Analyst Insight

Aberdeen Group
A Harte-Hanks Company

September 2012

Fighting Fraud with Big Data Visibility and Intelligence

For industries with a full-time focus on fighting fraud – such as banking, insurance and healthcare, among others – intelligence is invaluable. Although direct financial losses from fraud can vary significantly from one industry to another or in different parts of the world, the annual cost of fraud is substantial and the benefits of reducing it are very real. The problem is not that there is too little information, but too much – and most of it in disparate stovepipes and silos. Manual processes to aggregate, correlate and analyze this information are costly in terms of both time and resources, and often result in human error and crucial missed connections. Next-generation solutions for predictive analytics are solving the "big data" challenge, and are providing enterprises with the visibility and intelligence they need to move from post-incident forensics to a more proactive and predictive approach to fighting fraud.

Business Context: The Cost, and Complexity, of Fraud

In their *2012 Global Fraud Study*, the Association of Certified Fraud Examiners (ACFE) defines **occupational fraud** as "the use of one's occupation for personal enrichment through the deliberate misuse or misapplication of the employing organization's resources or assets." Based on their analysis of 1,388 cases reported by Certified Fraud Examiners between October and December 2011, the ACFE describes a taxonomy of 45 subcategories of occupational fraud, grouped into three major categories:

- **Corruption** – the misuse of influence in a business transaction in a way that violates duty to the employer in order to gain a direct or indirect benefit (e.g., schemes involving bribery or conflicts of interest)
- **Asset misappropriation** – the theft or misuse of the organization's resources (e.g., theft of company cash, false billing schemes or inflated expense reports)
- **Financial statement fraud** – the intentional misstatement or omission of material information in the organization's financial reports (e.g., recording fictitious revenues, understating reported expenses or artificially inflating reported assets)

In looking at these findings, two high-level points quickly become very clear: fraud is costly; and fighting fraud is complex:

Analyst Insight

Aberdeen's Analyst Insights provide the analyst's perspective on the research as drawn from an aggregated view of research surveys, interviews, analysis and industry experience.

Analyzing a Caseload of Fraud

The **Association of Certified Fraud Examiners** (ACFE) describes a taxonomy of 45 categories of occupational fraud, grouped into three major categories (corruption, asset misappropriation, and financial statement fraud).

In the realm of IT Security, this taxonomy is reminiscent of the *VERIS* model, created by **Verizon Business**, as seen in their *2011 Data Breach Investigations Report*. The VERIS "4 A's" model uniquely classifies each potential incident in terms of the *Asset* (what asset was affected), the *Action* (what action was taken on the asset), the *Agent* (whose actions affected the asset), and the *Attribute* (how the asset was affected) – resulting in a concise matrix of 630 distinct possible events.

Based on their 2010 caseload of 761 incidents, however, only 55 events were actually seen – which means that 91% of the threat-space was not in play.

Might a similar result be found in an analysis of the caseload data for occupational fraud?

- **Fraud is costly.** The average estimate of annual revenue lost to fraud is 5%, with a median loss per incident of $140K. More than 20% of all incidents led to losses greater than $1M.
- **Fighting fraud is complex.** With at least 45 different subcategories to be exploited, it's no surprise that fraud can be difficult to detect. The median time to detection was 18 months, with a range of 12 to 36 months.

Doing the simple mathematics of frequency times impact doesn't help much with the problem of complexity (see Table 1). Financial statement fraud, for example, is the least common (7.6%) but has the highest median loss ($1,000K), while asset misappropriation has the highest frequency (86.7%) but the lowest median loss ($120K) – and the choice of death by severe trauma or death by a thousand cuts leads to the same result.

Fast Facts

Public estimates for the average financial impact per incident can vary widely. Findings from Aberdeen's IT Security research include:

√ Blend of all incident types, from malware to loss or theft of intellectual property: $120K

√ Remediating an application security vulnerability: $300K

√ Data loss exposure: $640K

Table 1: The Complexity of Fraud – Occupational Fraud Frequency and Median Loss, by Type

	Corruption	Asset Misappropriation	Financial Statement Fraud
Definition	Misuse of influence in a business transaction in a way that violates duty to the employer in order to gain a direct or indirect benefit	Theft or misuse of the organization's resources	Intentional misstatement or omission of material information in the organization's financial reports
Example	Schemes involving bribery or conflicts of interest (6 subcategories)	Theft of company cash, false billing schemes or inflated expense reports (29 subcategories)	Recording fictitious revenues, understating reported expenses or artificially inflating reported assets (9 subcategories)
Frequency	33.4%	86.7%	7.6%
Median Loss ($K)	$250	$120	$1000
Weighted Loss ($K)	$84	$104	$76

Source: Association of Certified Fraud Examiners, *2012 Global Fraud Study*

How Fraud is Currently Detected is Inefficient

As shown in Figure 1, occupational fraud was most likely by far to be detected not by the organization's diligence and stewardship, but by virtue of an outsider tip, most commonly from employees of victim organizations. In general, "external" methods (including *tips*, *notification by law enforcement*, *external audits*, *accidental discovery* and *confessions*) are responsible for a higher percentage of detection and higher median losses than are "internal" methods (including *management review*, *internal audit*, *account reconciliation*, *document examinations*, and *IT controls*).

For these 1,388 cases, in fact, IT controls were the source of the *fewest* incidents detected. So fraud is costly, fighting fraud is complex … and current (mostly manual) methods are inefficient, even in a backwards-

Fast Facts

Attackers are becoming ever smarter, patient and disciplined; demographics have evolved to include:

√ Criminals (insiders, petty criminals, and organized crime)

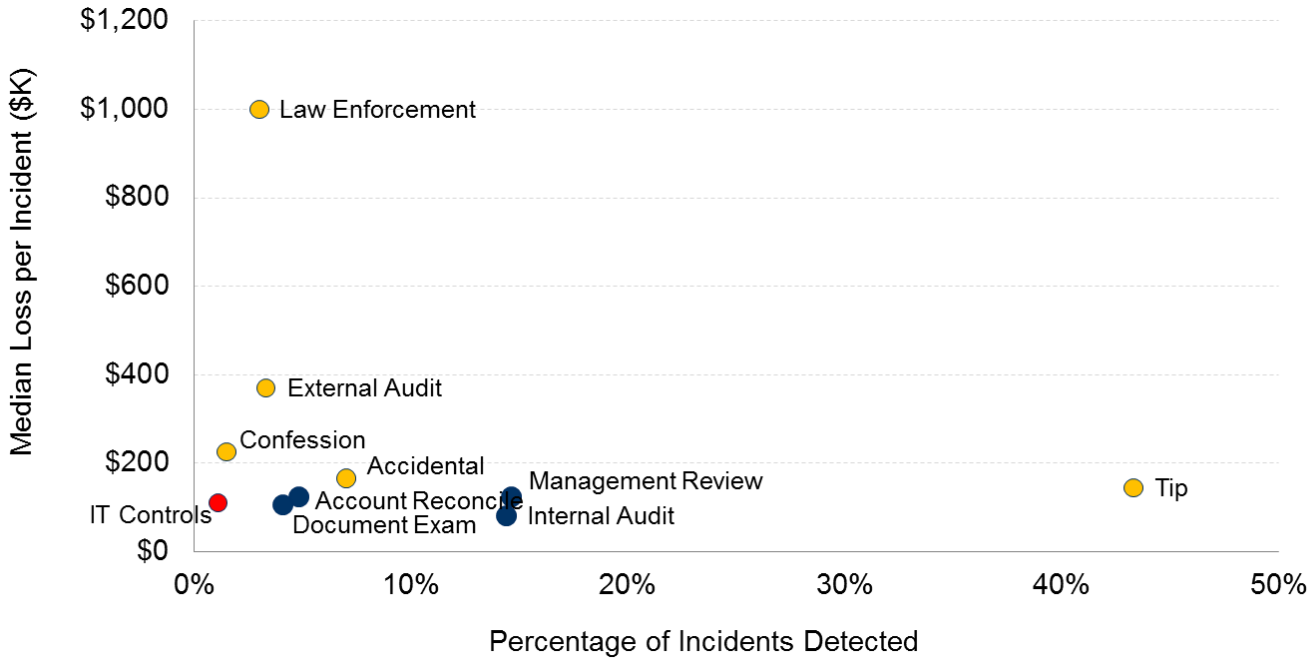
√ Terrorists

√ Anti-establishment *hacktivists*

√ State-sponsored initiatives

looking, forensic mode. This is a situation that cries out for a technology-based solution.

Figure 1: Fraud is Costly; Fighting Fraud is Complex; and Current Methods are Inefficient



Source: Association of Certified Fraud Examiners, *2012 Global Fraud Study*

But Success in Fighting Fraud Has Meaningful Rewards

A penny saved is a penny earned, as Benjamin Franklin made familiar in *Poor Richard's Almanac* in 1737 – and every penny of fraud loss recovered (or better yet, avoided) goes straight to the organization's bottom line.

A present-day example comes from the **Health Care Fraud and Abuse Control** program (HCFAC), under the US Department of Health and Human Services and the US Department of Justice. The HCFAC program was created under the **Health Insurance Portability and Accountability Act** of 1996 (HIPAA) to combat fraud and abuse in health care, including both private and public health plans.

Table 2: Health Care Fraud and Abuse Investigations and Returns

Investigations in FY2011	Criminal	Civil
New	1,110	977
Pending	1,873	1,069
Judgments and Settlements in FY2011	$2.4B	

Source: US Department of Health and Human Services and US Department of Justice, *Health Care Fraud and Abuse Control Program Annual Report for Fiscal Year 2011*, Feb. 2012

Over 5,000 investigations were in progress in FY2011, netting $2.4B in judgments and settlements. Between 2009 and 2011, the HCFAC program returned an impressive *$7.20 for every $1.00 expended*.

The description of the HCFAC program makes it very clear that fighting fraud is a multi-faceted initiative – incorporating all aspects of the *"people, process and technologies"* phrase that all of us have heard so many times. An illustrative example of a "program integrity" initiative under HCFAC includes the following elements:

- Developing risk assessment processes
- Identifying program vulnerabilities
- Conducting compliance and fraud audits
- Conducting ad-hoc studies and analysis with a special focus on select geographic areas
- Providing basic tips for consumers on how to protect themselves from potential scams
- Working with law enforcement, prescription drug plans, consumer groups, and other key partners to protect consumers and enforce rules
- Managing all incoming [tips and] complaints about fraud, waste, and abuse
- Utilizing new and innovative techniques to monitor and analyze information to help identify potential fraud
- Performing proactive research utilizing all available data to find trends in order to ferret out fraud, waste, and abuse activities

The last two bullets in particular are noteworthy, in that they highlight the critically important shift from fraud *detection and recovery*, to fraud *intelligence and prevention*.

New Approaches to Fighting Fraud are Imperative

The results of the 2012 survey on financial fraud published by the Information Security Media Group (ISMG) drives home the point that fighting fraud is a complex, multi-faceted initiative. Several types of financial fraud are displayed in Figure 2, based on the percentage of respondents experiencing each type of fraud in the last 12 months (on the x-axis), and the percentage of respondents indicating that they are well-prepared to detect and prevent each type of fraud (on the y-axis).

Visually, it's easy to see from Figure 2 that certain types of financial fraud occur frequently, but organizations feel relatively well-prepared – e.g., *check fraud*, *credit / debit card fraud*, *ACH / wire fraud*, *money laundering*.

It's also easy to see the cluster of several types of financial fraud that occur less frequently, and for which organizations do not feel prepared – and many of these are IT-based problems, such as *phishing* and *vishing*, *ATM*

Definitions

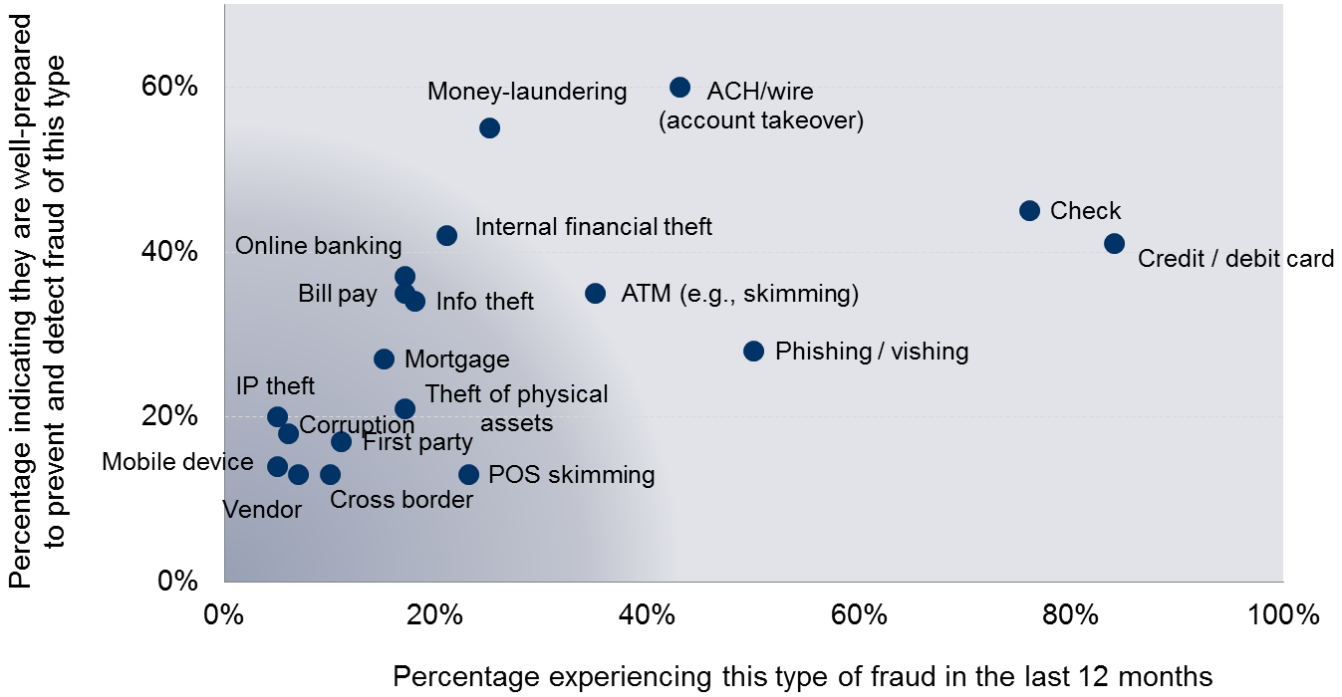
√ *Phishing* refers to the use of social engineering techniques for getting end-users to voluntarily give up private information. For example, the end-user may receive an email requesting that certain information be provided to resolve a problem with an account or verify a purchase.

√ *Vishing* (a combination of "voice" and "phishing") refers to the use of fake phone sites as part of the attacker's ecosystem for getting end-users to voluntarily give up private information. For example, the end-user may receive an email requesting that they call a toll-free number, or they may receive a phone call requesting that they call a toll-free number or visit a website.

√ *Smishing* (a combination of "SMS" and "phishing") refers to the use of short message service (SMS) text messages as part of the attacker's ecosystem for getting end-users to voluntarily give up private information. For example, the end-user may receive a text message requesting that they call a toll-free number or visit a website.

skimming, *online banking* and *bill pay*, *theft of intellectual property*, and *mobile devices*.

Figure 2: Frequency and Preparedness for Financial Fraud; Increasing Incidence of Cybercrime



Source: Information Security Media Group, *Faces of Fraud*, 2012 Fraud Survey

The inescapable conclusion is that it is increasingly difficult for enterprises to maintain high levels of preparedness simultaneously on all fronts. Rapid changes in information technology infrastructure require enhanced strategies for fighting fraud, specifically:

- From 100% success at prevention – to greater visibility, faster detection and incident response
- From "figure out what already happened" using post-incident forensics – to proactively "figuring out what's happening" using Big Data and predictive fraud intelligence.

No wonder that 3 out of 5 (61%) respondents in the ISMG study cited **fraud detection and monitoring systems** as planned investments in anti-fraud controls and measures over the next 12 months.

"We get good at looking at the historic impact of the [fraud] schemes, but none of us is great at being able to predict what the fraudsters are going to try next."

Sr. VP IT, Bank
ISMG, *2012 Fraud Survey*

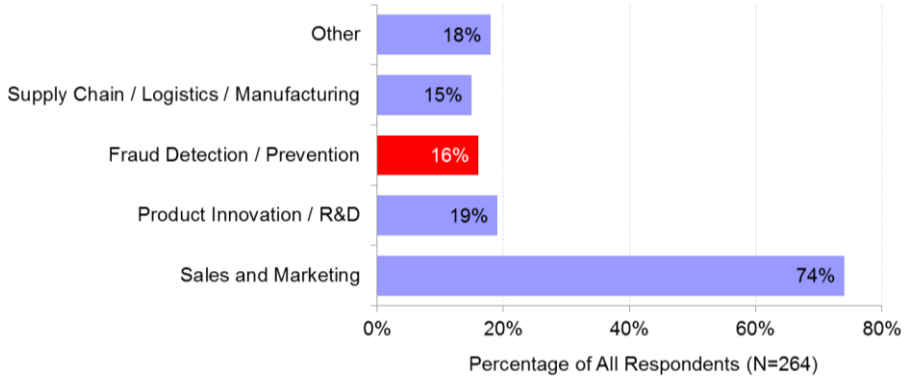
Aberdeen's Research Findings: Top Priority = Top Line

In Aberdeen's survey on *Predictive Analytics: Moving Beyond BI for Competitive Advantage* (1Q 2012), top line-oriented **sales and marketing** issues dominated the primary uses for predictive analytics, with specific drivers for current investments including:

- Tougher competitive environment (38%)
- Falling customer retention (29%)
- Increasing cost of customer acquisition (24%)
- Decreasing revenue (23%)
- Difficulty of forecasting demand (22%)
- Changing customer demographics (22%)
- Proliferation of channels (22%)

In contrast, roughly 1 in 6 (16%) respondents in Aberdeen's study indicated the current use of predictive analytics for the detection and prevention of fraud (Figure 3). The top priority is currently the top line, but as we have seen fighting fraud can yield significant bottom-line results.

Figure 3: Primary Use Cases for Predictive Analytics



Multiple responses accepted; does not add to 100%. Source: Aberdeen Group, 1Q2012

Aberdeen analyzed the responses from 29 companies using predictive analytics for the detection / prevention of fraud ("Fraud Users"), in comparison with 343 other organizations participating in the study ("All Others"). As seen in Table 3, companies using predictive analytics to fight fraud are investing more relative to all others – as part of their deliberate, strategic decisions to **reduce risk** and to reduce the net cost of fraud.

Fraud and the Supply Chain

"The production, packaging, and distribution of counterfeit software or hardware used by financial institutions or critical financial networks by cyber criminals could result in the compromise of proprietary data, system disruption, or complete system failure. Gaining physical and technical access to financial institutions could be accomplished by compromising trusted suppliers of technical, computer, and security equipment, software, and hardware.

Financial firms have become regular targets of supply chain attacks. For example, ATMs have been delivered with malware installed on the systems, fake endpoints on the ATM networks have been created, and individuals have posed as ATM maintenance workers. Additionally, vendors who supply services to the banking and finance sector are constant targets of cyber criminals, including those who provide services like security, authentication, and online banking platforms."

Gordon M. Snow, Assistant Director, FBI Cyber Division

Statement before the House Financial Services Committee, September 2011

Table 3: Fighting Fraud Requires Additional Resources, But Directly Benefits the Bottom Line

	Fraud Users	All Others	Difference
Percentage of annual revenue lost to fraud	1.5%	N / A	-
Increasing exposure to risk is a driver for current investments in predictive analytics	31%	6%	5-times
Identify high-risk customers and business activities is a strategy behind current investments in predictive analytics	38%	13%	3-times
Full-Time Equivalent staff or consultants involved in building and deploying predictive analytic models (FTE per 10K employees)	3.6 (47% increase)	2.4 (12% increase)	1.5-times

Source: Aberdeen Group, 1Q 2012

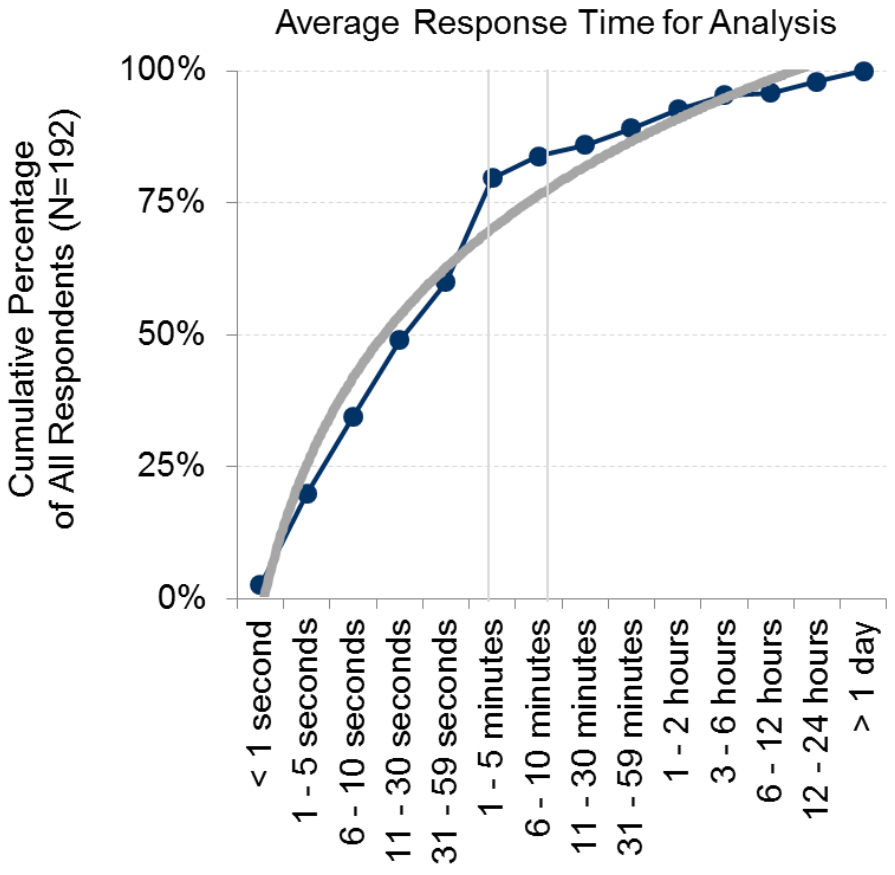
The average percentage of annual revenue lost to fraud by the Fraud Users in Aberdeen's dataset was 1.5% – recall that it was 5% in the ACFE study. Compared to all others, the current investments in predictive analytics by the Fraud Users were 5-times more likely to be from **increasing exposure to risk** as a driver, and 3-times more likely to be from **identifying high-risk customers and business activities** as a strategy. As we saw in the HCFAC example, incremental investments in fighting fraud can yield impressive returns that fall straight to the organization's bottom line.

It's worth noting that lack of data to be analyzed is not the problem; on the contrary, there is an overabundance of data – from both internal sources and external sources, both structured and unstructured, across multiple channels – and most of it is located in disparate stovepipes and silos. In other words, the data for successful use of predictive analytics solutions already exists … it's just not as integrated, available and easily accessible as needed. In Aberdeen's study on *Maximizing the Value of Analytics and Big Data* (September 2012), the average amount of active data (i.e., not archival or backup) in enterprise repositories was 730 terabytes, and growing quickly.

Definitions

√ The term **Big Data** has come to refer to the rapid growth of business data – in terms of not only *volume*, but also the variety of *formats* and the *speed* at which it needs to be captured and analyzed.

Figure 4: Average Response Time for Big Data Analytics



Source: Aberdeen Group, September 2012

The *integration* of data for analysis is a different story. On average, the percentage of data currently accessed for analysis was about 26%, but the percentage of data *desired* to be accessible for analysis was twice as high (52%). Similarly, the *performance* of predictive analytics systems can be an issue: respondents in Aberdeen's study indicated on average that 25% of dashboard refreshes and queries take too long. By inspection (Figure 4), this suggests that the maximum acceptable response time for analysis is somewhere in the neighborhood of 60 seconds.

Solutions Landscape (illustrative)

Solution providers for predictive analytics in the context of Big Data can range from smaller specialists to multi-billion dollar firms. For illustrative purposes, the solution providers with the highest *aided awareness* (i.e., respondents from Aberdeen's *Predictive Analytics* study had heard of these solutions) and the highest *footprint* (i.e., respondents had already deployed these solutions) are summarized in Table 4.

Table 4: Solution Providers for Predictive Analytics in the Context of Big Data (illustrative)

- IBM
- SAS
- Business Objects (SAP)
- MicroStrategy
- Nielsen
- KXEN
- TIBCO (Spotfire)
- Acxiom
- Siemens
- FICO
- Pitney Bowes (Portrait Software)
- Rapid Insight
- ThinkAnalytics
- Eloqua
- SAF AG

Source: Aberdeen Group, September 2012

Solution Selection Criteria

Leading factors identified by companies using Big Data and predictive analytics for the detection / prevention of fraud:

√ Automated data integration 54%

√ Ease of use 42%

√ Does not require specialized expertise 38%

√ Scales to accommodate future growth 31%

√ Ease of integration 27%

√ Performance 19%

Summary and Recommendations

- **Fraud is costly.** Direct financial losses from fraud can vary significantly from one industry to another or in different parts of the world, but the annual cost of fraud is substantial – in the ACFE *2012 Global Fraud Study*, an average of 5% of annual revenue, with a median loss per incident of $140K.
- **Fighting fraud is complex.** In general, external sources are currently found to be responsible for a higher percentage of detection – and higher median losses – than are internal sources and methods, with time to detection ranging from 12 to 36 months.
- **Success in fighting fraud pays off.** Every penny of fraud loss recovered (or better yet, avoided) goes straight to the organization's bottom line. In the example of HCFAC, every $1 expended in fighting fraud returned an impressive $7.20 in judgments and settlements.
- **Current (mostly manual) methods are inefficient.** Even in a backwards-looking forensic mode, current IT controls were found

to be the source of the fewest incidents detected. This is a situation that cries out for a technology-based solution.

- **New strategies for fighting fraud are emerging.** Rapid changes in information technology infrastructure are increasing the difficulty of maintaining high levels of preparedness simultaneously against all threats. In response, organizations are adopting enhanced strategies for fighting fraud: from 100% success at prevention, to greater visibility, faster detection and incident response; from "figure out what already happened" using post-incident forensics, to proactively "figuring out what's happening" using Big Data and predictive analytics.
- **Solution providers are leveraging "big data" for predictive analytics.** The problem is not that there is too little information, but too much – and most of it in disparate stovepipes and silos. Next-generation solutions for predictive analytics are solving the "big data" challenge, and are providing enterprises with the visibility and intelligence they need to move from post-incident forensics to a more proactive and predictive approach to fighting fraud.
- **Crawl, Walk, Run is a proven, pragmatic approach.** In the beginning, initiatives to leverage Big Data and predictive analytics for fighting fraud can sometimes get bogged down in debate over the optimal approach. Both of the following are examples of the pragmatic "crawl / walk / run" approach that is characteristic of the companies who are the most successful in their enterprise-wide initiatives, as seen consistently in Aberdeen's research:
    - Taking more time to integrate all data sources for a single application / process / workflow
    - Integrating fewer data sources that apply more broadly, and making rapid progress in the ability to analyze, understand and take meaningful action

Aberdeen Group
A Harte-Hanks Company

For more information on this or other research topics, please visit www.aberdeen.com

Related Research	
Go Big or Go Home? Maximizing the Value of Analytics and Big Data; September 2012	*Predictive Analytics: Moving Beyond BI for Competitive Advantage*; 1Q 2012
The State of Big Data; September 2012	*The State of IT (In)Security, and How to Avoid Costs by Spending More*; November 2010
Real-Time Data Integration: Driving Near Real-Time Analytics; September 2012	

Author: Derek E. Brink, Vice President and Research Fellow, IT Security and IT GRC (Derek.Brink@aberdeen.com)