Capgemini
CONSULTING.TECHNOLOGY.OUTSOURCING

Pivotal™

Detecting Anomalous Behavior with the Business Data Lake

Introduction

Detecting Anomalous Behavior with the Business Data Lake

Our Anomalous Behavior Detection Solution addresses security issues that conventional methods can't. It can help to detect and prevent theft of data or Intellectual Property (IP), for instance at the behest of nation states or organized crime, or by a disenchanted employee. The solution can quickly identify when a user is behaving in a way that is abnormal for them and take appropriate action to limit what they can do, or flag up the situation for managerial attention. It can also predict when anomalous behavior is likely to occur, and identify when a user's account has been hacked.

The solution builds on the Business Data Lake (BDL), a joint development between Capgemini and Pivotal. The BDL has the benefit of being able to store vast amounts of information, both structured and unstructured, affordably, and present analytics outcomes to the business. We apply data science techniques to the information in the BDL to achieve the Anomalous Behavior Detection Solution.

We provide a comprehensive range of implementation services, from a quick proof of concept to a full-scale service implementation.

Use Cases

- Detecting anomalous log-on patterns
- Network intrusion detection
- Detecting abnormal finance activities
- Advanced penetration detection
- Protecting web-based retail business
- Preventing leakage of sensitive data

DETECTING ANOMALOUS LOG-ON PATTERNS

Big data solutions can help organizations detect and react to both “rogue” employees and external attacks to the network, so as to protect their critical data assets.

Large organizations often face the complexities raised by having many employees with a frequent need to travel – from sales and marketing staff meeting customers and partners through to production and R&D teams operating in multi-country delivery centers.

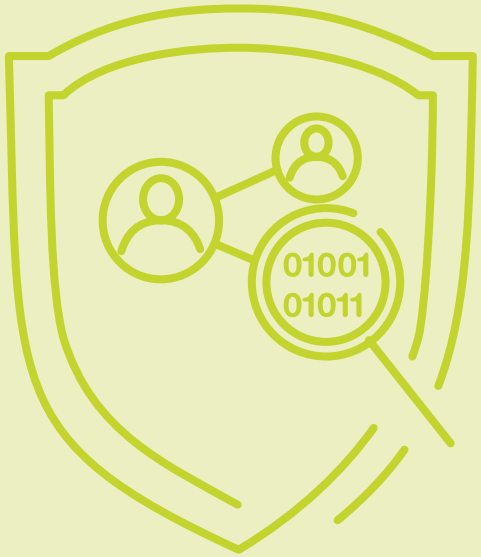
These employees often have a high level of access to company IP, including pricing and competitive insights. Because of their need to travel, they access information from a range of countries and systems – via laptops, mobile devices and even hotel computers – so that security is not always fully controlled by the company.

Existing security tools go part of the way towards defending the enterprise from traditional threats, but they are often unable to detect when an employee with the correct access rights is behaving abnormally, or appears to be doing so. For example, they might not be able to flag up that employee access credentials have been obtained via social engineering or targeted malware, and are being used by an attacker.

By using advanced data science based approaches, including machine learning and normalized behavior modeling, combined with a Business Data Lake, it becomes possible to create rich insights and take automated action to protect the enterprise from threats.

For example, it is possible to detect that an employee appears to have traveled an impossible distance between log-on attempts, or is suddenly accessing data sets they have never used before, albeit from the right location. If an anomaly like this is flagged up, you can take prompt action, perhaps suspending user access rights.

NETWORK INTRUSION DETECTION



Automated techniques can identify when malware is present on the corporate network and preempt some or all of the damage it could do.

Organizations often struggle to detect threats in large heterogeneous environments. Some of these are covert threats, i.e. they employ advanced techniques to bypass traditional security appliances and to be invisible to them. Malware can be targeted at one company, or even one employee at a company, to gain access to specific Intellectual Property or services within the business.

Absolute perimeter defenses are no longer possible in most organizations due to their complex global nature. This challenge is compounded by the fact that, to be competitive, companies increasingly need to work with wide ecosystems of partners and suppliers who have some level of network access. Instead of absolute defense, the main objective is to reduce the time to detection once the network has been compromised, and in some cases even to predict an attack.

By building behavioral intrusion detection frameworks based on machine learning, algorithms that analyze trends and connections, and with existing security data, it is possible to identify breaches and react fast enough to minimize corporate exposure.

In addition, predictive models can detect risks of advanced threats within the corporate environment and reduce malware's undetected "free time" on the network.

DETECTING ABNORMAL FINANCE ACTIVITIES

Conventional systems can tell when employees breach the rules associated with their roles, but not when they are acting in an abnormal way within the rules. Now it's possible to do both.

Finance controls present a challenge, as well as an opportunity, for many enterprises. Put too many controls in place and the process becomes unwieldy. Selected employees need to have some authority to execute finance decisions, such as authorizing of purchase orders. Most organizations will therefore sacrifice some control in order to delegate that authority and maintain operational speed.

However, sometimes the employees to whom authority is entrusted can be compromised – either via social engineering (such as manipulation or blackmail) or for personal gain. The enterprise needs ways of detecting and reacting quickly to abnormal approvals, even if the person in question is operating within their nominal ceiling of approval limits. However, it is challenging for conventional systems to identify that an employee is acting in accordance with their role and yet abnormally.

Through machine learning – defining the norm and mapping both historic and current usage from enterprise data feeds – anomalous behavior can be detected. The organization can then intervene or initiate further analysis.

This approach can achieve even more when additional data sets are used, for example to predict where loss might occur based on wider events and seemingly unrelated employee behaviors. You could use it to detect fraud in multiple centers by seemingly unconnected employee groups, or to take appropriate precautions when a cluster of employees have had poor appraisals.

ADVANCED PENETRATION DETECTION

The possibility of hackers getting hold of access credentials remains a significant threat to businesses, but new approaches to spotting abnormal behavior can guard against this form of attack.

Companies do their best to stop hackers, but at some point they will inevitably gain access to key user accounts as a "landing point" within the network. Once the landing point is compromised, they will use it to move laterally within the network and company systems. The intrusion and subsequent movement generates a lot of anomalous network traffic and systems access.

Once an external hacker gains access (using social engineering, targeted trojan or malware software) to any one area of an organization's systems, they can often use the credentials they have obtained to gain free reign over the whole network, gaining further access, changing access controls, creating back doors, and hiding their entry points.

Classic defenses do not spot that this has happened until it is too late, because the credentials used are valid. The best that companies have usually been able to do until now is get a retrospective view of what the hacker did based on logs and network data.

Now it is possible to tackle this problem proactively. Applying data science approaches to the business data lake can create a view of what is "normal" behavior for an individual or class of user. Then the organization can set up monitoring processes that continuously compare that view with current patterns of inter-device activity, account creation and changes, critical server activity, and privilege escalation.

This way, they can instantly detect abnormal actions that might indicate that a user account has been compromised. Detection can trigger automated actions ranging from informing line management and temporary user credential suspension through to mobilizing of a response team.

PROTECTING WEB-BASED RETAIL BUSINESS

Combining information about past Denial of Service attacks with wider network and social media data makes it possible to predict future attacks.

Retailers do an increasing proportion of their business via the web, selling to consumers via many electronic channels (web, mobile, etc). This means that they are highly dependent on platform reliability.

Denial of Service attacks on websites are increasing. These are large-scale, distributed attacks that usually involve bombarding the site with a high volume of requests that render it unusable. They can be launched by a range of attackers: organized criminal groups, “hacktivists” championing a cause, even nation states bent on causing disruption.

The impact on a business can be extreme, especially if the attack coincides with a peak seasonal event. Significant amounts of a company’s annual business are often executed in just a few hours, so they cannot afford to lose the website at that point.

Fortunately, it is now possible to identify patterns of external web traffic that may indicate when an attack is about to occur. This is done by combining information about known attacks with wider network and social media data. Supervised machine learning and data science techniques, building on the Business Data Lake, can help to analyze external traffic for evidence of activity by a known or new attacker.

If such evidence is found, security teams can take measures such as engaging their emergency response plans and activating upstream measures with their network operators, in order to protect revenue and maintain customer satisfaction. Attack information can also be shared quickly with partners and even rival companies to help maintain collaborative security information for the industry.

PREVENTING LEAKAGE OF SENSITIVE DATA

Sending email to the wrong recipient can have serious repercussions but the latest analytic techniques can warn companies before it happens.

When organizations communicate sensitive employee data (which might be HR- or finance-related), it is not uncommon for people to receive emails not intended for them. They may even go to someone outside the organization.

The use of private email accounts is ubiquitous and has further increased the risk because there are no easy ways to validate that a given email address is the correct one for the intended recipient. The danger is magnified when email addresses are “remembered” by an email client.

The impact ranges from a minor data breach to extensive data leakage that would enable an external party to start a social engineering attack on the business. Other dangers include reputational damage, financial loss by the employee, litigation and a loss of trust between employee and employer. Sending email to the wrong recipient can have serious repercussions but the latest analytic techniques can warn companies before it happens.

Now, however, by combining existing enterprise data (from finance, HR etc.), email information, and data science, it is possible to detect anomalous communications patterns. You can then temporarily hold a suspect email, while issuing a warning so that action can be taken to prevent data being sent to an inappropriate recipient.

About Capgemini

With almost 140,000 people in over 40 countries, Capgemini is one of the world's foremost providers of consulting, technology and outsourcing services. The Group reported 2013 global revenues of EUR 10.1 billion.

Together with its clients, Capgemini creates and delivers business and technology solutions that fit their needs and drive the results they want. A deeply multicultural organization, Capgemini has developed its own way of working, the Collaborative Business Experience™, and draws on Rightshore®, its worldwide delivery model.

Find out more at

www.capgemini.com/bdl
and **www.pivotal.io/big-data/businessdatalake**

or contact us at

bim@capgemini.com