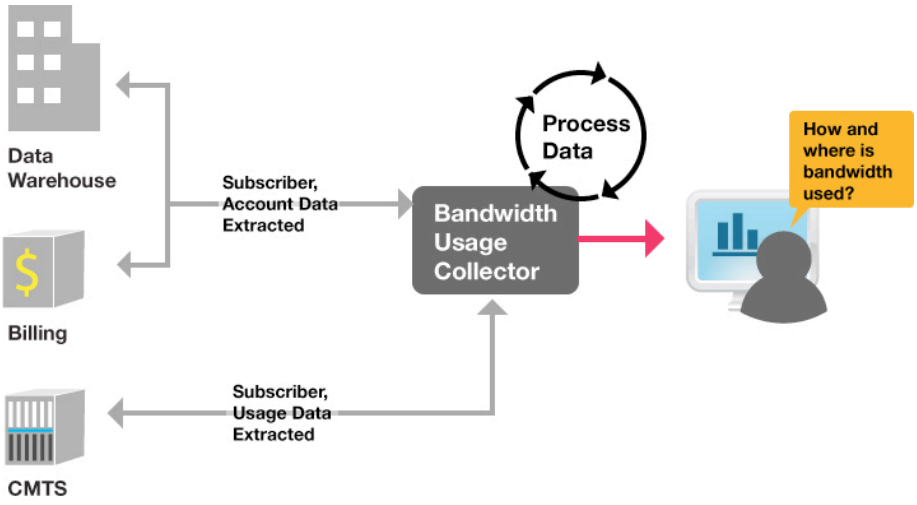
2 http://www.dell.com/learn/us/en/uscorp1/press-releases/2014-04-28-dell-software-big-data-midmarket-survey

degree of similarity, there are advantages and drawbacks to consider for each.

Overview of IPDR

IPDR is standard DOCSIS software-based technology used to collect and record network data traffic statistics from a CMTS. IPDR data contains information about every flow inside a CMTS and provides details of consumption usage for each subscriber device connected to the network. Usage details exposed by IPDR provide a view over subscriber and resource utilization and can be categorized by service type. IPDR produces this data without being intrusive to a subscriber's privacy because the level of detail exposed shows only what type of service is being consumed and not the specific piece of content.

Figure 1: The IPDR method extracts subscriber account details and usage data and sends it to a collector and processing



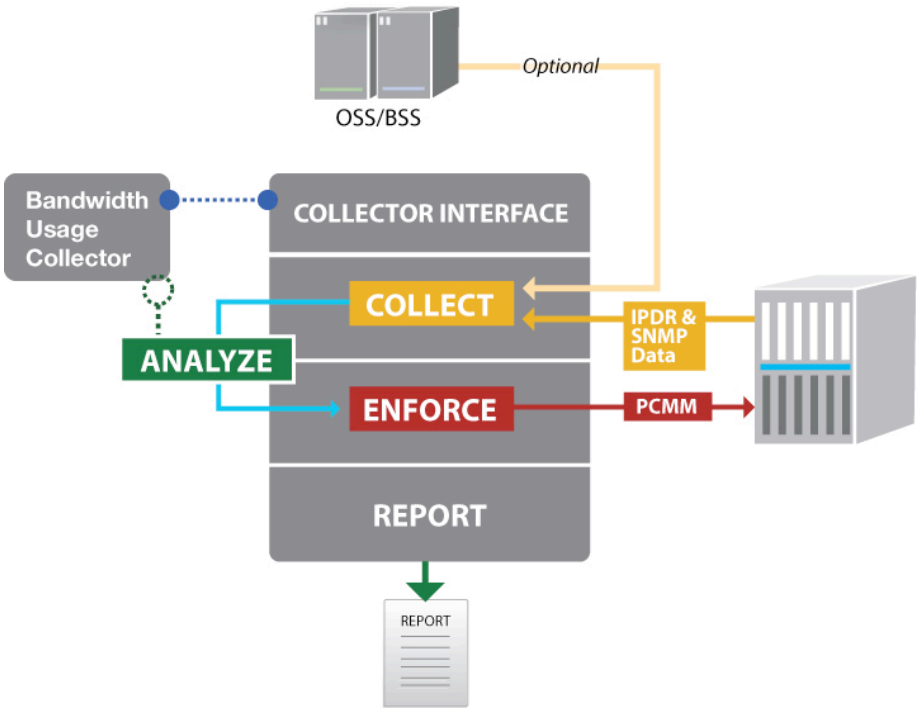
Conventionally, IPDR has predominantly been used to gather statistics to convert into billing events for usage-based policies. The information is gathered at regular intervals by a collector at low-impact to a CMTS. The data is then converted using the user database and DHCP lease information to be sent into an accounting database where it is aligned with subscriber service-level agreements. Beyond this, operators can use collected IPDR data to gain accurate information about their networks and subscribers, providing the insight required for:

- Bandwidth monitoring
- Policy enforcement
- Usage-based billing
- Capacity forecasting and network planning
- Marketing and sales campaigns

With IPDR collection, the CMTS is instructed to define, collect, encode, transport, and exchange cable modem-based usage records. These statistics are periodically and asynchronously reported to a central application called a collector, where data can be filtered, analyzed, and normalized by a solution for other departments to digest. IPDR information is collected and communicated out-of-band. Each CMTS vendor can implement their own

algorithms to assign the frequency at which gathered data is sent to the IPDR collector. This guarantees that IPDR will not interfere with normal service operations.

Figure 2: Once IPDR data is collected and analyzed. It can be used to enforce policies and produce customized reports



Using this method, IPDR offers an efficient and cost-effective approach for gathering and producing network and subscriber analytics. The benefits of IPDR data collection include:

- Essential non-intrusive insight into service consumption data for every device on a network
- Industry standard application-agnostic technology
- Collects data at low-impact to regular CMTS operations
- Highlights network topology when cross-referenced with SNMP data
- Little additional hardware required for implementation
- Cost efficient method

The drawbacks of IPDR:

- There are some discrepancies in interpretation of the standard between CMTS vendors
- Level of details exposed are limited, to protect subscriber privacy*
- IPDR is device centric and not subscriber centric

*In many regions, protecting subscriber privacy with non-intrusive data gathering is seen as a benefit to the IPDR approach

Overview of DPI

DPI is an application-based data collection method that sits between the subscriber and the Internet and has traditionally been used to throttle certain types of traffic on a network. DPI scans all network traffic and analyzes unencrypted data within a packet to create statistics based on specific parameters. It is a proprietary solution, and every DPI vendor implements different features. Usage details exposed by DPI show exactly what type of content or service a subscriber used, and all associated details. This deep level of insight has brought up privacy concerns in some regions.

DPI allows the network to discover what applications are using bandwidth and then depending on set policies can limit the amount of bandwidth used. Some providers use DPI to limit peer-to-peer (P2P) traffic and over-the-top services. Most analysis performed by DPI is done in-line with regular network traffic and is usually an appliance-based application sitting between every subscriber and the Internet. Flows are created internally and traffic policies are then applied. Data can be accessed later through an API. Because DPI is an in-line process, information is gathered from every single packet produced on the network. This means that using the DPI method is high-impact to a CMTS, creating a risk of affecting regular broadband services.

DPI provides network intelligence for:

- Insight into subscriber Internet usage and practices
- Policy enforcement to throttle certain types of traffic
- Data for marketing and sales campaigns

Many providers in North America use DPI to find and stop illegal download activity, but the level of detail gathered by DPI is sometimes viewed as controversial. Because each packet is inspected and analyzed, there is a growing sentiment in the user community that they are being spied on. Depending on the jurisdiction — and where privacy as well as net neutrality laws stand — this concern is more or less relevant to operators.

The benefits of DPI:

- Offers deep level of details about subscriber Internet usage and practices
- Allows service throttling on users who exceed their bandwidth quotas
- Can be used for network forensics and lawful intercepts

The drawbacks of DPI:

- Expensive and complex, usually requiring additional hardware
- Generally seen as too intrusive in the user community
- Not efficient for periodically reporting network statistics without an additional API
- Collects information in-line, which risks affecting regular network services

The table on the following page summarizes the information above and highlights the benefits and drawbacks of IPDR and DPI data collection:

	Internet Protocol Detail Record (IPDR)	Deep Packet Inspection (DPI)
What is it?	• Software-based technology used to collect and record network data traffic statistics from a CMTS • Contains information about every flow inside a CMTS • Provides details of consumption usage about subscriber devices on a network	• Application-based data collection method that sits between the subscriber and the Internet • Scans all network traffic and analyzes unencrypted data within a packet to create statistics based on specific parameters
Purpose	Provides network intelligence for: • Bandwidth monitoring • Usage-based pricing models • Capacity forecasting • Network planning • Marketing and sales campaigns	Provides network intelligence for: • Policy enforcement and network intelligence • Throttling certain types of network traffic to lower speeds (such as P2P file sharing)
How it works	• CMTS is instructed to define, collect, encode, transport, and exchange cable modem-based usage records • Statistics are periodically and asynchronously reported to a central application called a collector, where data is filtered and analyzed for other departments to digest	• Allows the network to discover what applications use the most bandwidth and depending on set policies limits bandwidth usage • Gains intelligence by examining every data packet as it passes an inspection point • Flows are created internally and traffic policies applied and data can be later accessed through an application
Benefits	• Application-agnostic • Essential service consumption data about every device on a network • Cost-effective • Industry standard, integrated into the DOCSIS protocol • Highlights network topology when cross-referenced with SNMP data • Little additional hardware required for implementation	• Provides a high level of detail about subscriber Internet usage • Can be used for policy enforcement and to throttle certain traffic • Can be used for network forensics and lawful intercepts
Drawbacks	• There are some differences in interpretation of the standard between CMTS vendors • Level of details exposed do not show what content a subscriber is viewing, only what type of content.* • IPDR is device centric and not subscriber centric	• Expensive and complex • Generally requires additional hardware integration • Level of details exposed is seen as too intrusive in the user community • Can burden the CMTS and risk affecting regular services

*In many regions, protecting subscriber privacy with non-intrusive data gathering is seen as a benefit to the IPDR approach

The case for IPDR vs. DPI

Usage Collection:
While both methods of data collection will provide similar intelligence for service providers, collecting network and subscriber data should never impact network performance. Subscriber quality of experience is the most crucial element when remaining successful in a competitive industry landscape. Because IPDR data is collected out-of-band, information can be normalized and compressed at low-impact to CMTS hardware. IPDR is a standard, and therefore several CMTS vendors can report the same type of information to one collector. Unlike DPI, IPDR is not intrusive. This is because it only measures packets sent and received — rather than inspecting the actual contents of the packet — and information is automatically streamed from the CMTS to an external collector at regular intervals.

Network Analysis:
Embedded in the IPDR data, you will find information about the actual interfaces and ports where the data is being pulled or pushed. Since IPDR data is local to a single CMTS, you can isolate a single subscriber and/or groups of subscribers per interface per CMTS. Such information is key in network planning and congestion prevention. Simply knowing the top users on the network is not enough. You must know where they are located and how many of them are accessing the same port or CMTS at the same time. DPI contains no data relevant to each subscriber besides the IP address, so you cannot infer anything else from the numbers. Operators know the throughput in a central part of the network but cannot see closer to the edge where problems may actually occur.

Costs:
DPI appliances usually run proprietary networking hardware in order to achieve the speeds required to run. IPDR is a software based approach, meaning data can be collected and analyzed at a much cheaper cost. The IPDR collector software runs on a separate server and you can typically run hundreds of CMTS units on a single server. The software runs on affordable commodity hardware which is easily sourced by the provider, and the data is already produced free of charge on any CMTS.

Conclusion

Due to the cost savings, ease of use, and level of detail gained, IPDR has been gaining traction as the better choice for Big Data collection and usage-based billing. It offers information that can be used for network engineering and trending analysis as well as bandwidth congestion management and network optimization. IPDR is also an easy add on and can be tested as needed by the end-user.

Although DPI could be used as a platform to perform the basic Big Data functions attributed to IPDR, it does so at a very high price and is often seen as very intrusive — increasing the risk of subscriber churn amidst growing privacy concerns — and is no longer acceptable in some regions. Many companies in regions where perceived privacy violations can affect subscriber churn rates have opted to remove their DPI platforms and simply enable IPDR software to accomplish usage-based billing and network traffic analysis. In other regions, IPDR can be run in parallel or in place of DPI as a

more cost-effective, less hardware-intensive approach to complete any of the functions enabled by accurate Big Data analytics.

Incognito Software Systems Inc.
Phone: 604.688.4332 or US/Canada toll free 1.800.877.1856
Fax: 604.688.4339
Email: solutions@incognito.com
Web: www.incognito.com