patternex

Detecting Data Exfiltration

A PatternEx Solution Guide

Problem

Data exfiltration has traditionally been one of the most difficult problems facing InfoSec professionals. The methods vary widely, from posting stolen data embedded in images to social media platforms, to an extremely slow trickling of data to an anonymous cloud storage providers. The techniques used are often tailored specifically to the particular victim, making them even more problematic. But once in, all attackers share one common goal: to control your data. Be it customer data, employee data, credit cards, intellectual property, or other critical information, data control is their goal.

Today's solutions focus on protecting this data by tagging critical data or files, then watching it at rest or as it moves over the network. Unfortunately it is often hard to distinguish between good and bad data, as well as what is legitimate use versus what is not. Furthermore, there are many ways to elude tagging technologies. This is reflected by the fact that 1 in 5 Americans have had their personal data compromised at least once in the last 3 years, while 90% of all companies state they have been a victim of a breach last year alone.

Today's solutions focus on protecting this data by tagging critical data or files, then watching it at rest or as it moves over the network.

These attacks can sometimes be state-sponsored, seeking competitive information or company or state secrets. But a large driving factor today has been the rise of criminal marketplaces, where the buying and selling of information is big business. From well-funded criminal organizations to lone wolves to foreign governments, if you have data of value, sooner or later someone will try and take it.

Current Solutions Are Fundamentally Lacking In Key Areas

Currently, there are two kinds of InfoSec technology solutions: those that are 'analyst-driven' and those that are anomaly detection-driven. The former are legacy products with significant limitations while the latter are an interesting emerging set that merit consideration and evaluation.

Approach	UnSupervised Learning (No Labels)	Supervised Learning (Static labels)	Active Learning (Dynamic labels)
Real-time (learning continuously)	Anomaly Detection for UBA		(AI)^2: Active Learning + Analyst Intuition
Batch (Learn offline)	Anomaly Detectors	Offline Fraud models	

Analyst-Driven Solutions:

Analyst-driven solutions include such products as network firewalls, Fraud Detection/ Prevention systems, SIEM's, and Intrusion Detection and Prevention Systems (IDPS). They're based on allowing analysts to create rules that allow or deny access to a user. And while these are a necessary component to your security posture, they are not sufficient.

Rules-based approaches have a high rate of undetected attacks. A rule, by definition, is something that reacts to a new attack; it prevents historical attacks from happening again. Rules-based solutions can't "see"—much less prevent—new and emerging attacks.

A second problem is the delay inherent to analyst-driven solutions. Once the attack is discovered, a human must create, test, and validate the rules that prevent the attack from happening again—in the meantime your company remains exposed.

Finally, rules are detectable and solvable. While rules remain static, cyber-criminals are constantly adapting their behavior, testing the rule until they find a way to circumvent it, starting the detection/delay/resolution process all over.

Anomaly-Driven Solutions

The new class of emerging solutions based on machine learning hold promise, but fall short in critical areas. These "anomaly detection" products discard static rules in favor of statistical models that determine when certain events are outliers. While these solutions tend to do better in the detection of new attacks, they suffer from high levels of false positives while still containing some important vulnerabilities.

The major failing of these Anomaly Detection products is not in their capability but in their approach. While they can execute complex statistical analyses to identify anomalous events, the results they generate lack context. Not all outliers are malicious. Generating alerts tied to anomalies means you generate many false positives.

The impact of too many false positives is severe: analysts become fatigued from seeing too many alerts; they then begin to distrust the system that generates false alarms, and eventually abandon the system.

Time to solution is also an issue. Anomaly detection (AD) products require a “baseline” of behavior upon which they can identify outliers. To do so, AD solutions require 90 days of data to form the baseline. Malicious behaviors in the baseline data may not appear as outliers to the ML solution, leading to blind spots in your defenses.

> Machines can process massive amounts of data and extract complex patterns, but without a human to guide them, they generate false alarms.

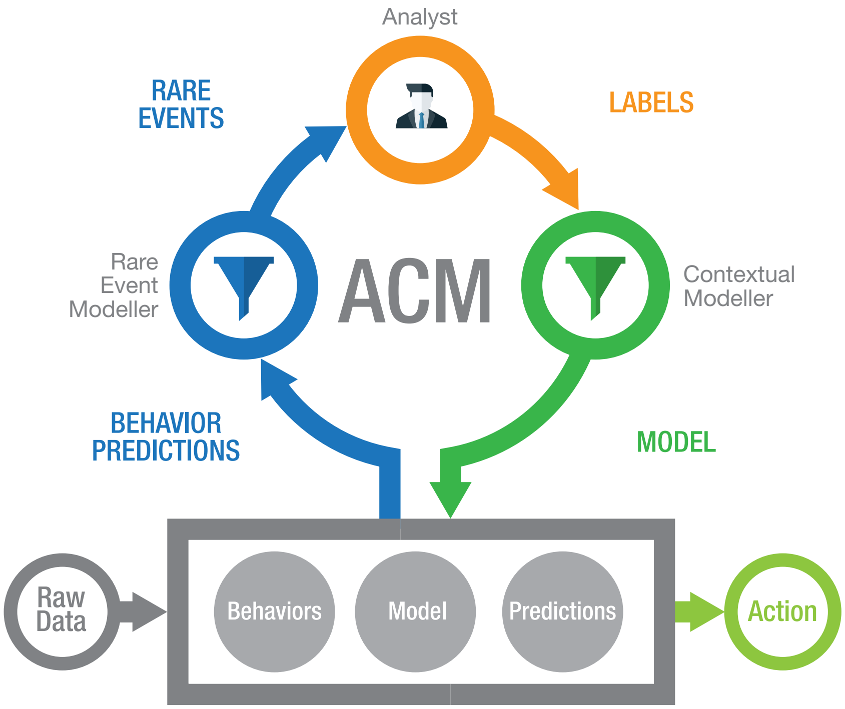
A Change In Perspective—And Approach

What is needed is a fundamental change in thinking and approach. Where analyst-driven or anomaly detection systems have important shortcomings, PatternEx combines the two into a new Artificial Intelligence concept called (AI)^{2}.™

The heart of this concept is the notion that humans and computers each have different—and complementary— strengths. Machines can process massive amounts of data and extract complex patterns, but without a human to guide them, they generate false alarms. Humans excel at things like context, nuance and understanding the implications of what the machine finds. With a human training a machine, the machine will eventually be able to mimic the human’s intuition to make value judgments about the significance of complex patterns at tremendous scale.

Patternex Threat Prediction Platform

The PatternEx Threat Prediction Platform combines big data infrastructure with human security expertise and artificial algorithms into a proprietary solution that can predict, in real-time, when you are being defrauded through business logic loopholes. The result is the industry's first Threat Prediction Platform.



Active Contextual Modeling

If the vision is (AI)^2, then Active Contextual Modeling TM (ACM) is that vision in code. ACM is the patent-pending system that takes raw data and ultimately delivers highly actionable alerts for your Security teams.

How it works: The first step in the ACM process is to ingest all the relevant log data from your IT infrastructure. The ACM then transforms those logs into quantifiable identities and behaviors. From there, the data are run through a combination of Artificial Intelligence models to produce a highly focused list of alerts for the analyst. The number of alerts addressed is determined by your analyst team based on their bandwidth and perceived importance of the alerts. The analysts review the alerts, then use the ACM-generated data to assess the alerts and label them as an attack or not. That analyst feedback updates the Artificial Intelligence models to be even more precise. The AI Platform then can instruct your system to challenge, delay, or block the malicious user automatically via an API connection.

Benefits

By augmenting your InfoSec team with the PatternEx Threat Prediction Platform, you can:

- Gain visibility into previously invisible cyber attacks
- Receive real-time alerts enabling you to take immediate preventative action
- Reduce the risk of brand damage
- Reduce the risk of data exfiltration

Conclusion

To find and beat Data Exfiltration schemes, you need to be able to recognize abnormal behaviors exhibited around your avenues of data removal, predict user intent, and stop the data from leaving your network before it happens.

Today you assume an attacker will get in, and their goal is to first control your data, and then to move it outside your network. By looking at the behaviors of the data in motion, endpoints, and users, PatternEx is able to not only find these abnormal behaviors associated with data exfiltration techniques, but can actually predict attacks before they occur.

The enabling technology here is Active Contextual Modeling, a patent-pending technology that learns from your team what is important to you, and what you consider a risk. As the system is trained by your team, it becomes a virtual analyst for you, analyzing behavior patterns and assessing the implications of those patterns as massive scale.

Using this hybrid approach enables the PatternEx platform to pinpoint data exfiltration schemes and help stop them before any damage is done.

Deployment Options

We offer the PatternEx solution as a software package, available as a service, in your private cloud, or on premise. Your requirements are basic: stream log data to into PatternEx, review the real-time user behaviors, and provide feedback.

How To Contact Us

For more information or to request a demo, send your email to sales@patternex.com or info@patternex.com.

PatternEx is an Artificial Intelligence security company that has developed a Threat Prediction Platform to identify malicious user intent. The PatternEx solution enables security analysts to detect and prevent cyber attacks in real time, at scale.

Trademarked names appear throughout this paper. Rather than use a trademark symbol with every occurrence of a trademarked name, names are used in an editorial fashion, with no intention of infringement of the respective owner's trademark.

The information in this paper is distributed on an "as is" basis, without warranty. Although every precaution has been taken in the preparation of this work, neither the author nor the company shall have any liability to any person or entity with respect to any loss or damage caused or alleged to be caused directly or indirectly by the information contained in this paper.

patternex

PatternEx
4620 Fortran Dr., Suite 202
San Jose, CA 95134

www.patternex.com