THE SECURITY STANDARD

Adapting Enterprise Security to New Realities, Threats and Endpoints

September 10-11, 2012 | New York Marriott at the Brooklyn Bridge | New York City

Produced by
CSO

Big Data for Security Intelligence

Ravi Devireddy
Senior Business Leader
Security Analytics
Visa Inc

What is Big Data

“Big Data refers to datasets whose size and/or structure is beyond the ability of traditional software tools or database systems to store, process, and analyze within reasonable timeframes”*

Characteristics of Big data (IBM) :

1.	Volume	From Terabytes to Zettabytes
2.	Variety	From relational data to semi-structured or unstructured data
3.	Velocity	From batch to streaming data

Credit: Understanding Big Data, Eaton et al. (IBM definition)

* McKinsey global institute – Big data – may 2011

Big Data - Hadoop

- Hadoop

  Is a computing environment built on top of a distributed clustered file system (HDFS) that was designed specifically for large scale data operations (MapReduce)

- MapReduce

  Is a programming framework, in which work is broken down into *mapper* and *reducer* tasks to process data that is stored across a cluster of servers for massive parallelism

Big Data Analytics

Big Data Analytics is the application of advanced analytic techniques to very big data sets

- **Data Science**

  “The ability to take data and be able to understand it, to process it, to extract value from it, to visualize it, and to communicate it” – Hal Varian

  “Data Science is a blend of hackers arts, statistics and machine learning” – Hilary Mason

- **Data Mining and Machine Learning**

  Analysis of large quantities of data to extract previously unknown interesting patterns such as groups of data records (cluster analysis), unusual records (anomaly detection) and dependencies (association rule mining)

- **Data Visualization**

  is the study of the visual representation of data to graphically illustrate data to understand and glean insights from the data.

Big Deal about Big Data

- The world is creating ever more data
  - Large Hadron collider generates 40TB/Sec
  - 30 billion pieces of content shared on facebook every month*
  - By 2013 the amount of traffic flowing over the internet annually will reach 667 exabytes

- Machine Data (Data Exhaust) is one of the fastest growing segments of big data
  - Website click streams
  - Network devices
  - IT Infrastructure
  - Mobile devices

The Economist

Obama the warrior
Misgoverning Argentina
The economic shift from West to East
Genetically modified crops blossom
The right to eat cats and dogs

The data deluge

AND HOW TO HANDLE IT: A 14-PAGE SPECIAL REPORT

Credit: McKinsey Global Institute, May 2011

Information Security Philosophy

“ There are known knowns; there are things we know that we know.

There are known unknowns; that is to say there are things that, we now know we don't know.

But there are also unknown unknowns – there are things we do not know, we don't know. “

Information Security Philosophy

known knowns	known unknowns	unknown unknowns
Rule based	Correlations	Intelligence
Signature based	Trends	Context
Dashboards	Analysis	Data Science

Big Data for Information Security

| | unknown unknowns

| | Intelligence

Context

Data Science

Information Security is a big data problem

- Volume, variety and complexity of the data is growing rapidly

  Vulnerability scans, configurations, identity and access, log data, threat Intelligence feeds, network flow and packet analysis, user activity, database activity, transaction data, operational data, etc.,

- Security intelligence requires interaction, correlation and integration of various security tools and data for increased accuracy, optimized decision support, and risk based prioritization

- Gradual shift from monitoring 'silos' towards more comprehensive and integrated approach.

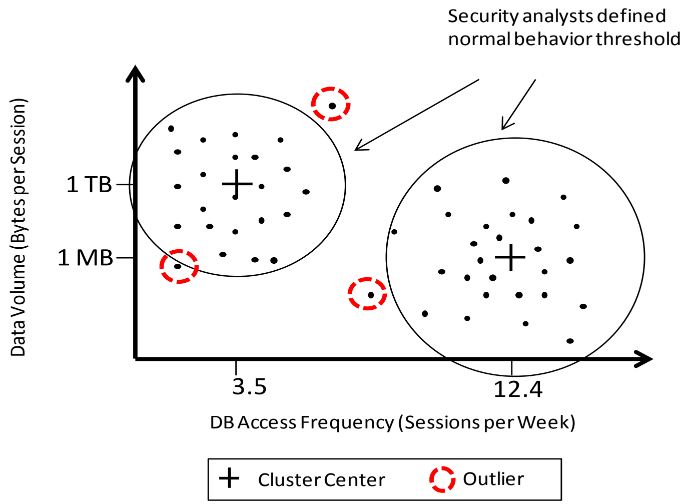
Applications of big data for Information security

- Contextual Information for optimized decision support, investigations, forensics and response
  - Enable information security analysts and incident responders to be more effective by providing a comprehensive view of security data
  - Overcome ‘silod’ data, monitors and scanners
  - Security data warehouse
  - Historical patterns and trends
  - Predictive analytics
  - Deeper drill down with detailed data

Applications of big data for Information security

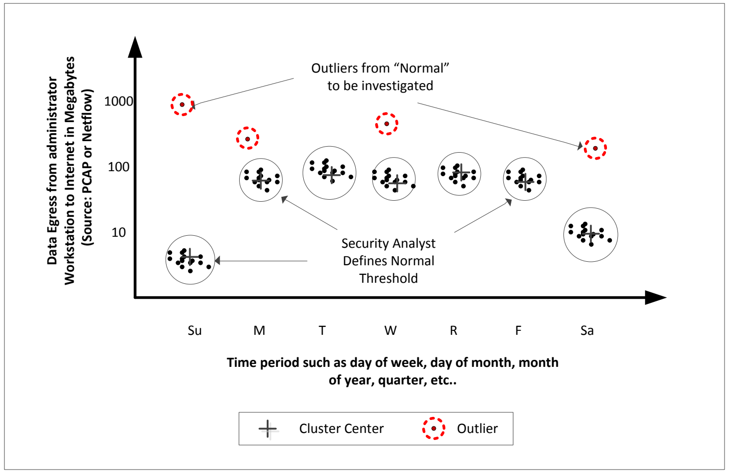
- Detect advanced targeted attacks
  - Signature based patterns may not detect
  - Anomaly detection systems model normal or expected behavior in a system, and identify outliers or anomalies by detecting deviations of interest that may indicate a security breach or an attempted attack.
  - application of statistical segmentation, association rule mining and clustering algorithms

Exhibit 1: Two User Behavior Clusters for "Normal" DBA behavior
Illustrative Example



Security analysts defined normal behavior threshold

Data Egress to Internet
Illustrative Example (Continued)

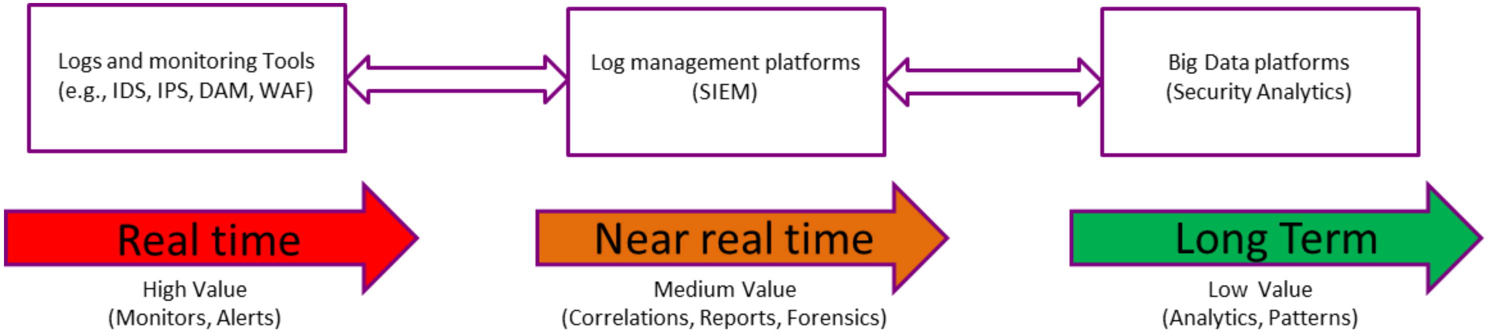


Outliers from "Normal" to be investigated

Security Analyst Defines Normal Threshold

SIEM 2.0

- SIEM (Security Information and Event Management) to evolve into a comprehensive security analytics platform
- Overcome current limitations on storage and processing capacity of SIEMs with big data technologies
- Time decay value of a log or event record



Challenges

- Big data technology maturity
- Difficulty of architecting a big data analytics system and problems with making the data usable for end users
- Batch oriented or near real-time
- Inadequate analytics tools, algorithms and applications
- Security, compliance and risk
- Shortage of talent
    - Technical skills and expertise in statistics and machine learning
    - Data savvy managers and analysts to frame right questions and act on insights from big data
    - Engineering and support of big data platforms and analytics tools

Challenges

- Access to data, need to integrate information from multiple sources
- Interoperability : Inadequate standards for integration of security scanners and monitors
- Lack of industry wide best practices for collecting, storing and querying security data and contextual information.
- Security tools that do have an API, Query or Export functionality of the data.
- Data silos and data integration challenges

Prepare for Security Analytics with big data

- Strategic Objectives of Information Security program
  - Data Breaches ? Insider Threats ? IP theft ? Fraud ?
- Consider the potential benefits of security analytics with big data
- Problem definition – Clarity
  - Before starting the 'how'
  - Start with 'Why' and 'what'

  Otherwise big data gives us just that : 'lot of data'
- Review the tools in use and data availability
- Interaction, Correlation and Interoperability as a criteria of selection for security tools

Security data scientist = Security specialist + Data analytics

- Emergence of ‘Security data scientist’
- Emerging role focused on applying scientific or mathematical analysis on large data sets to support security analytics
- Strong academic background in mathematics or statistics, with experience in information security functions, and passion for data science, data mining and machine learning.
- Design algorithms, build models, analyze and interpret the information by using mathematical or statistical methods and applying machine learning methods to detect anomalies or deviations

Recommendations

- Reevaluate current portfolio of monitoring and analytical tools
- Leverage big data with advanced analytics
- Big data is an opportunity, not a problem
- Beware of the challenges with big data

Thank you.

Q&A

Contact :

rdevired@visa.com

twitter : ravred1