

On the Benefits of Vulnerability Data Consolidation in Application Security

Santanam Kasturi¹, Xiaolong Li², Peng Li³ and John Pickard³

¹Technology Management, Indiana State University, Terre Haute, USA

²Dept. of Electronics and Computer Engineering, Indiana State University, Terre Haute, USA

³Dept. of Technology Systems, East Carolina University, Greenville, USA

skasturi@sycamores.indstate.edu

xiaolong.li@indstate.edu

lipeng@ecu.edu

pickardj@ecu.edu

Abstract: This research aims to build upon a conceptual idea of consolidating all application security vulnerability data from monitoring, detection, and discovery tools into a physical system that allows for convergence of observation and response to an event that is a threat. Multiple application security testing and monitoring tools are deployed at different layers of an application architecture and capture activities that occur at that layer. This multi-layer data capture is disconnected without any analysis of data lineage from the externally exposed web attack surface to deep down into the application and data layers. It is only through this data consolidation can one provide a reliable statistical analysis of correlating multiple vulnerability information and synthesize an attack pattern and predict possible events accurately. The benefits of such a system are discussed in this paper that includes how one can organize the data, identifying temporal and spatial correlation of events, focusing on specific web requests that point to a specific vulnerability, and formulating a fast response to such events. Advantages of integrating with Security Information and Event Management (SIEM), Security Orchestration, Automation, and Response (SOAR/XSOAR), Extended Detection Response (XDR) are briefly discussed. The analysis can be further used to develop a predictive system using deep learning (DL) techniques using correlation of application security vulnerability information.

Keywords: Data Consolidation, Attack Surface, SIEM, XSOAR, SOAR, XDR, Vulnerability Correlation, Deep Learning

1. Introduction

According to a survey, 77% of CISO's report that prioritization of vulnerability management remains a challenge as risk assessment is slow and lacks enough information, while another 88% say that combining application runtime context, vulnerability analysis, and risk impact assessments would make the job easier (Dynatrace, 2023). A typical 'breakout time' - the time an attacker takes to move laterally from the first penetration point in a compromised host to another host within the system - declined from 98 minutes in 2021 to 84 minutes as measured in 2023 (CrowdStrike, 2023). If one does not respond with an action to stop the spread to minimize the costs of damages caused by hackers, within this breakout time of 84 minutes, the loss could be catastrophic especially in the case of a ransomware attack. It is recommended that a 1 - 10 - 60 rule be established where the first minute should be in detecting the threat, the next 10 minutes be spent in understanding the threat, and the final 60 minutes in responding to the threat with appropriate action to minimize losses (CrowdStrike, 2023). Even more critical is a zero-day vulnerability commonly found in open-source software libraries and components which is a software code available to all to use and share updates and helps developers to use code developed by others for certain features that are readily available and can be used with little or no modifications. These components are again small functions or features that can be used among multiple applications within an organization. Chances are that everyone using a version of a component will introduce the same vulnerability across the enterprise, as was seen with the Log4Shell, a zero-day vulnerability that was discovered in 2021, and one that had the potential for an exploit that had no grace period to remediate, and the exposure was instantly open to attackers, leaving 'zero-days' to fix the vulnerability.

2. Challenges with Application Security

Multiple challenges exist and three are discussed here. By multiple challenges we mean there are operational challenges in integrating a host of tool technologies, managing information gathered from each, maintaining data integrity and quality, producing metrics and key performance indicators, and presenting visualizations for leadership are many and are not discussed here. However, it is worth mentioning that there are challenges beyond the four described below.

2.1 Application Security Testing in a Siloed Manner

Due to the nature of the applications that are being developed and approaches to different test and scan types that include Static Analysis Security Testing (SAST), Software Composition Analysis (SCA), Dynamic Analysis Security testing (DAST), Application Ethical Hack (AEH) /Pen test, Application Programming Interface (API) testing, Container testing, and even Infrastructure as Code scanning (IaC), in most large organizations data consolidation is not integrated to study correlation of vulnerabilities end-to-end and analyze if there is a pattern to attacks happening from outside. One solution proposed discusses elimination of false positives by running correlations from vulnerabilities found from different test and scan types within an application, all housed within one tool (Checkmarx, 2022). This however does not resolve the challenge of external attacks that target a specific vulnerability within an application or even a collection of applications that serve a single business domain that have common nodes of transactions. We must also consider that some organizations employ external agencies to scan their exposed web attack surface to discover vulnerabilities that are not detected by any of the test types. This can be due to a new vulnerability that has been added by National Vulnerability Database (NVD), or due to missed coverage in case of risk-based testing. So, other capabilities that discover vulnerabilities that complement those found using the standard test and scan methods employed within, must be included to cover the entire attack surface. Unless one brings in all kinds of application security vulnerabilities detected by internal tests, external agency discoveries, and external monitoring by Web Application Firewall (WAF) to consolidate all vulnerability data it will be difficult to visualize or synthesize an attack pattern (Checkmarx, 2022; Akamai, 2023; Carielli et al, 2022).

2.2 Security Testing Tool Sprawl

As more and more organizations adopt tools and automation, the abundance of tools in the market can lead to multiple security testing solutions that have over the years become outdated with the ever-growing complexity of applications. Added to this challenge is the amount of data that comes out of all this testing and the analytics still relying heavily on spreadsheet-based skills and reporting. The tool sprawl, semi-automated security testing and reporting practices makes analysis of vulnerability information difficult, slow, and is a challenge to proactively respond to security incidents. With the advent of AI and ML one can hope for an autonomous analytics solution, however the challenge here is data quality. Studies show that 40% of challenges come from tool sprawl, and correlating alerts from different tools is largely manual and time consuming, resulting in too many false positives. These studies also show that 62% of organizations use four or more solutions to maintain the security of their applications (Dynatrace, 2023). Deploying the right monitoring and detecting tool becomes key to correlate attack requests with existing vulnerability information as shown in Table 1. The details of each type of analysis and the tools associated with it can be found in literature referred here (Carielli, et al, 2021; Checkmarx, 2022; Veracode, 2023; Primeon, 2018; Akamai, 2023; Signal Sciences, 2021; FASTLY, 2022; Na, 2021; SALT, 2023)

Table 1: Test / Scan Data

Capability	Deployment Layer	Vulnerability Information	Alerts
SAST	Code	CWEs, Severity, Exploitability	Point in time scan reports
SCA	Component	CVEs, Severity, Exploitability	Point in time scan reports
API SECURITY	API	CVE's, CWE's, Severity, Exploitability	Point in time reports
DAST	URL/UI	CWEs, Severity, Exploitability	Point in time scan reports
PENETRATION TEST	URL/UI	CWEs, Severity, Exploitability	Point in time test reports
WAF	Network Layer	Blocked Attack requests, Incoming threats, target IP and host information	Continuous monitoring
Hybrid WAF/ RASP	Application Layer	Blocked Attack requests, Incoming threats, target IP and host information	Continuous monitoring

2.3 Data Storage

A final challenge is storing such a vast amount of information from multiple sources of data with relational capabilities enabling rapid analysis and visualization. Data consolidation from multiple sources that reveal vulnerabilities at different layers of an application becomes a critical requirement to begin with. Without it, to think about identifying an attack pattern, synthesizing an attack, or even predicting and preventing an attack will be fraught with risks of too many false positives.

3. Data Consolidation Benefits

In view of the three challenges discussed above - application testing happening in a siloed manner, tool sprawl, and data storage - the need for an integrated platform that consolidates data from monitoring and detecting tools and efficiently manages and provides analytical insights is much desired.

The first step towards a predictive analytics system is to build a platform for data consolidation from all the different monitoring and detecting methods employed for application security testing. The final solution is to design an integrated analytics and predictive system using autonomous processing techniques, like a Deep Learning (DL) technique such as an artificial neural network (ANN) that will predict and preempt an attack and is beyond the scope of this work and requires a separate treatment. Creating a data warehouse that can support data storage from multiple sources and ensure providing data marts for individual lines of business use is essential. The information from all the test, scan, and monitoring results provides insight of vulnerability and attack information as various attributes as seen in Table 1. To understand and correlate the data, one must look at a fast and easy way to store and retrieve information that can easily be fed to a visualization tool like Power BI or Tableau. The same information also needs to be fed to a predictive system that runs an algorithm for DL and assumes reliability and accuracy. The entire flow of data and the system is shown in Figure 1.

Combining all types of vulnerability information in a single data warehouse and transforming the data into visualization models to be presented using a BI Analytics solution has specific advantages. The first advantage is the ability to correlate vulnerabilities discovered in multiple layers through multiple monitoring and detection capabilities. Presenting vulnerability information discovered in multiple layers provides a view to the association of vulnerabilities to the transaction path.

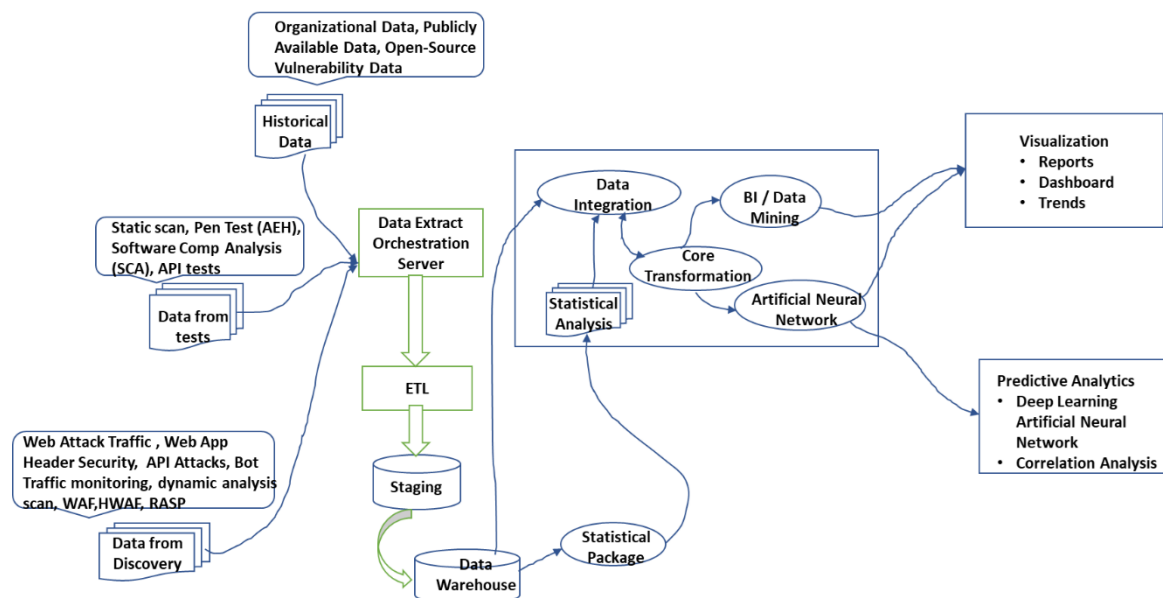


Figure 1: Conceptual Integrated Data Analytics System

It still does not give you a full picture of the threats facing your attack surface. It still does not tell you the present and the future; all it tells you is the past. While data consolidation will help in understanding the vulnerability lineage, it still will not help you in responding with the speed needed in the event of an attack. It will require a continuous monitoring system, and combined with all the data gathered will provide a better picture of what may happen. This continuous monitoring tool can be a WAF, or a Hybrid WAF, or even better a RASP tool (Akamai, 2023), (Carielli et al, 2022), (Signal Sciences, 2021), (FASTLY, 2022), (Na, 2021). In our research we have

used a WAF. Further, this can be integrated with an incident response system like SIEM/XSOAR/XDR that are primarily designed for networks, endpoints, and sensors, and not ideally for applications.

3.1 Correlation Analysis

Here is why data consolidation is key because processing all such correlations, which may be hundreds, is humanly impossible and to provide immediate response is not easily realizable, and not without a lot of manual analysis. Here is also why AI and DL are necessary to effect autonomous processing and replace manual processing and make decision support systems meaningful. Analyzing requests monitored using WAF and vulnerability data from SAST, SCA, DAST scans combined with a penetration test, one finds correlations between types of attack requests and corresponding vulnerabilities found in the application (Kasturi, et al, 2023). It is not possible, when dealing with dozens of applications with transactions running into several thousands in each, to manually identify what is a normal transaction and what is not. A funnel down approach is needed which becomes part of an extended analysis. Once specific request types that are considered as valid and applications with corresponding vulnerabilities are identified, analysis can be done at an application-level correlating with specific vulnerabilities to study if they are indeed valid requests or missed attacks. This requires a deeper analysis of the data and is part of continuing research to provide key metrics and a possible lead up to synthesizing an attack pattern by studying inter-application correlations. This analysis becomes a key piece in the integration with incident management and response systems as discussed in the following sections.

3.2 Integrating SIEM/XSOAR/XDR

When it comes to security and compliance, organizations must keep track of a vast array of data sources and threats. This is where SOAR (Security Orchestration, Automation, and Response), XSOAR (Extended Cybersecurity Orchestration, Automation, and Response), SIEM (Security Information and Event Management), and XDR (Extended Detection and Response) come into play. SIEM is a technology that is used to collect, store, and analyze security data from various sources. The data is then used to detect and respond to security threats. SIEM technology can be used with other security solutions, such as firewalls, antivirus software, and intrusion detection systems for detecting and alerting anomalies. SOAR on the other hand is a system for automating, orchestrating, and responding to security threats after collecting data from same sources that the SIEM gathers. These responses can be customized to specific threats. If the data is not primed for an immediate response, it can be used for further analysis to make a more accurate response. XSOAR is an extended improvement over a SOAR in that it incorporates additional features like machine learning (ML), threat intelligence (TI), and security analytics for improved automation and orchestration capabilities, with a dashboard for unified security information for better threat response. XDR combines multiple detection and response techniques like SIEM and XSOAR into a single platform, combining data from different sources (network traffic, endpoint logs, and threat intelligence) to provide a more comprehensive picture of security events. Security teams gain speed through this technology integration in responding to threats by getting a full view of the security posture. XDR is not a replacement to SIEM or XSOAR but acts as a complement to them.

A survey of application of SIEM/XSOAR/XDR technologies in a variety of industries is discussed briefly. The literature covered starts with challenges identified (Cinque, et al, 2018) to the latest review on the state of SIEM (Velasquez et al, 2023). The industries and domains covered are Air Traffic Control (ATC) (Cinque, et al, 2018), Intrusion Detection Systems (IDS) (Muhammed, et al, 2023), Industrial Control Systems (ICS) and critical infrastructures (Mern, et al, 2022; Gonzalez-Granadillo, Gonzalez-Zarsosa and Diaz, 2021), and Wind Energy Systems (Johnson, et al, 2023). This gives a good sweep of systems that are not just IT, but also includes critical infrastructures outside of IT. In addition to SIEM, the literature also includes scope of XSOAR, EDR, and XDR. Tables 2 - 5 below gives a comprehensive summary of SIEM, SOAR/XSOAR, EDR, and XDR (Nour, Pourzandi and Debbabi, 2023; Olteanu, 2022).

Table 2: Security Information and Event Management (SIEM)

Characteristics	Advantages	Challenges	Deployment
Collect security event logs and telemetry data in real time for threat detection and compliance use cases, analyze data in real time, analyze incidents and impact, report, store logs and relevant information.	Irreplaceable, provides a holistic view, establishes a threshold for critical control, provides a data lineage to trace back to the initial attack and ensures	Vulnerable to attacker countermeasures, expensive to deploy and maintain, correlation to attack source and target is a challenge, can generate a lot of alerts and false	Network, devices, sensors, and all infrastructure components leading to event collection; event normalization; set action rules for protect, remove, and respond; event storage and monitoring.

	hardening of the security posture and provides a sure means for real-time analysis. Supports audit and mandatory regulatory requirements in managing threats.	positives, requires skilled analysts, attack variants can pose problems, a new and evolving market with too many players, scalability is a big factor and has a multiplier effect and with an ever-growing number of connected devices and assets leading to alert fatigue with little context to the security operations team to understand and implement a response.	
--	---------------------------------------------------------------------------------------------------------------------------------------------------------------	-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--

Table 3: Endpoint Detection and Response (EDR)

Characteristics	Advantages	Challenges	Deployment
Endpoint monitoring and collecting potential threat activity, analyzing data, and identifying malicious patterns, provide automatic response with action to stop or remove threats with alerts.	EDR's can provide critical context to detect advanced threats, can run automated response activity to isolating an endpoint from the network in real-time, to stop and prevent further spreading of the issue.	Standardization of unstructured data; setting correlation rules are business specific and depends on analyst knowledge; behavior analysis is based on knowledge of system, environment, and a-priori knowledge; requires extensive ground truth data to create behavior rules.	Resources located on the endpoint like collected events, logs, and binaries, are correlated, and analyzed to determine whether a suspicious activity occurs on the host, are signature-based solutions for pattern recognition. Does not cover networks.

Table 4: Extended / Security Orchestration, Automation and Response (X/SOAR)

Characteristics	Advantages	Challenges	Deployment
X / SOAR systems (Extended / Security Orchestration, Automation, and response can be used to collect data on information security events from multiple monitoring targets, process them, and configure an automated response using typical response scenarios, can be run as a playbook.	Plays together with SIEM, uses threat and telemetry data across a wide range of target monitoring systems to identify events and alert threats. Automating routine and repeatable incident response tasks and workflows	Playbooks can become repetitive and if not reviewed can become obsolete. Although they save a lot of time in response actions, they are created by humans from experience and a-priori knowledge of systems. As systems and attacks change, playbooks need to be revisited.	Host based intrusion detection systems (HIDS) and Network based intrusion detection system (NIDS) are used to create playbooks for responses based on what is detected. Streamlines incident response through an interface.

Table 5: Extended Detection and Response (XDR)

Characteristics	Advantages	Challenges	Deployment
Extended Detection and Response is a technology for threat detection and response, it unifies security products for detection and response and threat intelligence into a single platform.	Uses machine learning, correlation, and analytics capabilities to enhance the response time and the efficiency of the security teams. While SIEM creates alerts, XDR does a deeper analysis using AI/ML for a better representation of the threat.	It is a new concept; the technology is evolving and has not yet matured.	Covers endpoints, servers, emails, cloud, and networks.

3.3 Monitoring: Application Security vs Infrastructure Monitoring

In recent times, web applications have become the primary targets of attackers with their widespread use that even attackers identified as script kiddies with little knowledge can perform very sophisticated and damaging

web attacks using ready-to-use attack tools. In addition, according to the analysis performed on all vulnerabilities, the rate of critical vulnerabilities in the web application layer is approximately four times that of the network layer (Edgescan, 2019). Yet, most literature today reports studies on network-based intrusion detection systems (NIDS) than studies on web-based attack (Sevri and Karacan, 2019).

Applications have “normal” attributes and behavior. This is the “descriptive” portion of the evolutionary state of analysis (Descriptive, Diagnostic, Predictive, Prescriptive, and Cognitive). Attributes are scripts, languages, interpreters, compilations, configurations, security artifacts, connections (to other apps or databases), and infrastructure topology. Behavior points to availability, stability, response time, throughput/volumes, utilization (hard disk, memory, CPU), load conditioned activities (batch vs. pseudo-batch), workflow, error rates, and error response activities (or lack thereof) and indicates a state of normal behavior or a state of a compromised system. Basically, the artifact that defines a workflow profile would be considered an attribute, while the actual processing statistics of a transaction through the workflow would describe a set of behaviors (Kasturi, 2020). Monitoring can be at the infrastructure or at the application level. While network and device monitoring are in the domain of infrastructure monitoring, network traffic associated with a specific application is in the domain of application monitoring. Similarly, CPU utilization or heap size measure is in the domain of infrastructure monitoring but associating that to a business transaction from an application is in the domain of application monitoring. The challenge, however, is that monitoring tools do not provide the correlation between an infrastructure analysis system behavior to an analysis of application behavior. The reason is applications, and their transactions do not reside or are contained in one system in an enterprise that has distributed application architecture. Applications are multi-tiered and are spread across vast global geographies. An application is also not a network, device, or a host, it is a piece of code that is not a physical entity for an agent to monitor and report to a SIEM, and when we have hundreds of distributed applications and many talking to each other, and all having hundreds of weaknesses it can be visualized as a spaghetti code with trapped vulnerabilities. There is a disconnect between a security operations infrastructure and application behavior, and correlating both needs to be addressed to bring in application monitoring within the ambit of security operations and threat response using SOAR or XSOAR, or even XDR.

3.4 Integrating Application Security Vulnerability Information Data Platform with SIEM/XSOAR/XDR

Incident response using SIEM/XSOAR/XDR are all reactive systems, these come into play after the fact that a threat has been detected. This is where threat hunting combined with threat detection complemented with historical data on existing application vulnerabilities and technical debt plays a pivotal role in making a comprehensive system. The proposed data consolidation platform is the first step towards that as it enables creating a data warehouse producing datamarts, dashboards, and a decision support system. The data, however, becomes a static representation of vulnerabilities prevailing in the application code, and this must be used with a correlation analysis (Kasturi et al, 2023). The end goal is to integrate this system into an incident response system as shown in Figure 2. The correlation analysis data can be fed into the incident response platform using STIX and TAXII to be analyzed as a threat pattern. Structured Threat Information Expression (STIX) is a standardized language that uses a JSON-based lexicon to express and share threat intelligence information in a readable and consistent format across incident response systems. Trusted Automated Exchange of Intelligence Information (TAXII) is the format through which threat intelligence data is transmitted. TAXII is a transport protocol that supports transferring STIX insights over Hyper Text Transfer Protocol Secure (HTTPS). Conversion of vulnerability data into a reliable threat indicator requires analysis and understanding, and interpretation of the vulnerability and the associated risk prior to creating a threat indicator. Once the feeds are ingested into the incident response platform, while SIEM creates alerts, XDR does a deeper analysis using AI/ML for a better representation of the threat, and X/SOAR runs a playbook for automated response.

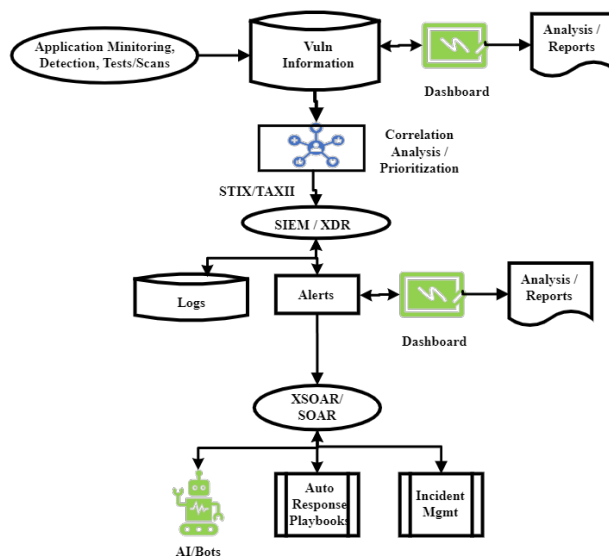


Figure 2: Conceptual Integrated of Vulnerability Data Platform and Incident Response Platform.

4. In Conclusion

This paper has proposed a data consolidation platform that will be foundational to build a comprehensive predictive system. We have also proposed how this data platform can be integrated with SIEM, SOAR, XSOAR, and XDR to establish a fully automated attack detection and response system. As mentioned in the paper, SIEM, SOAR, XSOAR, and XDR are more constructed to suit networks, infrastructure and devices, and sensors; not meant for application security vulnerability information as collected. So, this paper makes a special case that must be made for integration of application security information as part of threat intelligence and threat response.

Acknowledgements

This research has been supported and guided by Dr. Xiaolong Li of Indiana State University, USA, and Dr. John Pickard and Dr. Peng Li of East Carolina University, USA.

Dr. Xiaolong Li is a professor in the Department of Electronics and Computer Engineering Technology at Indiana State University. He received his PhD in Computer Engineering from the University of Cincinnati in 2006. His primary areas of research include modeling and performance analysis of MAC protocol, Internet of Things, Wireless Ad Hoc networks, and sensor networks.

Dr. John Pickard is a professor of Information and Cybersecurity Technology at East Carolina University, North Carolina, USA. He received his PhD in Technology Management from Indiana State University in 2014. His main research areas are internet protocols, convergence of information and operations technologies, and Internet of Things applications.

Dr. Peng Li received his Ph.D. in Electrical Engineering from the University of Connecticut. His professional certifications include CISSP, RHCE and VCP. Dr. Li is currently an Associate Professor at East Carolina University. He teaches undergraduate and graduate courses in programming, computer networks, information security, web services and virtualization technologies. His research interests include virtualization, cloud computing, cybersecurity, and integration of information technology in education.

References

- Akamai, (2023) "Slipping Through the Security Gaps: The Rise of Application and API Attacks", *Akamai*, <https://www.akamai.com/blog/security/the-rise-of-application-and-api-attacks>
- Carielli, S., DeMartine, A., Provost, A.C. and Dostie, P. (2021) "The Forrester Wave™: Software Composition Analysis, Q3 2021-The 10 Providers That Matter Most And How They Stack Up", *Forrester*, August, https://www.forrester.com/report/the-forrester-wave-software-composition-analysis-q3-2021/RES176091?ref_search=3502061_1674835391293&utm_source=PANTHEON_STRIPPED&utm_medium=email&utm_campaign=summit21na&utm_content=blog&categoryid=a89c0000000AKp1AAG%3Futm_source%3DPANTHEON_STRIPPED

