

Research on Security Defense Methods for Big Data Centers

Jing Wang ^a, Yunjie Gu ^b

China Electronic Device System Engineering Corporation, Beijing, China
^a 18911684617@163.com, ^b 3844543747@qq.com

Abstract: By integrating and analyzing massive amounts of data, big data centers can effectively address key issues in multiple fields. However, as the scale and importance of big data centers continue to grow, security issues such as high-risk vulnerabilities exposed due to inadequate security configurations in physical computing facilities, malicious attacks stemming from lax access control at cloud boundaries, effectiveness issues in virtual resource isolation, and insufficient security protection measures for cloud platforms have become increasingly prominent. Based on the existing big data center architecture, this paper designs a security protection framework from four dimensions: physical computing facilities, cloud boundary access zones, cloud tenant zones, and cloud platforms. It integrates security protection functions and measures based on traffic analysis, including access control and risk assessment.

Keywords: Big Data Center; Defense Architecture; Traffic Analysis Engine.

1. Introduction

With the development of information technology, we have ushered in the era of big data. Big data centers have addressed issues related to the storage, high-speed computation, security management, and efficient sharing of massive amounts of data, serving as the "super brain + secure vault" of the data era. By integrating and analyzing vast amounts of data, big data centers can effectively solve key problems in multiple fields. However, as the scale and importance of big data centers continue to grow, security issues have become increasingly prominent. These issues primarily center around four dimensions: physical computing and storage facilities, cloud boundary access zones, cloud tenant zones, and cloud platforms. The risks are more concentrated and have a wider impact [1].

(1) The security policy configuration of physical computing facilities (including servers, network equipment, network security equipment, etc.) does not adhere to the principle of minimization, which may lead to the exposure of high-risk vulnerable ports, resulting in security threats such as vulnerability exploitation, ransomware infection, and malicious implantation in the supply chain.

(2) Due to lax supervision by operation and maintenance personnel, incomplete safety regulations and systems, and issues such as abuse of authority and operational errors, illegal data theft, tampering with system configurations, and other violations have occurred.

(3) With the widespread application of network virtualization technology in cloud data centers, physical boundaries have become blurred, and traditional north-south security protection schemes are no longer effective for east-west traffic and access [2]. Virtual machine drift, rapid node creation and deletion, etc., pose challenges to the speed of security protection strategy deployment and operational maintenance workload.

(4) As the foundation of protection, traffic monitoring is integrated and fused in various current security solutions. In the security operation of cloud data centers, the issue of traffic invisibility, especially the invisibility of east-west access

traffic, needs to be addressed.

(5) The identity authentication of the cloud platform is not strict, leading to unauthorized access and use of resources and data.

2. Design of Security Architecture for Big Data Center

Big data centers typically feature high-bandwidth access, massive nodes, and software-defined characteristics. They implement security protection from multiple levels, including cloud boundary access areas, cloud tenant areas, cloud platforms, and physical storage and computing facilities. The designed security architecture for big data center is as shown in Fig 1:

2.1. Cloud Boundary Access Area

In the network security system of a big data center, the network boundary serves as the fundamental security guarantee and the first line of defense in the security protection system. At the cloud boundary exit side of the data center, there is a DDoS traffic cleaning device, which consists of a cleaning device and a detection device[3]. The cleaning device is responsible for implementing functions such as DDoS attack filtering and abnormal traffic cleaning, while the detection device is responsible for real-time detection of traffic in the network. In terms of traffic redirection, traffic is replicated to the detection device through port mirroring configured on network devices, where it undergoes traffic analysis and detection[4]. When the detection device detects abnormal traffic, it sends information about the abnormal traffic to the cleaning device through the management port. The cleaning device then generates the corresponding redirection route and sends it to the network device, thereby redirecting the abnormal traffic to the cleaning device. After cleaning, the clean traffic is re-injected into the network device for normal forwarding. Boundary access control, intrusion prevention devices, and firewall devices collect network traffic through port mirroring, monitor and block attack and intrusion information in network traffic in real time,

and adopt a serial deployment to monitor all traffic entering the cloud [5].

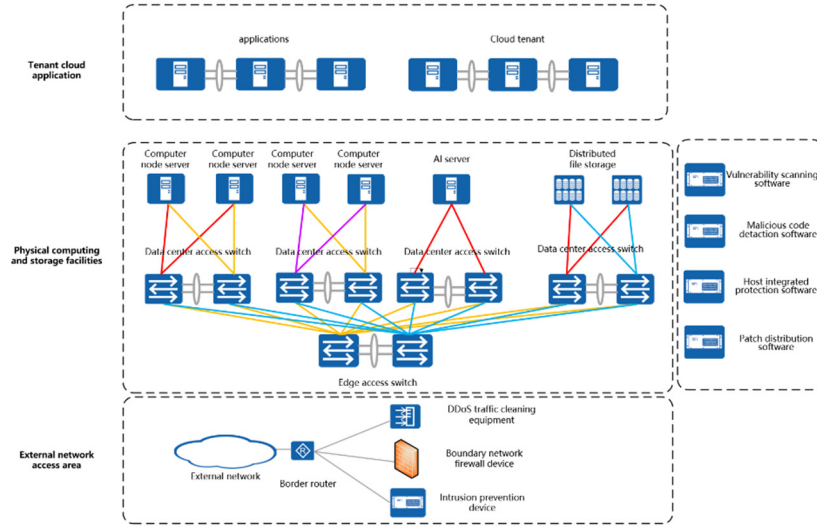


Fig 1. The security architecture for big data center

2.2. Cloud Tenant Security Protection

Cloud tenant security protection primarily refers to the security measures provided for tenant applications on the cloud (including within VPCs). In terms of form, it primarily adopts a virtualized software resource pool solution, encompassing but not limited to security group firewalls, cloud bastions, and so on. Initially, it is essential to ensure the effectiveness of security configurations for internal virtual hosts (including operating systems, databases, and other applications) and host firewall policies. There should be no pre-installed advanced backdoors or viruses on internal hosts. Based on this foundation, a security resource pool is established, allocating protection resources and issuing personalized protection policies on a tenant or application basis. Monitoring is conducted through traffic analysis devices, intrusion warning systems, and other means to ensure the effectiveness of isolation between different cloud tenants or applications.

2.3. Cloud Platform Security Protection

In addition to its primary functions such as monitoring and auditing abnormal states, managing and allocating computing resources, cloud platforms typically come equipped with numerous security features. These include but are not limited to virus protection, intrusion detection, security auditing, vulnerability scanning, patch distribution, and vulnerability remediation[6]. These security measures provide protection against web application layer attacks and also encompass capabilities such as malicious request filtering, sensitive data leakage prevention, and HTTP protocol compliance verification.

2.4. Security Protection for Physical Computing Facilities

The security protection of physical computing facilities is essential, with the core aim of ensuring the stable operation of the host. We should strengthen the physical security protection capabilities of computing facilities from the perspective of physical deployment environment; enhance the network security protection capabilities of computing facilities from the aspects of network equipment

configuration, network security equipment configuration, and network access control configuration; strengthen the system security protection capabilities from the perspectives of server operating system security, software supply chain security, and application service security; and enhance the security management capabilities from the aspects of operation and maintenance personnel, security regulations and rules, with the core aim of comprehensively strengthening the security protection capabilities of computing facilities.

3. Big Data Center Defense Technology

3.1. Analysis of Big Data Center Security Defense Technologies

With the popularization of cloud-native architectures, Docker containers have become core components. However, traditional security monitoring methods struggle against dynamic, short-lived container instances. eBPF (extended Berkeley Packet Filter), as a secure and efficient virtual machine technology running within the Linux kernel, allows user-space programs to dynamically inject and execute custom logic without modifying kernel source code.

A cloud-native security architecture built on eBPF employs a four-layer defense model to achieve comprehensive monitoring and protection, covering the core security risk points in cloud-native environments:

(1) Container Runtime Security: Monitoring Key System Calls

By tracing key system calls such as `execve`, `clone`, and `fork` via eBPF, real-time awareness of process creation and program execution behavior within containers is achieved. Combined with a process whitelist mechanism, malicious program execution can be quickly identified and blocked, preventing container escape attacks [7].

(2) Network Layer Isolation: Enhancing the Effectiveness of K8s NetworkPolicy

Traditional K8s NetworkPolicy relies on iptables for network isolation, which suffers from low rule-matching efficiency and poor adaptability in complex scenarios. eBPF optimizes the data plane, enabling efficient implementation of NetworkPolicy:

1) eBPF-based traffic filtering reduces rule-matching latency to under 1 microsecond, supporting real-time matching of tens of thousands of rules.

2) L7 layer traffic identification is supported, allowing for fine-grained isolation strategies targeting protocols such as HTTP and gRPC.

3) Network traffic anomalies, such as abnormal port access or high-frequency connection attempts, are monitored in real-time.

(3) File System Integrity: Real-time Tracking of Sensitive File Changes

By tracing kernel events related to the file system (e.g., inode, vfs) via eBPF, the creation, modification, and deletion of sensitive files like /etc/passwd and /proc/cmdline are monitored in real-time[8]. Once an abnormal change is detected, an alert is immediately triggered, and the complete operation chain (including process, user, and container information) is recorded, providing a solid basis for security traceability.

(4) Process Behavior Analysis: Detecting Anomalous Process Trees and Privilege Escalation

A process behavior baseline is established using eBPF. By analyzing dimensions such as process tree structure, parent-child relationships, and permission changes, anomalous behaviors are identified. For example, scenarios such as a root-privileged bash process appearing within a container, an unexpected explosion in process forks, or a process accessing across container namespaces are all quickly detected and intercepted.

4. Conclusion

The research findings of this paper aim to design a security protection architecture based on the existing big data center architecture, encompassing four dimensions: physical

computing and storage facilities, cloud boundary access area, cloud tenant area, and cloud platform. It integrates security protection functions and means based on traffic analysis, with access control and risk assessment as extensions.

References

- [1] Liu, F. X. (2024). Data security in big data centers: Regulatory and compliance research.
- [2] Wan, Q. (2019). Practices in security architecture design for cloud data centers. *Information & Communications*.
- [3] McKeown, N., Anderson, T., Balakrishnan, H., & et al. (2008). OpenFlow. *ACM SIGCOMM Computer Communication Review*, 38(2), 69-74. <https://doi.org/10.1145/1341421.1341426>.
- [4] Chiosi, M., Clarke, D., Willis, P., & et al. (2012). Network functions virtualisation introductory white paper [White paper].
- [5] Huang, M., Luo, W. B., & Wan, X. (2019). Research on network security of campus network. *Journal of Physics: Conference Series*, 1187(4), 042113.
- [6] Yusof, M. A. M., Ali, F. H. M., & Darus, M. Y. (2018). Detection and defense algorithms of different types of DDoS attacks using machine learning. In *International Conference on Computational Science and Technology* (pp. 370-379). Springer.
- [7] Wu, B., Zhong, E. H., Tan, B., & et al. (2014). Crowdsourced time-sync video tagging using temporal and personalized topic modeling. In *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 721-730). Association for Computing Machinery.
- [8] Lv, G. Y., Xu, T., & Chen, E. H. (2016). Reading the videos: Temporal labeling for crowdsourced time-sync videos based on semantic embedding. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*.