2014

The bad guys are using it, are you?

Hong-Eng Koh
Oracle Corporation, Singapore, China Public Security University, China, hong-eng.koh@oracle.com

DOI: 10.4225/75/57a830e7c833c
Originally published in the Proceedings of the 7th Australian Security and Intelligence Conference, Edith Cowan University, Perth, Western Australia, 1-3 December, 2014

THE BAD GUYS ARE USING IT, ARE YOU?

Hong-Eng Koh
Oracle Corporation, Singapore
China Public Security University, China
hong-eng.koh@oracle.com

Abstract

From Occupy Wall Street to 2011 England riots to Arab Spring to Mumbai 26/11 to the ethnic cleansing rumors in India and increasingly used by pedophiles, social media is a very powerful tool for pedophiles, troublemakers, criminals and even terrorists to target individuals and even to go against the establishment. On the other hand, social media can save lives in a disaster, and its a natural extension of community policing or engagement. Community engagement is a must-have strategy for any public safety and security agency. However, this strategy requires the removal of stovepipe processes and systems within an agency, allowing better process integration and information sharing, thereby offering improved services to all stakeholders. Improved information sharing within an agency is also a stepping stone towards cross-agency information sharing. Criminal and terrorist organizations do not follow and do not commit crimes and terrorism based on how public safety and security agencies are structured. They commit crimes based on maximum returns, be they financial ideological or to instill fear. This is why information sharing or intelligence fusion is crucial across agencies especially with more illicit use of technologies and social media by the 'bad guys'.

Keywords
Big data, analytics, social media, mobility, security, ontology, semantic, pattern recognition, event processing, data discovery, sentiment analysis, business intelligence, Hadoop, NoSQL.

INTRODUCTION

Over the past twenty years the level of organized crime and terrorism directed against states, groups and individuals has increased significantly. It is not just the number, but also the sheer scale of attacks perpetrated that has grown. A glaring example is ISIS. In many cases, responsibility for these attacks lies within a new breed of criminal and terrorist organizations that has emerged during this period. These organizations have a global reach and employ sophisticated techniques to spread their doctrines and to organize, plan, communicate and execute organized crime and terrorist acts. They exploit the widespread availability of global mobile communications and take advantage of World Wide Web based applications such as email, websites, blogs and social networking to develop and maintain contact with globally dispersed networks and cells, and to direct operations.

On the other hand, social networking can save lives in a disaster, and it is a natural extension of community policing or engagement. Community engagement is a must-have strategy for any public safety and security agencies. The power of social networking from Boston Marathon[1] terrorist attack to China Szechuan Ya'an[2] earthquake have been illustrated.

To combat these new threats or master social networking for better community engagement, intelligence and law enforcement agencies need to be able to digest, analyze and make sense of hyper volumes of data, including structured and unstructured, in real-time speed. And also the ability to sieve out noises, coded phrases and false positives from open source intelligence.

A good smart surveillance and intelligence system has to address major requirements such as:

[1] See http://www.fema.gov/media-library/assets/documents/33747
[2] See http://www.tealeafnation.com/2013/04/social-medias-role-in-earthquake-aftermath-is-revealing

Event Processing & Pattern Recognition

With huge quantities of data flowing into the system, it is vital that it can automatically and continuously monitor the data flows for specific data occurrences. Through the establishment of policies and rules, these data occurrences can be identified as significant events requiring further attention by an appropriate analyst. For example, the system could automatically monitor all incoming data against watch lists, pushing the data to the appropriate analysts immediately when an event is detected. Through the use of natural language processing, increasingly sophisticated rules can be quickly developed to describe more complex events, reducing the number of false positives and more accurately determining the correct level of threat.

Pattern matching is the ability to look across many events and look for patterns of occurrences. For example, one or two mentions of a public place over a day in a social network feed maybe insignificant, but 10,000 in one hour could be indicative of an unexpected and potentially unlawful gathering.

Ontology-based Expert System

Ontology is the semantic representation of an agreed vocabulary of information terms, meanings, concepts and relationships within a particular domain. The defined concepts and relationships represent knowledge about that domain. The scope of ontology can be as broad as intelligence or as specific as terrorism, cyber crime, explosives or money laundering, and ontology may span several domains. An ontology represented in this way is used in extracting entities from documents and forms the basis of search, reasoning and inferencing functionality in the solution.

Ontologies from different sources can be combined to represent the full spectrum of domain expertise both inside and outside the intelligence and law enforcement agency. The combination of data and the ontology (i.e. the meaning of the data) is what makes the knowledge base a semantic data model. This enables machine driven inference models to be applied to the semantic data model to further extend the knowledge base, revealing relationships that were previously implied or non-obvious.

Discovering the Unknowns

Traditional predictive analytics using business intelligence tool depends on expert knowledge on cause-effect relationship in order to develop the modeling required to predict the outcome. What if we do not know what we do not know; do not know what questions to ask; do not know the cause-effect relationship? With new technologies and crowd-sourcing (or rather crime-sourcing) through social networking, such unknowns are very real.

This is where we need a new class of technology that allows analysts to slice, dice, search and hypothesize massive volume of data from multiple sources, to discover the unknowns and patterns.

Big Data Ready

In 60 seconds, there are 98,000 tweets, 695,000 Facebook updates, 370,000 minutes of Skype calls, 13,000 iPhone applications downloaded, 600 new YouTube videos, 168 million emails sent, 694,445 Google searches, 6,600 new photographs uploaded on Flickr, and the list goes on.

The effectiveness of the above three analytical features are limited if the underlying infrastructure is not Big Data ready. We are talking about analyzing massive volume of structured and unstructured data in the speed of light, and one that can be executed with very efficient infrastructure and data center. Furthermore, the current state of Big Data sees organizations having separate data sources such as traditional relational database, NoSQL and Hadoop, each having different query tool. We need a familiar and efficient tool, such as SQL query, to holistically search and analyze across the various data sources and platforms.

Cyber Security

One cannot safeguard privacy without cyber security. A system of this nature deals with very sensitive data. It needs to include the functionality required to protect the system from all internal and external access violation and denial of service threats, and to ensure that all information, physical and software assets are not misused or corrupted. The activities of intelligence and law enforcement agencies are of general interest to the public but of the greatest interest to foreign powers, terrorists, criminal organizations and computer hackers. Therefore, surveillance and intelligence system must adhere to the highest levels of security in order to maintain the sensitivity of information it holds as well as the secrecy of its own operations and of collaborative operations it carries out with other intelligence and law enforcement agencies.

Security functionalities need to be based on de facto security standards and industry best practice. Starting with identity management to facilitate the implementation of security policies governance, authentication, and authorization through the centralized administration of identities together with the associated mapping of roles, credentials, rights and entitlements.

There needs to be fine grained access control to relational data, both at the row and column level, and encryption of all data online, at rest and in motion. And capability to reduce the risk of insider threats by preventing highly privileged users, e.g. database administrators, from accessing application data, enforcing separation of duty, and providing controls over who, when, where, and how applications, data and databases can be accessed. Security design from grounds up needs to have centralized management of audit settings and automates the collection, consolidation and securing of all audit data. It also produces standard and customized compliance reports.

This paper presentation will come with various demos of social media monitoring and analysis.

CONCLUSION

Each intelligence solution is a unique complex solution. It must capture, process and maintain accessible online hyper-volumes of data. In doing so it must support the work of all types of staff through automating labor intensive and time consuming tasks carried out under strict security policies. Implementing a solution invariably takes place in a heterogeneous environment and involves integration with existing legacy systems and the introduction of new technologies.

This paper addresses these requirements. It is a design reference for building the next generation intelligence solution, which is repeatable, open, flexible and extensible, and is based on a technology stack, which is available today. Its use of semantic technology and domain specific ontologies provides more powerful automated insight and inferencing than is available in purely relational and NoSQL technologies. Complex security policies and business rules can be defined through natural language documents and maintained on the fly by business users. The system is performant and highly scalable, both in terms of processing and data storage and data is always available online. The system is protected from internal and external threats by an end-to-end security. It also detects misuse and abnormal behavior

REFERENCES

Intelligence Hub & Alerts, Hong-Eng Koh, Oracle Corporation May 2013
http://www.oracle.com/us/industries/public-sector/intelligence-hib-alerts-br-1536069.pdf

Boston Marathon Terrorist Attack:
http://www.fema.gov/media-library/assets/documents/33747

Szechuan Ya’an Earthquake:
http://www.tealeafnation.com/2013/04/social-medias-role-in-earthquake-aftermath-is-revealing