SAFARI

Situational Awareness Framework for Fraud Detection

Platform Description

Situational Awareness for Fraud Detection

Many data analysts rely in ACL-like data analytics packages for the detection of fraud within the data managed by their ERP (Enterprise Resource Planning) systems. However, once tools such as ACL are run, an auditor still needs to find individual suspicious transaction associated with fraud to prove fraud and make a legal case. Usually, doing a run does not identify the suspect case, but rather points out a broad area of the possible fraud location where fraudulent transactions should be manually found.

An effective data-driven approach for fraud detection on millions of unlabeled payment registers should incorporate different views or perspectives of available data. To achieve this, SAFARI introduces the concept of situational awareness (SA), conceived as the "information gathered from a variety of sources that, when communicated to emergency managers and decision makers, can form the basis for incident management decision making".

SA builds a path from massive data to information to understanding, allowing for appropriate analysis and sharing at each point of the fraud detection process.

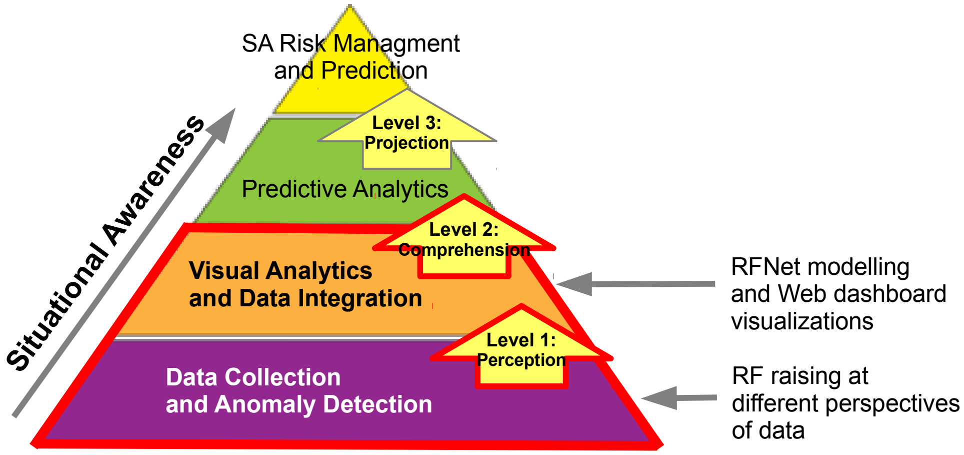


Figure 1. SAFARI SA approach (Level-1 and Level-2).

Current Approach:
Levels 1 and 2 Situational Awareness

MIT GDC is exploring the SA approach for fraud detection by exploiting a software prototype that implements the first and second levels of a three-level SA solution (Fig. 1):

> **SA Level-1: Environment perception**. This is the basal constituent of SA: the data is analyzed by exploiting domain-specific anomaly detection techniques on different perspectives of the data (personal information, monetary quantities, spatial information, etc.) in order to produce red flags (RFs) that point out to suspicious yet not conclusive and disparate behavior on unlabeled data.
>
> **SA Level-2: Situation comprehension.** Is the combination, interpretation, and retention of anomalies to form a coherent picture of the situation whereby the significance of events is understood. This is achieved in SAFARI by exploiting the RFNet approach and rich Web-based Visual Analytics techniques.

Red Flag Networks for Evidence Integration

To implement the concept of SA into a flexible fraud detection work-flow, SAFARI implements a novel yet intuitive Red Flag Network (RFNet) link analysis approach (Fig. 2) to integrate the data generated by a variety of anomaly detectors. An RFNet is a network of entities (e.g. payment documents) where entities are connected if they participate in automatically raised red flags (e.g. payment documents whose vendors are in the same geographical location or whose legal names are suspiciously similar).

Existing link analysis solutions, such as those from Palantir, SAS Institute, and Centrifuge Systems, offer social network capabilities to construct social networks from different sources of data. However, networks have to be manually built by SMEs and anomalies have to be manually discovered, severely limiting the quantity of data and the insight that can be achieved in a single session.

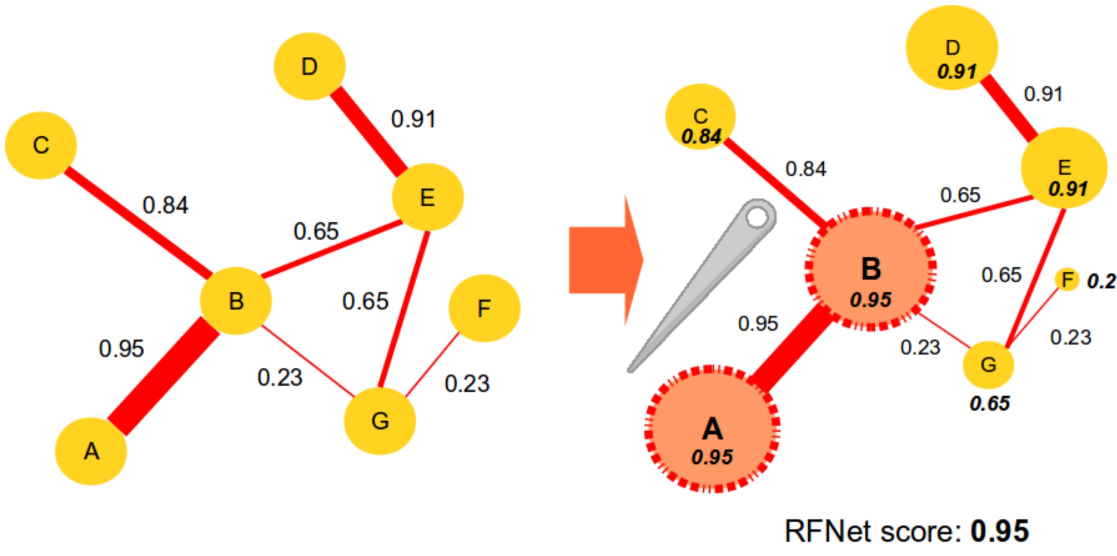


Figure 2. RFNet SAFARI approach

In contrast, SAFARI's RFNet-based modeling approach is to connect disparate but potentially related entities and red flags to automatically build fully contextualized potential fraud cases from hundreds of thousands of registers and automatically detected anomalies. In addition, the RFNet approach leverages the advantages of well-established network science techniques, such as node centrality quantification, successfully used by government agencies, consulting companies and researchers for the investigation of a wide variety of criminal activities.

Belief Networks for Risk Ranking

The occurrence of any RF does not necessarily indicate that fraud has occurred. Rather, they identify the possible occurrence of fraud and lead the analyst to look at the right places. Fraud schemes arise when the RFs are related and the case is solved when the loose ends are tied up.

To achieve this, SAFARI implements a Bayesian Belief Network (BBN) approach for data fusion (Fig. 3). A BBN can be thought of as a probabilistic graphical model that allow us to deal with fraud occurrence uncertainly by representing causal relationships among various red flags to which observed occurrence is posted as evidence.

The BBN is retrained in an unsupervised way using concrete evidence (e.g. observed RFs between two documents) to recalculate the probability values for the BBN nodes. MIT is experimenting with the BBN approach to encode the expertise of SMEs, correlate the occurrence of RFs, and measure the probability of the occurrence of fraudulent activity.

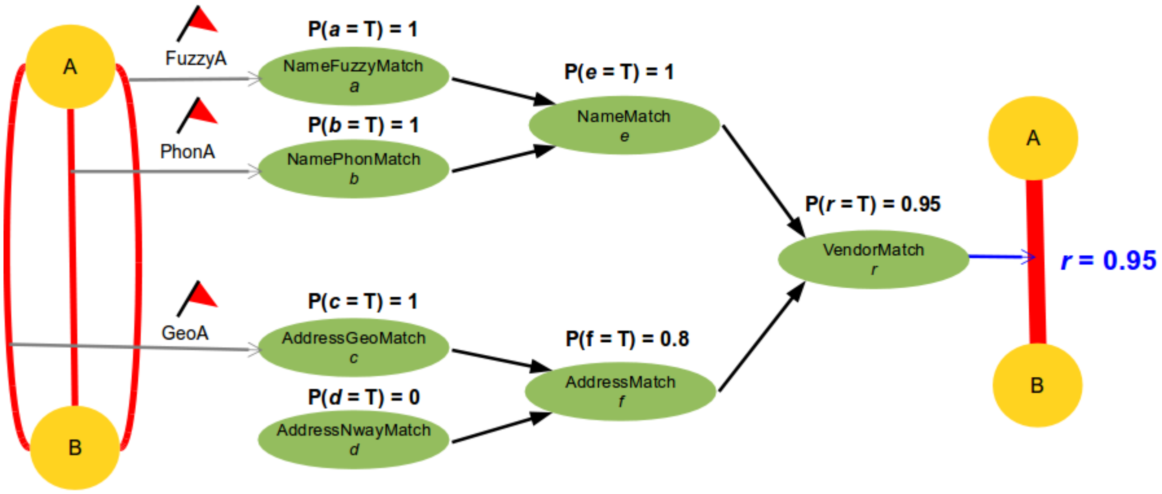
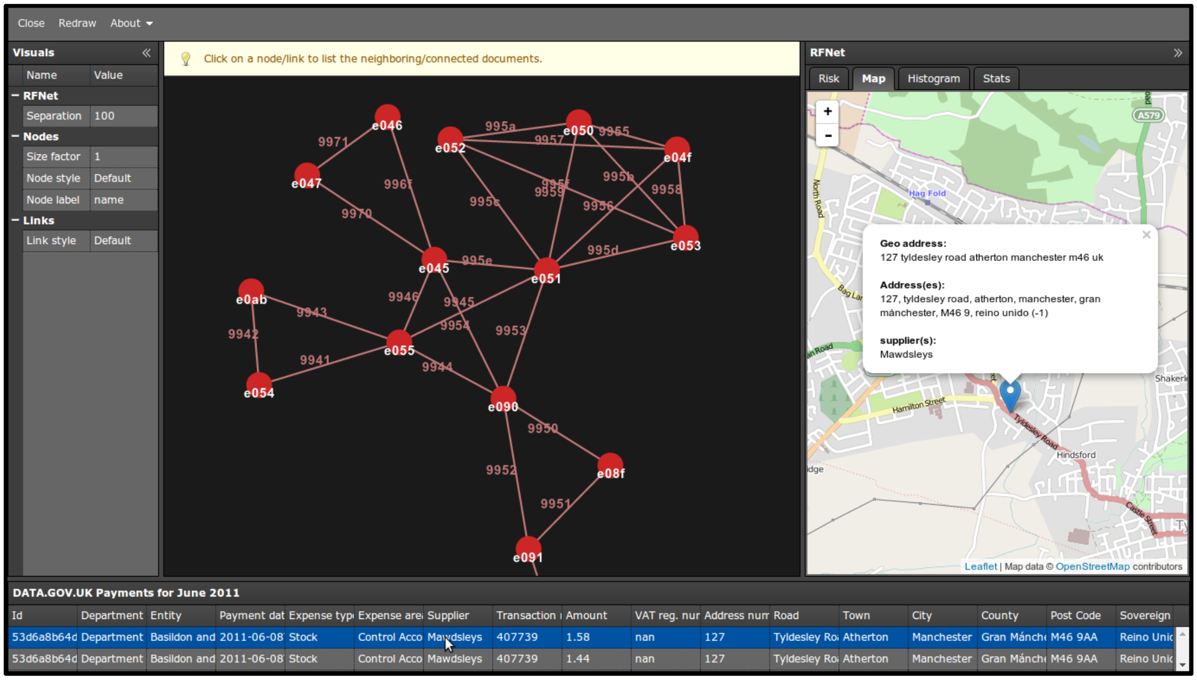


Figure 3. BBNs calculates fraud rankings for pair-wise matchings.

Rich Web-Based Visual Analytics

SAFARI's RFNet and map data visualizations are enablers that provide insights into the related entities being inspected and unveil relationships that are hard to be discovered otherwise. The objective of SAFARI data visualizations is to help SMEs in perceiving patterns, gain insights into the data being inspected and make sense of disparate but potentially related phenomena in order to gather evidence to build a strong fraud case.



Click on a node/link to list the neighboring/connected documents.

Name	Value
RFNet	
Separation	100
Nodes	
Size factor	1
Node style	Default
Node label	name
Links	
Link style	Default

Geo address:
127 tyldesley road atherton manchester m46 uk

Address(es):
127, tyldesley road, atherton, manchester, gran mánchester, M46 9, reino unido (-1)

supplier(s):
Mawdsleys

DATA.GOV.UK Payments for June 2011

Id	Department	Entity	Payment dat	Expense typ	Expense are	Supplier	Transaction	Amount	VAT reg. nur	Address num	Road	Town	City	County	Post Code	Sovereign
53d6a8b64d	Department	Basildon and	2011-06-08	Stock	Control Acco	Mawdsleys	407739	1.58	nan	127	Tyldesley Ro	Atherton	Manchester	Gran Mánche	M46 9AA	Reino Unic
53d6a8b64d	Department	Basildon and	2011-06-08	Stock	Control Acco	Mawdsleys	407739	1.44	nan	127	Tyldesley Ro	Atherton	Manchester	Gran Mánche	M46 9AA	Reino Unic

Figure 4. SAFARI Web-based GUI

SAFARI implements a user interface inspired on the ideas of a dashboard system to show a complete set of information (relationships, geographical locations, tabular data) and facilitate the prioritization of tasks, the evaluation of red flags and the sense making process of fraud scenarios (Fig. 4). For maximum reach, the graphical user interface is delivered by using state-of-the-art Web technologies and protocols. As a consequence, SAFARI is accessible without the need of a specialized client or runtime. A browser on desktop will be sufficient to run the full session.

Flexible and Extensible Framework

SAFARI's flexible framework architecture offers SMEs and decision makers the best of different worlds in an unified platform: link analysis ideas inspired in Palantir, Centrifuge Systems, SAS Institute and BKD Forensics solutions; geolocation analysis/visualization capabilities inspired in Palantir products; and a Guardian Analytics-like scoring concept to rank entities involved in a suspicious scenario.

Additionally, the SAFARI platform and work-flow (Fig. 5) can be upgraded and extended to take advantage of new advances and add more capabilities in data and visual analytics for fraud detection. Being a flexible software framework, SAFARI can keep up with business evolution and dynamics, having the potential to be a valuable advisor in decision-making in both the short and the long-term.

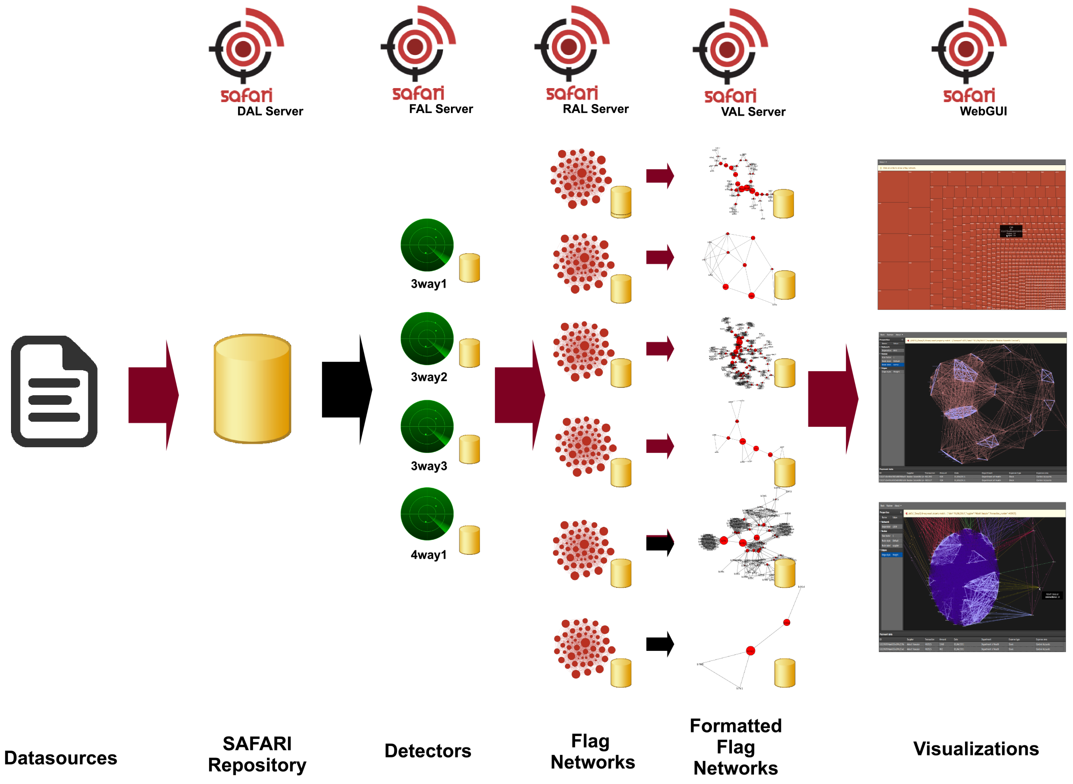


Figure 5. SAFARI workflow

Our Vision:
Holistic View of Risk Scenarios

In SAFARI, we envision a future where different data sources, data analytics techniques, and visualization infrastructure is integrated in a way that gives SMEs and decision makers a holistic and accurate view of fraudulent behavior at multiple levels of granularity. We believe that by collecting and integrating data produced by a variety of financial anomaly detection techniques SAFARI can prevent financial and government institutions costly fraud-related incidents caused by poor situational awareness.

Additionally, we believe that the SA-based SAFARI approach can be useful for detecting other type of financial deceptions, such as credit card and health insurance fraud. Even more, SAFARI can be adapted to be applied in a wide variety of environments where the assessment of risk is critical for decision making, such as the risk assessment of scale formation in oilfield facilities, the investigation of criminal activity, counter-terrorism, computer cyber-security analysis, and the minimization of industrial health/safety/environmental incidents.

SAFARI Team at MIT GDC

Alberto Garcia-Robledo
Abel Sanchez
Rongsha Li
Juan-Carlos Murillo-Torres
John Williams
Sascha Boheme