Securing the new era of big data

In the digital age, information is the new currency. And in order to get information, enterprises are mining data – and lots of it – for the knowledge that it can yield. On the scale of Internet commerce or social networks, the amount of data can be pretty large - think of the hundreds of millions of smartphones and end-user devices. On the scale of consumer, medical, scientific, or research data, it can be gigantic, as sensors and instruments can collect vast amounts of raw data, whether from a single source (such as instrumentation of a GE aircraft engine during a flight) or from the projected 26 billion devices that will make up the Internet of Things.

The Gold Rush we currently see for collecting and analyzing Big Data, which in turn is being fed increasingly by the Internet of Things, is creating greater challenges for the networks and security of the data centres in three key areas:

First, there is Aggregation. Increasingly, rather than processing and reducing the raw data at the data source to a more manageable volume, raw data is being transferred and stored centrally – because it now can be – so that it can be analyzed in different ways over time. Today, enterprises are transferring terabytes of data over long distances every day. The sheer quantity of data is forcing core network and data centre upgrades, such as 100GbE switching fabric, to deal with individual data transfers at10Gbps or even higher. This also creates challenges for perimeter security, such as firewalls, as many vendor solutions are not designed to handle such large in flows and sessions. For example, a firewall that boasts 10 GbE ports or 40Gbps aggregate throughput may not actually have internal processing paths all the way through to handle an individual 10Gbps flow. LAN congestion from normal enterprise campus traffic may further saturate network appliance CPU or memory resources, causing large flows to stall or even drop.

Next comes Processing. Big data flows are not symmetric – the raw data that goes in does not necessarily go out in the same form and

volume. Instead, the data kept in storage arrays is analyzed typically by an intermediary set of servers, then further reduced and delivered – often by web server front-ends – as a reduced set of insights before exiting the data centre. This means higher bandwidth with an increasing proportion of lateral or east-west traffic, within the data centre, instead of north-south traffic that is going out to the Internet or elsewhere. Many studies show that east-west traffic now accounts for up to 70% of the data centre traffic, and this trend will continue to increase with the growing amount of big data analytics.

East-west traffic needs to be segmented and inspected, not just for blocking lateral movement of advanced persistent threats and insider attacks, but to secure the data itself, some of which can be sensitive if disclosed or leaked. Network security architectures need to evolve from perimeter or gateway security oriented to a multi-tiered, hybrid architecture where more east-west traffic becomes virtualized and abstracted with the adoption of server/network virtualization and cloud computing.

Last, there is Access. As part of Big Data, data is being archived for long periods, which is authorized to access which data, and for what purposes? Often there is not just a single data set, but rather multiple repositories of data that may be combined and analyzed together. Each set of data may contain certain sensitive or confidential information, and may be subject to specific regulations or internal controls. Further, there is often not just one group of analysts or researchers, but over time many constituents seeking to gain different insights. A large pharmaceutical company provided a good example where their Big Data research efforts were open to not just internal employees, but also contractors, interns, and visiting scholars. For each of them, a separate analytics sandbox needed to be created, authorizing and auditing specific entitlements to identified data sets that could be accessed and combined.

In such context, IT organizations need to fundamentally re-think network security instead of taking incremental steps to meet

evolving data centre security needs. In many cases, the data centre infrastructure itself is presently being consolidated and transformed due to not just Big Data, but ongoing cloud computing and SaaS initiatives as well. As part of this transformation, IT should consider an architecture that is:

[if !supportLists]· [endif]High-performance– support the larger volumes of data with higher network throughput and with high-speed ports (e.g. 40G/100G fabric) and high port density, but also be scalable and elastic to accommodate ever-growing data sets

[if !supportLists]· [endif]Secure – augment perimeter security with increased internal segmentation to secure east-west movement of data and monitor for advanced and insider threats

[if !supportLists]· [endif]Consolidated – integrate multiple security functions from core security functions like firewalling/VPN, anti-malware and intrusion prevention to advanced threat protection, strong authentication and access control

Finally, customer may consider the opportunities to leverage Big Data itself for better security. With more control and monitoring points being deployed throughout the network and the data centre, plus with SIEM (security information and event management) and log management tools being able to aggregate increasing amounts of security logs and event data, more security analytics and insights are possible to better protect not only Big Data but also the data centre as a whole.