Analytics

Craig Wentworth

Premium report

Technology & Suppliers

On the Radar: ExtraHop

October 2014

ExtraHop provides a platform for capturing, decoding and analysing network wire data in real time for the purposes of business and IT Operations Analytics. It can also stream this data to a NoSQL database for multi-dimensional post-hoc analysis when combined with data from other sources.

This *On the Radar* briefing note follows a simple "ten questions" format, which we designed to provide a concise but thorough overview of a company and its products and services. We use this format to focus on the capability and suitability of small, specialist vendors – to help you build the best possible vendor shortlists when looking to make new technology investments.

Find out how to access related research at http://www.mwdadvisors.com/ec/membership.php.

MWD Advisors is a specialist advisory firm which provides practical, independent industry insights to business leaders and technology professionals working to drive change with the help of digital technology. Our approach combines flexible, pragmatic mentoring and advisory services, built on a deep industry best practice and technology research foundation.

1. Who?

Based in Seattle, WA, USA, ExtraHop (www.extrahop.com) provides a platform for capturing, decoding, analysing, and streaming network wire data 'in flight' for the purposes of business and IT Operations Analytics. In the context of MWD Advisors' Big Data vendor landscape[1] technical capability clusters, ExtraHop occupies a position as a provider of 'specialist source' Big Data analytics, especially as it's now able to stream its data to a NoSQL store for consolidation and analysis in combination with other sources.

2. What does it do?

The ExtraHop Operational Intelligence[2] Platform for wire data analytics is a real-time stream processor of network traffic. It takes a copy of the hex code 'in flight' (by tapping into a mirrored system without the need for agents, so there's no load applied to production traffic itself), reassembles network packets (accounting for IP fragments, out-of-order segments, and microbursts), combines these to form a data stream that reconstructs network traffic flow, and applies decoders for individual protocol layers in order to extract meaningful transactional information.

The platform decodes the layers from L2 through to L7 (i.e. Data Link, Network, Transport, Session, Presentation, and Application) as standard. Additionally, the company has recently begun releasing decoders for specialist layers, launching with an HL7 Analysis Module in ExtraHop's Healthcare Edition which decodes the standard messaging protocol (a subset of the L7 Application layer) for transferring clinical and administrative data between Hospital Information Systems. It's these proprietary decoders that enable the ExtraHop platform to observe real-time behaviour between systems at a deeper level than traditional packet inspection and header sampling alone would have allowed. It's able to pull together request and response information (including selective payload) from transactions between applications as they communicate across each operational tier of an organisation's physical or virtual infrastructure.

ExtraHop's base framework covers those protocols common in web environments, e.g. HTTP/HTTPS, SSL, and AMF. For HTTP payloads, for example, data could include order numbers, customer IDs, and/or structured data in XML, REST, JSON, AJAX, JavaScript, and HTML5 formats. Additional modules are available that support specific applications requirements. These include databases (Sybase IQ, Couchbase, IBM DB2, IBM Informix, MS SQL Server, MySQL, Oracle, and PostgreSQL); electronic securities and equities trading applications; MS SharePoint CMS; IBM MQ message-oriented middleware; Directory and Authentication, Authorisation and Accounting services (such as LDAP, RADIUS and Diameter); infrastructure services (such as DNS, FTP and SMTP); EMC backup and repositories; and now some industry-specific protocols, including healthcare (HL7), financial services (Financial Information eXchange), and telecoms (Short Message peer-to-Peer).

The platform is based on a proprietary data store and scales up to 20Gbps in a single appliance. It can perform real-time SSL decryption at that throughput rate without having to resort to offline post-hoc packet analysis by processing up to 35,000 handshakes per second for 2048-bit RSA keys and 200,000 handshakes per second for 1024-bit RSA keys.

[1] See the MWD report *Big Data for analytics: Vendor landscape* at http://www.mwdadvisors.com/library/detail.php?id=594

[2] ExtraHop describes 'operational intelligence' as the result of bringing a Big Data ethos (of breaking data out of siloes to affect multi-dimensional analysis) to the field of IT Operations Analytics, where the focus tends to be more on application performance, availability and security management. However, as this report illustrates, ExtraHop's platform has traditional business analytics applicability too – if deployed to facilitate accessing 'special sources' of data in a Big Data analytics environment.

ExtraHop provides an 'Application Inspection Trigger framework' that allows the platform itself to apply base-lining and trending to combinations of metrics, e.g. visualisations of data showing the frequency of sales overlaid onto a real-time geographical map showing the location of the users purchasing those products over a specified period. However, it's how ExtraHop is able to share the metrics it gathers from its wire data in real time with other analytics platforms (calling in turn upon other data sources) via its Open Data Streams technology that extends the solution from IT Operations Analytics (providing a side-show of fairly well-bounded native business analytics based on what the product itself can glean from the wire over a period of time) to one which can inject real-time, business-relevant data into a much wider analytics conversation – one with much wider applicability beyond IT operations.

The recently announced Open Data Stream for the NoSQL database MongoDB is designed to stream ExtraHop's wire data (in real time) into a highly scalable third-party store that brings together unstructured data sets from diverse sources across and outwith the enterprise for post-hoc transactional processing and non-invasive multi-dimensional analysis. There is a similar capability for Elasticsearch ELK, joining existing streaming technology for Splunk VMware LogInsight. ExtraHop cites numerous sources relevant to IT operations management, but Big Data analytics use cases certainly benefit from the input of a subset, such as agent data from instrumented systems, data, meteorological data, social media and a wealth of other human-generated data. A NoSQL database is a common component of many Big Data environments, and MongoDB has good integrations with analytics and visualisations platforms like Tableau, Pentaho, Chartio, Elasticsearch Kibana, Domo, and JSON Studio to generate and present that business insight.

Wire data streaming to MongoDB (and others) is limited only by the NoSQL database's sharding (i.e. clustering) capability; ExtraHop itself can stream an unlimited number of time-stamped transactions (up to 400,000 per second from a single appliance).

3. Who is it for?

Historically ExtraHop has marketed its platform as a way to monitor the 'application delivery chain', and indeed has been focusing on use cases within the field of IT Operations Analytics. To that end, platform bundles have been geared towards application management, with visualisations, trending and alerts capability, programmable metrics and analytics for enterprise applications such as MS SharePoint, MS ActiveSync and IBM WebSphere offered on top of its core Context and Correlation Engine.

However more recently, and more interestingly in the context of Big Data for business analytics, the company has also been starting to position the technology as a means by which real-time customer and other data-of-interest present 'on the wire' in the transactions between business applications can be woven together with data from other sources as part of a wider business analytics use case. Opening up a real-time streaming connection to a NoSQL database like MongoDB brings the potential for ExtraHop's customers to apply its wire data extraction and processing technology to provide a valuable 'special source' of Big Data for enterprise-wide business analytics.

So, whilst ExtraHop is still being predominantly driven by the needs of IT operations to develop a view of application performance across organisational infrastructure, the by-product is a solution – when deployed with a different use case perspective in mind – that also provides access to real-time transactional data for business analytics purposes.

There's also a scenario where ExtraHop's approach of tapping into the wire data can shortcut the extraction of historical data (as long as it's queried at least once in a contemporary setting). Since the data packets will travel on the network that ExtraHop's platform has a window on, with the appropriate decoding it can collect data which may have resided in a legacy store as if it were any other data in flight, without the need to install a connector and translate from the source system itself; it picks up the hex code after native extraction, as it passes through the network.

Of course this only works for historical data which becomes the subject of a live request – it has to be summoned from its store first before it'll be seen 'on the wire' – but it only has to be queried once before it's captured by the ExtraHop platform and can subsequently be streamed to a NoSQL database for further analysis.

ExtraHop's platform is available on a physical appliance (available as all-in-one rack units supporting either 3Gbps, 10Gbps or 20Gbps data throughput), or as an AWS-ready virtual appliances (1Gbps or 3Gbps varieties) and a Central Manager virtual appliance for managing and reporting distributed ExtraHop appliances (both physical and virtual).

ExtraHop doesn't publish comprehensive list prices on its website, but states that its enterprise edition starts at $7,500 – equating to $4-$12 per server per month, depending on what protocol modules are included. It's available either as a Pay-As-You-Go (i.e. pay-only-for-what-you-need-*today*) monthly subscription plan (covering all maintenance, support and software upgrades for physical or virtual appliances; minimum term 12 months); or a traditional perpetual plan (one-time hardware fee plus annual support and maintenance fees, for physical appliances only). It also supports a Pay-As-You-*Grow* option (subscription or perpetual), which permits flexible bursting without penalties and allows customers to add additional servers as needed.

There's also a freely downloadable 'Discovery Edition' for VMware, Hyper-V and AWS environments, which offers basic functionality with some restrictions. For example, it has limited throughput capability – only the L2-L3 layers are decoded, compared with full L2-L7 in the enterprise edition – and it sports only a 24-hour history, compared with 30-days in the full version.

4. Why is it interesting?

ExtraHop's take on the field of Big Data is that there are three main areas where innovation can flourish: data extraction, the data store (how scalable, searchable, real-time friendly, etc. it is), and the advanced analytics/visualisations that deliver up actionable business insight. It also maintains that up until now, it's mainly been these last two stages that have benefited from most of the attention. Its argument is that however good the store or the analytics applied to the data within it is, what you're able to do is still dependent upon the how good the extraction method is at getting you the data you need (when you need it). Whether you're processing data-at-rest in a giant data lake, or streaming data-in-motion, for the most part the extraction principles are broadly similar: data is generated by some process (either by machine, or requiring some form of human interaction), then it's collected, logged, possibly aggregated and then mined, and is either directed to an analytics processing engine or the compute comes to it. The analytics engine relies on a preceding Extract-Transform-Load process to supply the data it needs. Or, in the case of Hadoop and NoSQL stores, it relies on more of an Extract-Load-Transform one – since the data would typically be unstructured or semi-structured. When use cases describe real-time streaming scenarios, they still generally mean that the data gets to the compute via this route – it just might be happening too fast, or the insight required too soon, for it to be viable to store it anywhere first and perform post-hoc analysis at your leisure – so the streaming database still performs some sort of ETL, albeit very quickly and on a continuous stream of data.

Enter ExtraHop and its wire data analytics. The approach is interesting firstly because, rather than building connectors to acquire its data sources, the ExtraHop platform taps into the packets of uninterrupted data in flight around an organisation's dynamic network infrastructure. The company's mantra is "everything hits the wire[i]", and so it's from the wire that the platform reconstructs data streams for eventual analysis from data-in-motion, without recourse to any schema, meaning that any payload of any transaction between any applications becomes accessible by the same means. Any query, any message, that calls up data-of-interest from anywhere in the enterprise (including read-requests of historical information which might not normally be easily interfaced to a Big Data architecture) can potentially become part of an emerging analytics picture, providing ExtraHop's wire tap is downstream of the source.

The extent to which detailed information can be gleaned depends on the availability of decoders that extract rich, protocol-specific data; but the company is expanding its reach there, and basic packet analysis can at least reveal some specifics, where formats are known.

The second interesting aspect is that, whilst ExtraHop's technology was developed with IT Operations Analytics in mind, it's no great leap to picture how its wire data extract approach (whether with or without the benefit of native analytics processing) can be useful for getting at fast-moving streams of application transaction data that carries payloads-of-interest from a broader business analytics perspective – especially when it's combined with other data from other sources (for example, retail, customer behaviour, meteorological, regional income, demographic) to complete a more rounded view of the customer experience.

5. How established is it?

ExtraHop was founded in 2007 by engineering veterans from F5 Networks. Today the fourth generation of the company's platform product monitors over one million systems, undertaking trillions of transactions daily for customers across healthcare, e-commerce, financial services, and IT. Major customers include Alaska Airlines, Bet365, Concur, Purdue Pharma, and McKesson, though most are primarily using ExtraHop for application management and IT Operations Analytics, rather than in Big Data business analytics contexts.

ExtraHop is a privately held company funded by venture capital, having raised $61.6m in three funding rounds. The latest round in May 2014 raised $41m and was led by Technology Crossover Ventures, alongside existing investors Madrona Venture Group, Meritech Capital Partners, and Sujal Patel; other investors include Andreessen Horowitz. ExtraHop has declared twin priorities for this new funding: aggressive expansion into new geographical markets, and innovation within its wire data analytics platform to further "establish the IT Operations Analytics category". Regarding the latter, note that the emphasis is not primarily on business analytics and intelligence applications at this stage – though recent innovations, such as the HL7 module for healthcare use cases do open up wider Big Data applicability.

The company doesn't announce its financials, but reports that between 2012 and 2013 it achieved 150% year-on-year revenue growth.

6. How open is it?

ExtraHop's SDK documents the same APIs its web interface uses, so third-party developers have access to any metric in its internal data store. Application Inspection Triggers expose a programmatic interface for the ExtraHop parsing engine; when used to initiate syslog exports metrics can be sent to third-party analytics platforms or management tools according to policy-based and/or event-driven scripting.

AI Triggers expose a programmatic interface for the ExtraHop parsing engine, so the platform can be extended to monitor new types of metrics specific to customer's business, applications and IT environment.

7. Who does it partner with?

ExtraHop's technology partners focus on IT management and include Splunk, VMware, MongoDB, and Elasticsearch (all of which are destinations for wire data via its Open Data Stream channels), Amazon Web Services (the hosting partner for its virtual appliances), Couchbase, Citrix, F5, Cisco, Microsoft, Ixia, Keynote, SevOne, jSonar, Gigamon and Arista. It doesn't currently have any explicit partnerships to develop more industry-specific protocol layer decoders.

The company also has channel partners (local systems integrators and value-added resellers) that provide expert training, support, and other services.

8. Are there areas for improvement?

ExtraHop's technical basis is a good one for those looking for a way to access data in real time from within their own infrastructure that could bring additional richness to Big Data analytics. The base framework provides a window on many of the protocols underpinning transactions which serve up the sort of customer experience common in a digital enterprise, and growing specialisations in some verticals (particularly healthcare, through the HL7 layer decoder) have the potential to provide additional detail.

Given the trajectory ExtraHop is on with HL7 and the Healthcare Edition, it's not difficult to see how it could replicate this approach in other verticals – especially where it already has a significant customer presence (and can therefore work with them and learn their priorities) and has made inroads with support for pervasive protocol layers already (such as Diameter for telecoms and FIX for financial services).

However, what's currently still lacking – despite some product description nods in the direction of its applicability in providing 'business insight' – is any serious push by ExtraHop to highlight the platform's capabilities as being a good fit for this purpose. Until the company gets more solidly behind the product's strengths beyond IT Operations Analytics (i.e. bringing those capabilities in innovative data extraction and stream processing to wider business analytics) the latter market will remain largely untapped. That point is tantalisingly close, with industry-specific protocol layer decoders which can surface data items of business interest buried in real-time data streams; and the Open Data Streams technology for storing wire data in a NoSQL database, which itself could be configured as the go-to repository for unstructured Big Data prior to analytics processing. However, at every promotional turn, whilst the business analytics applicability is mentioned, it's done so almost in passing; the focus remains on ExtraHop's platform drawing upon these capabilities to bring 'Big Data thinking' (in terms of bringing together disparate data sets) to the field of IT Operations Analytics, i.e. doing so to undertake application performance, availability and security management better, not to improve business-level insight based on what can be learned from the data in these streams.

9. What's next?

ExtraHop is starting to provide industry-specific protocol layer decoder modules as a way of tailoring its platform to particular vertical markets. The company already cites a number of prominent healthcare customers, and a Healthcare Edition of its product (with real-time analysis of HL7 data) was announced in September 2014. Whilst promotion of this, along with the bulk of ExtraHop current marketing, still focuses on industry-specific application performance, availability and security monitoring, little mention is made of the potential for it to uncover business insights beyond IT Operations Analytics and situational awareness. It's not hard to see how a real-time window on transaction data pertaining to patients, hospital operations, and clinical experiences could become a valuable component of any Big Data analytics use case in a healthcare environment to improve patient services, especially as it can now be streamed into a NoSQL database, which a customer could use as its enterprise-wide analytics melting pot.

Whilst there's no corresponding specialist packaged analytics module for any other vertical yet, several industries are already well served by the protocol framework (and for which there's therefore a shortcut to verticalisation if considered a priority). These include financial services (ExtraHop can already decode the FIX protocol and covers many of the others often seen in banking environments – SSL, CIFS, etc.) and telecoms (the Diameter protocol can contain carrier and payment plan information amongst its Attribute-Value Pairs). ExtraHop's website also illustrates use cases across retail (focusing on HTTP, SSL, IBM MQ, SQL access, web service APIs), government, and SaaS Providers – though of course some of these (particularly SaaS Providers) will mainly be concerned with IT Operations Analytics (to help tune their infrastructure for customers), rather than wider business analytics connotations.

The company may decide to productise more of these as industry-specific offerings to enhance its appeal in key verticals, especially if it decides to devote resources to developing native decoders for those protocols deemed to be pervasive enough. There's no roadmap for this at present, but ExtraHop is working with its customers across these industries to explore priority protocols that may be ripe for decoder attention. Its payload analysis alone can go so far, even into protocol layers for which there isn't yet a native decoder, but it's a lightweight process which can't delve as deeply or reveal as much information since the hex format detail would be unknown to ExtraHop, and that's the extra discovery a dedicated decoder can bring.

10. Should I consider it?

Should you use ExtraHop's Operational Intelligence Platform for Big Data business analytics? The product is certainly technically well suited, and anyone using it for this purpose will have added a useful string to their Big Data analytics bow; but for the moment the company seems reluctant to pursue that cause, preferring instead to focus on promoting the IT Operations Analytics category (generally) and its own place as a significant player within it (specifically), lest notions of 'dual use' risk weakening its brand identity, and confusing its market positioning.

Organisations from some vertical markets (healthcare in particular, following the launch of the HL7 Analytics Module) will feel more comfortably well served by ExtraHop's protocol layer coverage. There's specialisation elsewhere too (e.g. financial services, telecoms), though not yet to the same extent; the focus of feature-richness here will inevitably lend weight to the product's applicability in some industries more than in ones not yet so specifically catered for.

If you have the aptitude and inclination to take a product currently being marketed primarily (though not exclusively) for one purpose, and put it to good use in a more business insight orientated Big Data context, your decision could pay dividends by bringing an innovative approach to real-time streaming data acquisition, consolidation, and analytic processing. However, you'll need to ensure that you have the in-house talent to realise the benefits for yourself since ExtraHop and its channel partners and resellers will remain focused on that primary IT Operations Analytics market for the time being, at least.

[i] Of course, that's not strictly true; not everything does "hit the wire" (for example, policy and configuration changes, resource utilisation metrics, and internal application execution details don't pass between systems on the network – but these aren't typical of the data and metrics most sought after in a Big Data business analytics use case). A fairer statement would be "all transactions hit the wire", and this is what ExtraHop is focusing on (which is good for when it's primarily the payload within application communications which are under scrutiny). Also, whilst (some of) the data might well hit the wire, it may also fall off before ExtraHop can tap into it – an overloaded network infrastructure should prioritise production traffic over mirrored traffic and since it's the latter which the ExtraHop platform feeds on, this will have the tendency to lower the quality of its input, leading to loss of integrity in wire data output.