Anomaly Detection vs. Pattern Recognition

By Dennis Zimmer
CEO opvizor GmbH,
VMware vExpert, VCP, VMware VCAP
**Table of Contents**

1. Introduction ................................................................................................................................. 3  
   1.1 Virtual infrastructures are becoming increasingly complex ............................................. 3  
   1.2 A wide range of virtualization solutions and infrastructure components ........ 3  
   1.3 Predictive issue detection ...................................................................................................... 3  
2. Anomaly detection .......................................................................................................................... 4  
   2.1 Anomaly detection in VMware environments ...................................................................... 4  
3. Pattern recognition ......................................................................................................................... 5  
   3.1 Crowd-sourced pattern recognition ..................................................................................... 6  
3. Mixing both worlds ....................................................................................................................... 6  

DISCLAIMER .................................................................................................................................. 8
1. Introduction

1.1 Virtual infrastructures are becoming increasingly complex
Virtualization is an integral part of a modern data center. A few hosts can now manage what previously required a large number of servers. In fact, frequently, the degree of virtualization is 90 percent or more. With the high rate of virtualization and the associated increase in complexity, resulting complications can often be difficult to identify. What’s more, under certain circumstances, even minor problems can cause a significant negative impact on the entire infrastructure. Therefore, it becomes important to effectively surveil the infrastructure to both ensure smooth operations and avoid costly errors.

1.2 A wide range of virtualization solutions and infrastructure components
Virtualization solutions are many: the selection ranges from suppliers such as KVM and Citrix to Microsoft Hyper-V and the market-leading provider VMware, with its vSphere solution. In terms of basic functionality, each of these solutions is quite similar. They mainly enable resource partitioning for optimal and cost-effective use of physical hardware. In addition, they often enable new methods of high-availability designs. When these solutions are paired with other components of the infrastructure, the combinations are limitless.

1.3 Predictive issue detection
Every IT department running a virtual environment struggles with the complexity of the configuration and the constant goal to prevent downtime and other negative effects. Unfortunately, the most common and yet difficult way to manage the infrastructure is a reactive one – waiting for something to happen before the culprit can be found. That means the damage already exists, but the question is how fast the system is up and running at full speed again.

A more proactive way to manage the infrastructure, including investigating log files, checking for configurations, events and performance trends would necessitate a tremendous amount of work for the typical size of virtual systems, requiring the hiring of additional resources. This is because virtualization adds the network, storage and applications layers to the system administrator’s working desk.

As a result, software vendors attempt to offer predictive handling of the environment via software. This has been accomplished in the security space, known as intrusion detection and anomaly detection. Unfortunately, there is no self-learning super-software that takes care of all the predictive work of the IT environment. Fortunately, however, opvizor offers a real, practical solution, one that favors the system administrator and people involved. Opvizor builds an entire platform on the knowledge of people, instead of the scientific way of predicting issues that doesn’t account for human interaction.
2. Anomaly detection

Typical anomaly detection products have existed in the security space for a long time. Most of them deal with intrusion detection and try to locate uncommon network traffic. To better understand what uncommon means, you need to understand that these products run in silos. So every customer needs to have a separated environment (the smaller the better) to run the anomaly detection service to create a common baseline. This baseline consists of hundreds or thousands of entries of what should be normal behavior. Whenever a new service comes into the network, the intrusion detection sends an alert. If this alert is critical, the system administrator needs to react; if the alert just informs about a correct situation that was unknown before, it’s a false positive and the system administrator needs to train the intrusion detection system not to raise that alert again. As most administrators will likely get thousands of false positives, there is a high likelihood that administrators will start ignoring all alerts of the anomaly detection system, rendering it useless.

All the information so far assumes there is a symptom that the detection software can jump on and create an alert for. Unfortunately, most misconfigurations have no symptoms until it’s too late and a system goes down, a service fails or an attacker attacks. Therefore it’s impossible for an anomaly detection system to detect misconfigurations and misaligned best practice settings that lead to the majority of support cases.

Anomaly detection has some amazing benefits:

- Finds uncommon behavior in the environment potentially uncovering bad things
- Acts dynamic and can be trained
- Is not limited to a specific set of rules

Anomaly detection also has significant disadvantages:

- Can only find symptoms, not the root cause
- Cannot locate issues that don’t have symptoms, which is most
- Baseline is only as good as the setup

2.1 Anomaly detection in VMware environments

You may think – thanks for the refresh about intrusion detection systems that most of us have already abandoned, but what about virtual environments? Actually there are also products for virtual environments, especially VMware, such as VMware’s own flagship product vCenter Operations.

When speaking to customers, basically the same situation is true as it was mentioned for intrusion detection systems. As you typically don’t start using anomaly detection solutions with a clean and perfect environment, it’s necessary to train the system every single day to get the most out of it. Of course the system is self-learning, but it learns bad things as well as good things. Most of the time there
are more anomalies a day that are good (planned upgrade, new VMs, Host systems replaced) than bad (error log entries, strange performance behavior). Therefore as a system administrator you need to handle all the false positives. We’ve seen situations where customers replaced their ESX Hosts and based on the trends, alerts were sent out stating that no ESX Hosts remained, or virtual machines that are intentionally powered off but powered back on every month for 2 days also generated an email alert. If an anomaly detected is really a critical situation, you need to find out what the root cause is, which requires a good deal of product and operational knowledge. Last, but not least, the first real result out of good anomaly detection software doesn’t arrive until after a week or two.

Now imagine a bad configuration has no symptoms, i.e. a virtual machine is configured with 16 GB of memory, despite its 12 GB limitation. The symptom happens when ballooning comes into play because the application wants to access 16 GB but is blocked by the VMkernel at 12 GB. First of all, when the symptom is there, you need the knowledge to find this misconfiguration quickly, as the application already has problems. Finding this change is hard and time consuming. Pattern recognition is needed in that case.

Anomaly detection can be great, but not if the system isn’t taking care of continually. Depending of the size of the infrastructure, a minimum of 1-3 people are needed just to manage anomaly detection and troubleshooting.

3. Pattern recognition

Pattern recognition is the old way of detecting issues and doesn’t look very modern when compared to anomaly detection. It’s not really cool from a marketing perspective.

Pattern recognition has some amazing benefits:

- No false positives
- Instant results (time to value)
- No manual root cause analysis needed as root cause is already known most of the time due to the pattern
- No training necessary

Pattern recognition also has its disadvantages:

- Doesn’t find anomalies in a dynamic way
- Only finds exact patterns (no dynamic behavior)
- Needs high expertise when defining patterns

To get a better idea about a pattern to be recognized, think about a snapshot that is older than 1 month, growing faster than 10% a day or a CPU limitation set lower than the configuration. Patterns can be a lot more complex and involve metrics such as configuration, performance, log data, events. Furthermore, it can combine data from other applications, hardware components and more.
When I started thinking about opvizor, I asked myself whether we should go for anomaly detection, pattern recognition, or both.

I decided on something more advanced: *crowd-sourced pattern recognition*.

### 3.1 Crowd-sourced pattern recognition

As pattern recognition platforms are erected on the number of patterns that can be checked, it’s hard to come up with hundreds of patterns with just 5 people in a room brainstorming. The issues also need to come with information about the thread and how to solve things – which is one reason why most of the current existing analysis and monitoring tools consist of 100 rules or less. opvizor consists of more than 730 rules and is continuously growing.

That’s why opvizor is setup to be intelligence-centralized and community-driven. That means our customers and system integrators, consultants and other community peers send us information and support requests to be embedded in our rulebase. And that’s our job - adding rules and enriching the descriptions to explain the issue and give hints to solve it. This allows us to always be up to date and provide the most advanced pattern recognition engine to find misconfigurations, security leaks and performance bottlenecks in our customer’s environment.

### 3. Mixing both worlds

Why mix both worlds?

As I mentioned before, both ways to recognize issues come with benefits and disadvantages. We believe that anomaly detection is the more convenient way to think about environment optimization and health checking, but it’s definitely not the best approach. As anomalies arise mainly from misconfigurations, you need to find the misconfigurations first to avoid issues. Anomalies are the symptoms of a misconfiguration.

If you don’t want to renounce on anomaly detection you must consider that only a healthy baseline is ready to properly be surveilled by anomaly detection.
If the misconfiguration is fixed, the anomaly will be eliminated. On the other hand, it’s impossible to have all issues that can potentially happen covered as a pattern, no matter how many people contribute knowledge.

Therefore, we typically recommend also using anomaly detection, providing a better chance of finding strange symptoms in an environment as a predictive engine before your real-time monitoring tool jumps on an exceeded threshold or a system downtime.
DISCLAIMER

Copyright 2014 opvizor Inc., all rights reserved

The content and the information in this document are protected by copyright. This emphasis, processing, distribution or duplication (copying by any means) of this work or portions thereof, are not permitted without the consent of the publisher. The information in this document is provided together with the VMware opvizor analysis software.

This document is for informational purposes only. opvizor Inc. assumes no liability for the accuracy or completeness of the information.

To the extent permitted by applicable law, opvizor Inc. provides this document as is without warranty of any kind, including in particular the implied warranties of merchantability, fitness for a particular purpose and non-infringement. In no event shall opvizor Inc. shall be liable for any loss or direct or indirect damages arising from the use of this document, including, without limitation, lost profits, business interruption, loss of goodwill or lost data, even if opvizor Inc. has been advised of the possibility of such damages.

The opvizor Inc. reserves the right to make changes and improvements to the product in the course of product development.

opvizor, Inc.
4141 Southwest Freeway, Suite 265, 77027 TX, Houston
www.opvizor.com
CEO : Dennis Zimmer

Date: June 11, 2014