



## **Preventing Data Leaks At The Firewall**

A Simple, Cost-Effective Way To Stop Social Security  
and Credit Card Numbers From Leaving Your Network

December 2008

Palo Alto Networks  
232 E. Java Dr.  
Sunnyvale, CA 94089  
408.738.7700  
[www.paloaltonetworks.com](http://www.paloaltonetworks.com)

## Table of Contents

Executive Summary .....	3
Data Leakage Continues to Be a Problem For Enterprises .....	4
DLP Technology is Cumbersome, Incomplete, and For Many, Overkill...And Don't Forget Expensive.....	4
For 90% of Enterprises, Keeping Social Security and Credit Card Numbers From Leaking Would Be Enough .....	4
The Perimeter Is Key, But Legacy Security Technology Can't Help .....	4
First - Block "Bad" Applications .....	5
Second - Scan "Good" Applications. Including SSL. ....	5
Lastly - Know Users, not just IP addresses.....	5
A Word on Performance .....	5
Palo Alto Networks™ Is First to Include DLP Functionality In The Firewall .....	6
It's Time To Fix The Firewall.....	9

## Executive Summary

Numerous examples of accidental and deliberate data leakage continue to make headlines, and data leak prevention (DLP) technologies are being touted as a panacea. Unfortunately, given the scope, size, and distributed nature of most enterprise datasets, just discovering where the data is and who owns it is a challenge, and many DLP projects are proving slow to progress. Furthermore, given the absence of control over applications in most enterprises, it isn't clear that DLP technologies would have helped in several of the recent cases (US Army, Pfizer, etc.). Some organizations, for business model or cultural reasons, will have to go through the effort of large-scale DLP implementations. For most organizations, however, controlling the applications most often used to leak sensitive data and stopping unauthorized transmission of credit card and social security numbers and their ilk would be acceptable. Exerting that control at trust boundaries is ideal – whether the demarcation point is between inside and outside or internal users and internal resources in the datacenter – the firewall sits on the perfect spot, seeing all traffic. Unfortunately, legacy port- and protocol-based firewalls can't do anything about any of this – being ignorant of applications, users and content. To do this correctly, enterprises should first achieve a degree of control over applications – thus limiting the avenues of data leakage. Second, organizations need to scan the applications they do want on their networks, for confidential data. Third, organizations should be able to understand which users are initiating these application transactions.

Palo Alto Networks™ next-generation firewalls incorporate three key technologies that enable enterprise customers to incorporate some of the most commonly needed DLP functionality at their network perimeter – easily, and without adding more appliances. App-ID™, User-ID, and Content-ID, coupled with the high-performance, multi-gigabit next generation firewall platform offers immediate relief to the most common data leakage pain (SSN/CC moving over unauthorized applications), allowing enterprises to complete their large scale DLP projects at their leisure.

## Data Leakage Continues to Be a Problem For Enterprises

Large scale, public exposures of personal information remain far too common. Practically weekly, headlines declare tens of thousands of credit card numbers leaking out of retailers, or social security numbers leaking out of government agencies, health care organizations, or employers. A recent example (December, 2008) showed a misconfigured and prohibited peer-to-peer file sharing application putting a database of 24,000 US Army soldiers' personal information in the public domain. This is similar to the Walter Reed Medical Center breach in June 2008, or the Pfizer incident from 2007 – all three involved data leaking through the perimeter via an application expressly prohibited by policy.

## DLP Technology is Cumbersome, Incomplete, and For Many, Overkill...And Don't Forget Expensive

Data leak prevention (DLP) technology has captured the attention of many IT organizations, with a promise to help organizations manage their confidential data. Project scope, for these technology providers, however, is a problem. Questions of access control, reporting, data classification, data at-rest vs. data in-transit, data ownership, desktop agents, server agents, and encryption have slowed DLP projects to a crawl in many organizations. Venture capitalists funded many data leakage prevention vendors, many of which have been acquired by larger security companies, who have further expanded the scope of an already unwieldy offering. Some of these vendors are now marketing data *loss* prevention, which incorporates practically the entire information security function (and even includes elements of storage management!). Needless to say, this broadened scope is beneficial, but adds complexity, time, and expense – both in hard costs and in staff time. Oddly enough, many of the recent breaches caused by unauthorized and misconfigured peer-to-peer file sharing applications wouldn't have been prevented by the typical implementation of DLP technologies on the market today – because control of applications isn't addressed.

## For 90% of Enterprises, Keeping Social Security and Credit Card Numbers From Leaking Would Be Enough

For a few highly intellectual property-dependent organizations, implementing a long-term, comprehensive DLP project – which should ultimately include data discovery, classification, and cataloging – is appropriate. But for the remaining 90% of enterprises out there, stopping a couple classes of confidential data (e.g., social security numbers and credit card numbers) at the trust boundary would be a great start. This would avoid the expensive and embarrassing public exposure of employee and/or customer personal data.

## The Perimeter Is Key, But Legacy Security Technology Can't Help

If enterprises could control the flow of confidential data at the trust boundary, it would stop a large percentage of incidents that regularly make the news. Unfortunately, the legacy security infrastructure at most enterprise perimeters is poorly equipped to offer this functionality. Most firewalls sit in a great position to help – they demarcate the trust boundary, they see all traffic, and they exert policy control (i.e., they can block traffic). But legacy firewalls don't understand content, don't understand applications, can't see inside SSL-encrypted traffic, and have no understanding of users. In fact, if it isn't source or destination IP address, source or destination port, or network protocol, firewalls don't understand it. Other firewall "helpers" (e.g., intrusion prevention systems, web proxies, URL filtering devices) only see a portion of the traffic, don't sit in-line, and/or have limited application and content understanding.

## **First - Block “Bad” Applications**

Examining most of the recent incidents, the first thing enterprises need to do is get control over which applications are running on the network. Every organization has a different view of desirable and undesirable applications. Each enterprise needs to look at applications from both benefit and risk perspectives. On the benefit side, an application might help an employee do their job better, faster, or cheaper, or improve customer relations, or make the workplace more pleasant. On the risk side, applications may harbor vulnerabilities, carry malware, be prone to misuse, or – particularly relevant to this discussion – transfer files. In some cases, organizations want to enable social networking applications for cultural reasons, or for business reasons – or block them for security reasons. Either way, the first thing to do with regard to stopping confidential data leakage is to identify which applications are moving across the network – regardless of whatever evasive tactic the application employs, and block undesirable applications (thus limiting the avenues through which confidential data can flow). In order to do perform this control effectively, the device needs to “see” all traffic.

## **Second - Scan “Good” Applications. Including SSL.**

The second thing to do is scan desirable applications for confidential data leakage. Once an organization has settled on the applications it wants on its network, the next step is to scan applications for confidential data leakage – including SSL-encrypted application traffic and compressed content. As an aside, any applications that use proprietary encryption (e.g., Skype) should be very closely evaluated, because if allowed, they cannot be scanned. Generally, this content scanning should be high performance (high throughput, low latency), unobtrusive, and easily implemented. More specifically, the scanning capability should be simple to enact in policy, and adjustable in sensitivity, to allow normal appropriate transactions without triggering response – yet still detect abnormalities.

## **Lastly - Know Users, not just IP addresses**

The third thing to do is to bring users into the picture. Understanding which users are using applications, and which are engaged in moving particular classes of content has two benefits – actionable visibility, and refined policy. The most efficient way to do this is to tie into the enterprise directory (identities and groups are already there). Often, when an organization hears that they’ve had a leak, the first thing they ask is, “who leaked it?” Having the ability to understand users – not just IP addresses – gives the granularity enterprises need to guide specific users about policy, and take more serious action if warranted. The second benefit understanding users offers is the ability to incorporate that understanding into policy – so certain users might be enabled to use certain classes of applications, and other users might not. This empowers enterprises to further compartmentalize and contain risk.

## **A Word on Performance**

Today’s applications are real time in nature. With the exception of email, for which delays go largely unnoticed, users expect instant response from their applications. Which means that the network that those applications run on cannot delay applications and content. The reason for highlighting this is simple – any solution that meets the above three requirements must be high performance in nature – high throughput, low latency – both in scanning, as mentioned above, but also in traffic processing.

In summary – if IT staffs know the application, the user, and the content (i.e., whether or not the traffic contains confidential information), they can act – block or alert – quickly, and appropriately, without sifting through dozens of log files.

## Palo Alto Networks™ Is First to Include DLP Functionality In The Firewall

Palo Alto Networks offers enterprises a unique approach – visibility and control over applications, the ability to scan application content, and visibility and control of users and groups. Palo Alto Networks next generation firewalls incorporate 3 key identification technologies into a high-performance platform. App-ID, Content-ID, and User-ID give organizations business-relevant control over applications.

**App-ID™ Classifies Applications.** App-ID technology identifies applications regardless of port, protocol, encryption, or evasive tactic. App-ID currently recognizes over 750 applications, and Palo Alto Networks adds 3-5 new applications per week – giving enterprises visibility and policy control over actual applications, not just ports.

**Content-ID Identifies Content – Including Confidential Content.** Content-ID technology incorporates 3 key content security elements – confidential data (DLP functionality), threat prevention, and a URL filtering capability. Content-ID is stream-based scanning engine using a unified signature format. Which means, first, that it doesn't buffer files, and second, that it only scans once, for all content security functions. Palo Alto Networks' single pass architecture is discussed further below. The data filtering feature in PAN-OS makes implementing DLP functionality in the firewall simple. As shown in Figure 1, adding a policy object that scans application traffic is a matter of assigning the data filtering profile to the policy, determining what sort of data to scan for, and assigning a weight (i.e., number of occurrences that trigger the policy action). For example, for one application, Yahoo Instant Messenger, an organization could say that a single credit card number would trigger a policy action ('block'), but another class of application, say, Social Networking, a user would need to transmit 3 credit card numbers to trigger the policy.

Note that the SSN and credit card signatures use validation algorithms to minimize false positives (e.g., Luhn algorithm for credit cards). Enterprises can also use the regular expression capability built into the data filtering feature to create custom patterns.

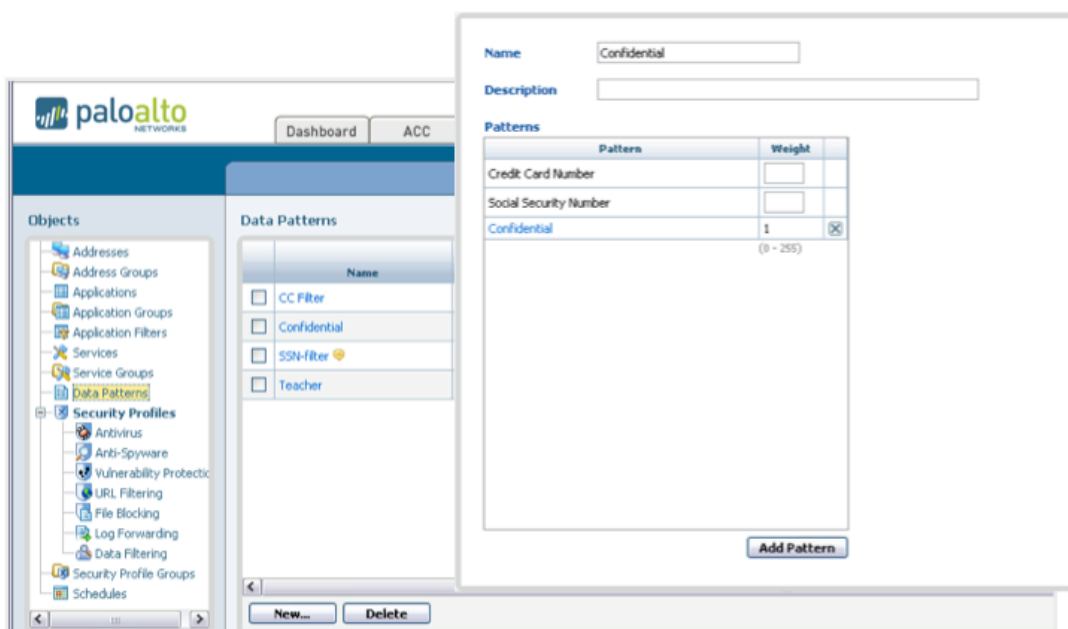
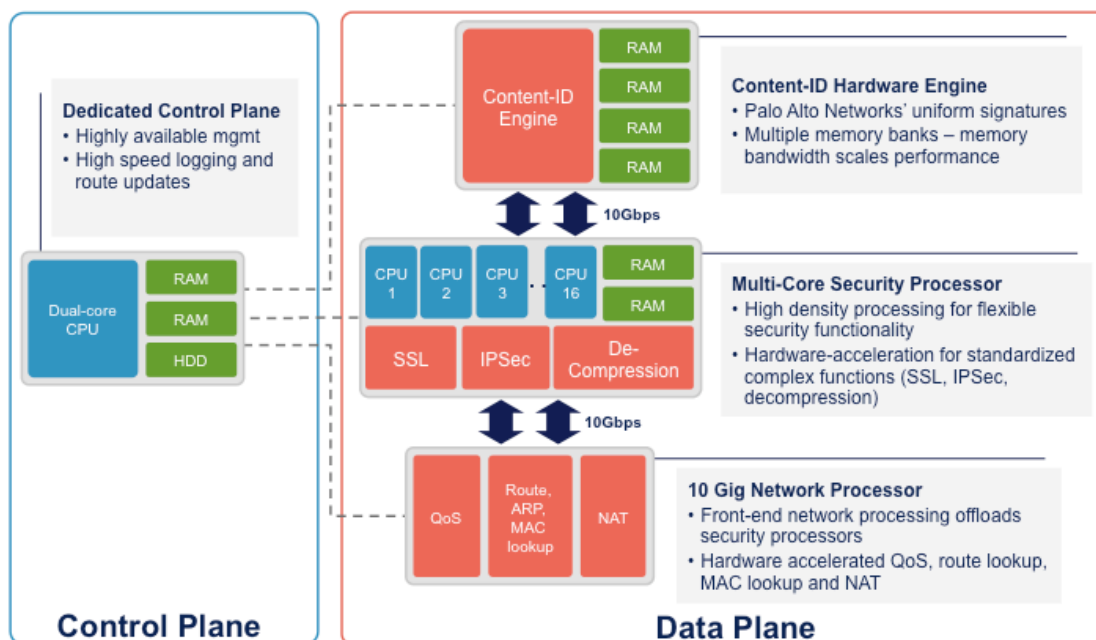


Figure 1 - Adding a Data Filtering Policy Object

**User-ID Integrates With Enterprise Directories.** User-ID technology integrates Palo Alto Networks' next-generation firewalls with enterprises' Active Directory implementations. Meaning that the single policy engine governing application and content security also has the ability to refine that policy with the user and group definitions already used in the enterprise.

**High Performance Platform.** Palo Alto Networks next-generation firewalls offer a level of performance not previously available in any device incorporating DLP functionality – up to 5 Gbps of throughput while scanning application traffic for confidential data. This is due to two key elements: Palo Alto Networks' hardware architecture and the single pass architecture used for the data path.

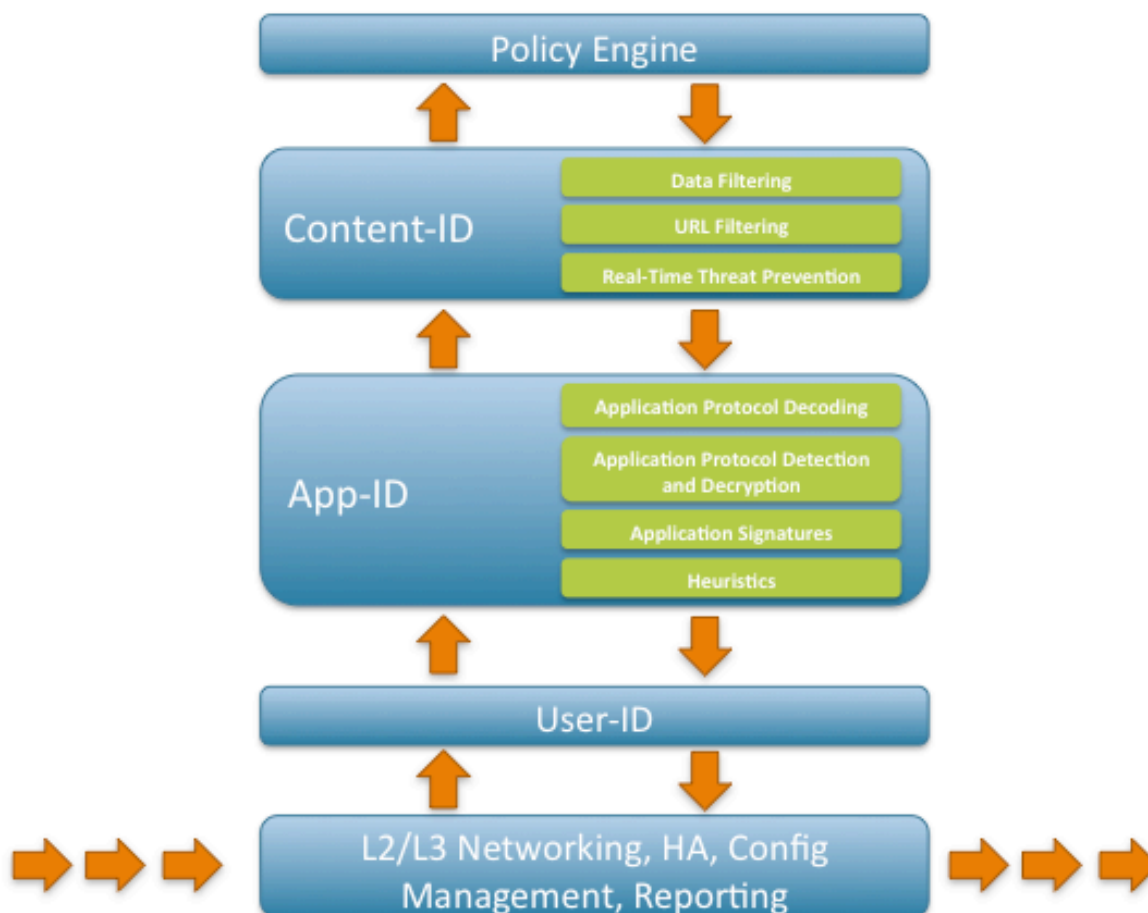
Examining hardware first, Palo Alto Networks uses principles commonly employed when building networking devices. Separation of data and control planes means that heavy utilization of one doesn't negatively impact the other. The control plane has its own CPU, RAM, and disk. Additionally, dedicated, specialized processing and memory for networking, security, and content analysis – all connected via a high-speed data plane (10Gb on the PA-4000 Series, 1Gb on the PA-2000 Series) means that traffic won't bog down. Figure 2 depicts the PA-4000 Series hardware architecture.



**Figure 2 - PA-4000 Series Hardware Architecture**

Second, Palo Alto Networks engineers addressed the path that traffic takes through the security infrastructure. In legacy network security infrastructure, traffic flows through several security devices, each with its own networking engine, classification engine, pattern matching engine, and policy engine. This duplication of effort is not only inefficient, but also slow. This poor performance is the key reason why enterprises are loath to put yet another device in the traffic flow – especially something as intensive as DLP scanning.

Palo Alto Networks next-generation firewalls utilize a single pass architecture, with traffic flowing through a single networking component, a single application classification engine, a user classification capability, and a single content/pattern matching engine – resulting in the ability to see and enforce policy control across applications, users, and content (including confidential data and threats) – without slowing traffic. Figure 3 is a graphical representation of Palo Alto Networks’ single pass architecture.



**Figure 3 - Palo Alto Networks' Single-Pass Architecture**

## It's Time To Fix The Firewall

Comprehensive DLP is a worthwhile pursuit, but is complex, expensive, and takes a while to implement. In the meantime, the legacy firewall is sitting in a prime spot to help out – but due to its inability to see applications, users, and content – can't do a thing. Firewalls should be able to:

- First, block undesirable applications.
- Then, scan allowed applications for confidential information.
- See and manage policy by users and groups, not IP address

Fortunately, Palo Alto Networks, with its identification technologies, delivers this in a high-performance firewall platform. The application visibility and control, coupled with the data filtering feature found in Palo Alto Networks' next generation firewalls can enable simple, high-performance DLP controls at the enterprise perimeter – which would have stopped the data leaks in several recent, highly publicized incidents. And that should free up staff to work on that data loss prevention project.