



# Web Traffic Capture

*Capture your web traffic – filtered and transformed, ready for your applications – without web logs or page tags and keep all your data inside your firewall.*

**Metronome Labs LLC**  
5401 Butler Street, Suite 200  
Pittsburgh PA 15201  
(412) 408-3167  
[www.metronomelabs.com](http://www.metronomelabs.com)

---

# Web Traffic Capture

## Understanding your web visitors

Understanding your web visitors is critical to the effectiveness of your site. Never before has there been such a wealth of data on the way your visitors interact with your website and react to your sales and marketing strategies.

Web analytics is becoming increasingly important for companies that sell or market through the web. In essence, web analytics packages are simply data management and reporting tools. What differentiates them is the way they collect the data. Initially, data was obtained from server log files and this is still a popular method. But log files do not give the whole story, so page tags are now in vogue. Page tags provide more information about your visitors but the data is often sent to a third party site which raises concerns about security and privacy. Because your web data is in a remote site it is difficult to correlate with your in-house sales and marketing databases.

But there is a better way. Metronome *Capture*™ traffic collection provides the richness of tag data with the security of log data inside your firewall with no changes to your website.

## Web traffic overview

When a visitor goes to your website, his browser sends an HTTP request packet. This is routed over the Internet to your server, which then replies with an HTML page carried by the HTTP protocol. On busy sites which have many servers, a load balancer routes the request to the least busy server. When the visitor's browser receives the HTML page, it loads it and reads all the links it contains to request graphics, style-sheets, etc.

Each page request from a particular visitor may be directed by the load balancer to a different server so the pages for one visitor session may be served by many different servers in your server farm.

## Logs and tags

### *Log files*

All web servers output log files although the actual content may differ slightly. They contain information about your visitor but the data is essentially about what the server is doing. If you use server log files to track your visitors, your analytics software has to gather the logs from all your servers, merge them together and then try to organize the page views and hits into visitor sessions. Server logs contain the hits for graphics which are usually uninteresting and so

---

Web logs need extensive filtering and processing to be useful.

They slow your servers down and do not really tell you what the visitor is experiencing.

Page tags become increasingly costly to maintain as your analytics needs grow.

Managed services send your data to a remote site where it is difficult to correlate with your enterprise databases.

there is a huge amount of additional data that must be filtered out. All of this takes a lot of time and expensive computer power and storage. Typically, processing is performed each night so you have to wait a day to get your information.

Web logs miss important data because servers do not see the underlying network protocol and they do not know when the page they sent actually got there. They don't know when it is complete with all its objects loaded ready to view. A web log does not show that a visitor clicked to a different page while the first page is on its way.

Outputting a web log slows your servers down and reduces your site capacity. If you can turn web logs off, you can save money on server hardware and software.

But one advantage of web logs is that the data they collect is secure inside your firewall and can be joined with your enterprise sales and marketing data to get a more complete view of your visitor.

### ***Page tags***

Page tagging is now in vogue. It works by placing a one-pixel dummy graphic on a page. The visitor's browser will request this dummy graphic from a special tag server which will store the request and associated data in a log which can then be used for analytics. Typically, the page has a script embedded in it that will gather information about the visitor's machine and add it as parameters to this request. The request is usually directed to a third-party managed site where the parameters are collected in that site's web server logs and then processed into a data warehouse. The data can then be viewed over the Internet through a portal.

Page tags are essentially visitor oriented and tell you much more about what your visitors are doing. Because the tags are operating from the visitor side, it is easier to relate the page views to visitor sessions and eliminate all the unwanted hits for graphic objects. There is less post-processing, so the data may be available sooner. In practice, sites have many pages that are changing often and it is not practical or cost effective to maintain custom tags on every page. The solution is standard tags. This makes maintenance easier but reduces the quality of the data to not much better than log files. A tagging solution requires that you make changes to your pages or the servers and becomes increasingly costly to maintain as your data analysis needs grow.

Because page tags work from the visitor's browser, they can miss some important server events. For example, if a page has a server error, it never gets to the visitor and the page tags do not fire, so you get no data on this important event.

Then there are the security and privacy issues that have prevented many financial and government institutions from employing page tags. The tag data is

usually sent to a third party site where it is warehoused with all the other clients' data. Sending potentially sensitive information off-site is often unacceptable.

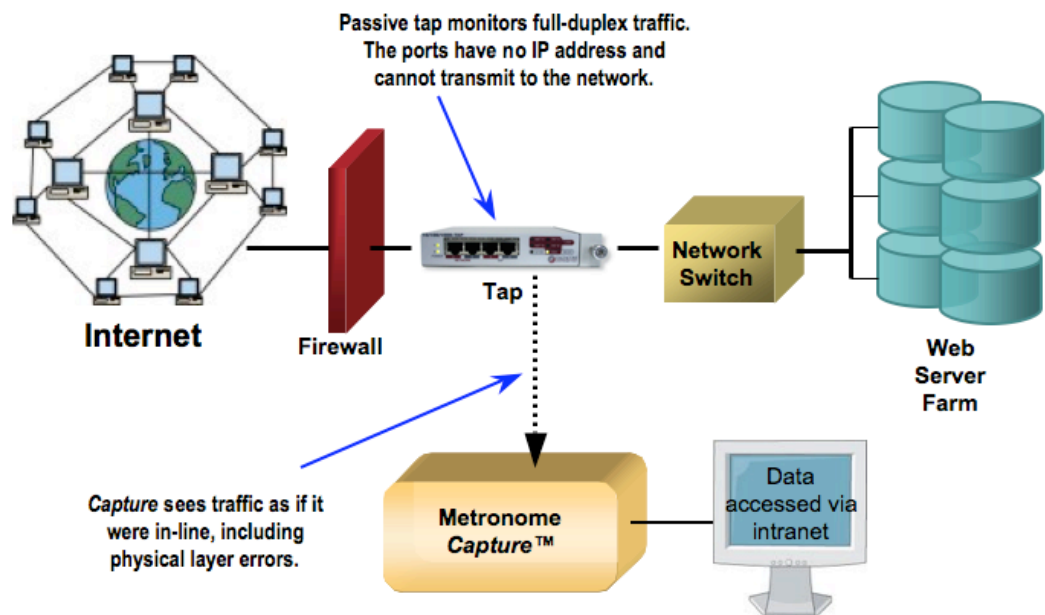
The next Internet boom will be serving content to mobile devices. Many mobile devices are not capable of executing JavaScript, so dynamic tags will not work. This means that sites will be forced to analyze log files or use cumbersome static tags.

Today, websites are seen as one customer touch point in an integrated marketing and sales strategy. To get a complete view of your visitor, web data must be joined to data in your corporate, sales and marketing databases. But data from web tagging is collected in a remote third party database and there is a vast amount of it. To make the join you must download this data over the Internet. Now you have to manage the web data that you are paying someone else to manage!

### **Metronome Capture – rich, secure and convenient data capture**

Metronome Capture collects the visitor and server activity, filters and sessionizes the data and keeps it all inside your firewall.

Metronome Capture is placed inside your firewall before your load balancer. It passively listens to all the traffic to and from your site regardless of which server actually handles the request. It collects all your clickstream data at one central location, even when you have multiple servers and domains. It produces a single log file (or data stream) for your whole site with the data already filtered and organized into visitor sessions. It is not a proxy and is totally passive. A failure in the TAP or the Capture application will not affect your website and it will never affect performance.



Capture sees all of the traffic flowing between your visitors and web servers so it sees the acknowledgements to requests plus the low level errors that the server

---

never sees. This enables *Capture* to calculate every detail of the transaction including precise load times for the HTML and each of its components.

Metronome *Capture* automatically links page views and hits into visitor sessions with a unique ID using a sophisticated algorithm. The data is available immediately.

### ***Filter and transform***

Metronome *Capture* has a sophisticated rule engine that is easily configured to give you the data you want and remove the data you don't. You can filter out hits you do not need based on any criteria. You can decide which fields you want and determine the format of your log. You can perform translations on the data and create custom fields to your specifications. For example, you could look for and extract a specific string from your cookie or a ZIP code from a form entry. You can pull out product IDs from a shopping cart or search parameters from a search request. You can categorize traffic by website, domain, etc. The rules are executed when the transaction occurs, so you get the results in exactly the format you want with no post processing required. *Capture* can even extract tags and data from the HTML pages, perform transformations and add them to the logs.

Cleaning data in real time consumes less storage space, less computer power and makes the data immediately available.

### ***Data channels***

Channels enable you to deliver different views of the data to different user communities such as IT, Marketing, etc. The combination of rules and channels enables you to feed analytics, load databases and perform traffic analysis any way you want.

Channels allow you to filter, clean and aggregate data in different ways, and allow you to deliver the results to different locations. You could create a "visit" log that contained one row per visit with information, another log for page views and a third to feed a fraud detection application with zip codes and IP location information. *Capture* typically sends the data to a log file, but it will also stream the data to an IP address for processing by another computer on your network in real time.

Information collected and managed by Metronome *Capture*, including all custom reports and errors, can be requested as XML via the HTTP protocol.

*Capture* can emulate a web server log or tag server log, even when there are custom tags. If want to keep your current analytics package but eliminate page tags or web logs, *Capture* can mimic the log format while creating just one pre-filtered log file. If you need great analytics, read about Metronome *Explain*<sup>™</sup>.

Web data is cleaned, filtered, transformed and organized into visitor sessions in one log file.

*Capture* data is ready to use without any post processing.

---

## More advantages

### **Data encryption**

*Capture* can track a visitor across your insecure and secure sites. You can load your master encryption key file onto the *Capture* platform so *Capture* can decode the HTML. Since *Capture* is secure behind your firewall, there is no security risk. If you need higher levels of FIPS compliance, *Capture* supports SSL cards from companies like nCipher®. *Capture* supports sites that use multiple RSA keys.

### **Triggering events**

Metronome *Capture* supports three types of events (report, error and session events). An event is triggered whenever the channel it is associated with allows transaction data to pass through its filtering rules. All of the channel's data cleansing rules apply to any data used by the event. Events may also have their own data filtering and cleansing rules.

### **Beacons and event sequences**

The Beacon feature enables you to detect and track sequences of important business events that occur within unique sessions. For example, you may want to track the particular sequence of placing an item in the shopping cart, viewing the shopping cart and then the item being removed. The sequences of events that have been detected so far for a particular session may be analyzed using the x-beacon data identifier. The complete sequence may be placed in a log variable and used to generate an event.

### **Geo-location**

Metronome *Capture* uses an integrated database from Quova® to instantly pinpoint each visitor's physical location (country, state, city) and identify their connection information (ISP, network carrier and connection speed). You will know where your visitors are coming from. You can monitor connection latency from different cities to troubleshoot response problems. You can extend this with events for real-time fraud detection applications such as correlating ZIP codes from applications with actual visitor location.

### **Metronome Web Console**

You can create real-time reports from any of the standard or custom fields that are being collected. Reports typically show aggregated data about what is happening on your website now. This might be the number of visitors currently online, the number of visitors per hour over the last few hours, longest page load times, etc. The reports show up-to-the second information and can be refreshed every few seconds. They are viewed through your web browser. The report data can be requested in XML format on demand.

By allowing Metronome *Capture* to decode secure data, trigger on events, and track sequences of events in your site you will develop a powerful understanding of key events like purchase or shopping cart abandons.

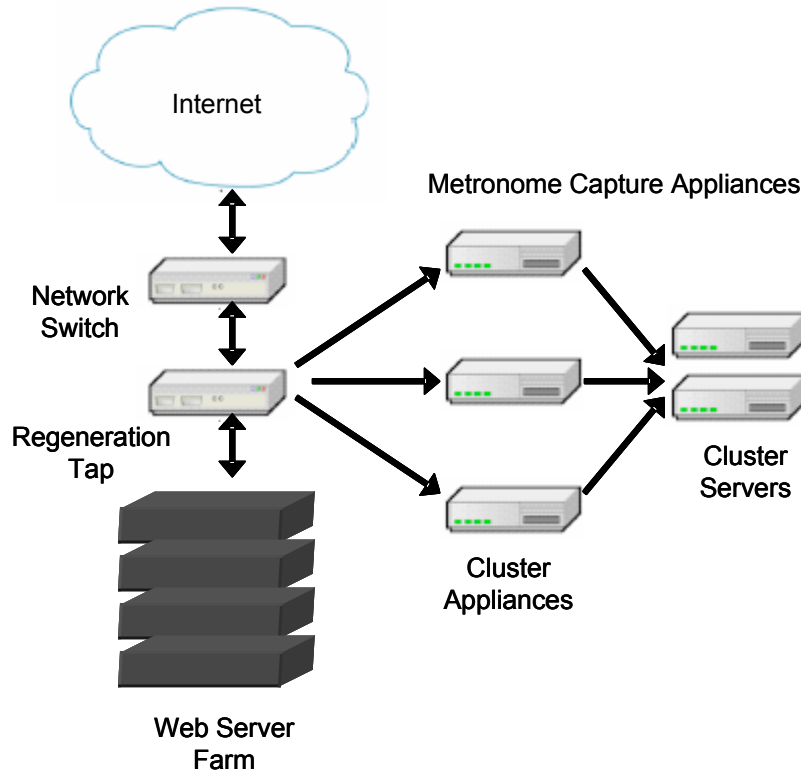
Instantly know where your visitors are coming from.

Track network latency, detect suspicious visitors.

---

### **Clustering and failover**

On busy sites, the data collection *Capture* functionality can be logically split into the functions that handle the packet capture, reassembly and filtering (appliance) and the functions that handle channeling and events (server).



### **Clustering**

*Capture's* dynamic clustering allows appliances to automatically share the load. When an appliance is added or removed the others dynamically adjust to share the load evenly. *Capture* senses when a member of the cluster fails and the members adapt automatically. Clusters are currently installed on some of the busiest e-retail sites.

### **Extensible**

Metronome *Capture* has an extensible Java layer that listens to an IP socket and receives data from a *Capture* channel. This layer can be loaded on a *Capture* computer or a different computer. By extending the Java classes, you can distribute channel data any way you want. Currently, there are standard plug-ins to load the data into a database and send alerts over email.

### **Technical**

*Capture* uses a passive network tap to promiscuously collect packets. The *Capture* ports cannot transmit data and have no IP address so they cannot be interrogated.

---

A network tap is typically inserted between the load-balancing switch and an edge router. This tap maintains a hard-wired connection between the two devices so that the flow of traffic is not delayed and a failure of the tap (e.g. due to a power loss) will not cause a network outage. Since the tap is one way into the *Capture* computer, it does not introduce a security risk. *Capture* also supports the use of SPAN ports and repeating hubs. SPAN ports are often available on routes and load balancers.

*Capture* is a Linux application that runs on in a dual processor server configuration and can utilize multi-core CPUs. Its multi-threaded architecture distributes work evenly across multiple processors, allowing a single application to scale to the full line speed of both copper and gigabit fiber networks.

### **A revolution in web analytics**

Web logs were never intended to be used in analytics, so it is very difficult and expensive to extract information from them. Building any but the most basic report from web logs takes hours or even days. Web logs also slow down your servers and offer virtually no insight into a visitor's actual experience.

Embedding page tags into your web pages allows you to eliminate web logs and receive reports faster but become increasingly costly to maintain as your needs grow. Page tags raise privacy and security issues, especially when using a hosted service. Tags only work on pages that actually get loaded, not on the ones that break.

Metronome *Capture* eliminates web logs and page tags by analyzing and extracting information from your network traffic inside your firewall. There are no security or privacy concerns and little maintenance. Plug it in and you are up and running. Metronome *Explain* extends the solution to provide powerful and sophisticated reporting that gives you a complete view of your visitors.

### **About Metronome Labs**

Based in Pittsburgh PA, Metronome Labs LLC develops value-added products including solutions for web analytics, IT forensics and web data capture and loading. There are hundreds of *Capture* appliances installed, including major e-retail sites. For more information, visit the website at [www.metronomelabs.com](http://www.metronomelabs.com).

Plug in Metronome *Capture* to understand what is happening on your website and what your visitors are experiencing.