Unit 10: From Distributed Computing to Microarchitectures

1. **Describe various distributed systems and the pattern in which they evolved.**

As technology advanced, the concept of distributed systems evolved from a single mainframe computer to multiple computers working together to complete tasks more efficiently. A distributed system is a network of components spread across different computers that work harmoniously to achieve a shared objective. Despite its many benefits, implementing a distributed system involves overcoming obstacles like managing failures, ensuring concurrency, maintaining security, standardising data, and ensuring scalability (Packtpub, 2022).

A distributed system is a group of computer programs that work together across multiple computation nodes to achieve a shared objective. This setup relies on these nodes communicating and synchronising over a network. Distributed systems facilitate resource sharing, simultaneous processing, scalability, error detection, and transparency by removing bottlenecks and central points of failure from a system (Zettler, N.D).

Zettler (N.D.) states that centralised computing uses a single computer to execute all computations in one location. In contrast, distributed systems involve multiple nodes communicating with one another. Centralised systems store state in a central node, leading to network congestion and a singular point of failure. Distributed systems, however, have multiple nodes and no single point of failure. One example of a distributed system is the microservices architecture, where applications are separated into distinct components or services, each with its business logic and multiple redundant copies to eliminate central points of failure.

Distributed system paradigms have evolved over the past 60 years to meet the demands of changing computing landscapes. They operate through message passing, without a shared clock, and require reliable and secure operation. Communication models like inter-process communication, remote invocation, and indirect communication support message exchanges. Consistency is a significant challenge, and distributed systems must be resilient to node failure and lost/delayed messages. Distributed systems evolve as scientific, technological, and societal factors progress, resulting in new computer systems and adapted paradigms (Lindsay et al., 2021).

According to GeeksforGeeks (2022), the distributed computing system illustrates how centralised systems periodically develop towards decentralisation. It is all about the movement from centralisation to decentralisation. In the beginning of 1955, we used centralised systems like mainframes; however, nowadays, we most likely use decentralised systems like edge computing and containers. For instance, below is the evolution of distributed systems throughout the years, according to GeeksforGeeks (2022):

1. Mainframe-based computing machines were used in the early 1960s to process large-scale data, providing time-sharing to local clients through teletype terminals. This client-server architecture allowed multiple resources to be sent over a single medium. However, their high cost led to the early development of disk-based storage and transistor memory.
2. In the early 1970s, packet-switching and cluster computing emerged as an alternative to mainframe systems, utilising similar workstations connected via high-speed local-area networks. This allowed for parallelism and global message exchange. ARPANET, created between 1967 and 1974, enabled services hosted on remote machines across geographic boundaries. TCP/IP protocol facilitated datagram and stream-oriented communication over packet-switched autonomous networks, primarily through datagram transport.
3. The evolution of the internet and PCs began with the introduction of TCP/IP technology, transforming it into multiple connected networks. As the number of hosts connected to the network grew, centralised naming systems like HOSTS.TXT couldn't provide scalability. Domain Name Systems (DNSs) were created in 1985 to convert hosts' domain names into IP addresses—early GUI-

based computers with WIMPs enabled home computing, including video games and web browsing.

4. The 1980s-1990s saw the development of HyperText Transfer Protocol (HTTP) and HyperText Markup Language (HTML), leading to the creation of web browsers, websites, and servers. The standardisation of TCP/IP provided infrastructure for interconnected networks known as the World Wide Web, leading to a significant growth in host connections. As PC-based application programs grew, communication became complex, posing challenges in application-to-application interaction. Network computing, enabling remote procedure calls over TCP/IP, became widely accepted for application software communication. Distributed computing applications emerged to address these challenges.
5. Peer-to-peer (P2P) computing is a distributed application architecture that partitions tasks or workloads between peers without a central coordinator. It was introduced in 1999 by Shawn Fanning and has evolved since then. Grid computing allows multiple jobs to be completed by computers connected over a network using a data grid. Web services enable platform-independent communication using XML-based information exchange systems. P2P networks are often created by collections of 12 or fewer machines sharing data and consuming resources. Proper security is challenging due to nodes acting as both clients and servers.
6. Cloud computing combines cluster technology, virtualisation, and middleware to manage resources and applications online, accessible from anywhere. It eliminates the need for updating servers, hardware, or software licenses. Mobile computing enables data transmission over wireless networks, with popular forms including smartphones and tablets. IoT technologies emerged from mobile computing, utilising sensors and software for data exchange. API management platforms evolved to implement scalability, flexibility, portability, caching, and security. Virtualisation, allowing multiple systems within one computer, has become a core feature of distributed systems. Virtualisation options include VM Ware Workstation, Microsoft Hyper-V, and Oracle Virtualization.
7. Fog and edge computing are two approaches to managing large amounts of data generated by mobile computing and IoT services. Edge computing moves client data to the network's periphery, reducing latency issues and improving efficiencies. Fog computing aggregates data at access points, reducing costs and improving efficiencies. Companies like IBM drive these approaches. Container-based application deployment allows applications to run on any environment with a host operating system, with popular platforms like Docker and Kubernetes providing large clusters and communication between services. Distributed systems are programmed by application programmers, with cloud providers managing infrastructure.

2. **Discuss the security attacks to which distributed systems are particularly vulnerable.**

Security breaches can occur due to weaknesses in system security procedures, design, implementation, or internal controls. A thorough analysis, classification, and formal modeling process are necessary to identify these vulnerabilities. A vulnerability lifecycle model has been utilised in case studies, revealing that systems may remain vulnerable even after applying security patches. Operational and information-based vulnerabilities are the two primary types. Threats to systems can exploit these vulnerabilities, resulting in security breaches. Mitigation can be accomplished through avoidance or tolerance, with the release being effective against less complex attacks (Bhargava & Lilien, N.D.).

Suri et al. (2019) discuss various peer-to-peer (P2P) systems attacks, targeting functional elements like P-OP and P-DS. These attacks disrupt connectivity access to other nodes and corrupt data structures. They exploit fundamental P2P features, such as message exchange-based decentralised coordination and partial peer views. Attackers aim to trick peers by providing incorrect data or colluding to create partitions hiding system views. The text also enumerates representative security attacks and their impact on Confidentiality, Integrity, and Availability (CIA). Examples include Denial of Service attacks, which limit access to nodes or communication routes, and collusion attacks, which involve a large subset of peers and override control mechanisms. Pollution and whitewashing attacks compromise P2P systems' integrity by adding incorrect information, endangering P-DS functionality. Routing attacks undermine message-passing mechanisms, modify peers' routing tables, and affect P-DS functionality. Sybil attacks compromise P2P networks' availability or confidentiality through spoofing, while eclipse attacks block peer views, impacting P-OP and P-DS functionality.

Hence, Baker (2023) defines a cyber-attack as an attempt by cybercriminals or hackers to access a computer network or system, aiming to alter, steal, destroy, or expose information. Common types include malware, DoS attacks, phishing, spoofing, identity-based attacks, code injection, supply chain attacks, insider threats, DNS tunnelling, and IoT-based attacks. According to Baker (2023), below are the ten most common types of cyber-attacks to which distributed systems are particularly vulnerable:

Malware is malicious software that harms computers, networks, or servers. It includes ransomware, trojans, spyware, viruses, worms, keyloggers, bots, and crypto-jacking. Ransomware encrypts data, while fileless malware uses legitimate tools. Spyware collects user information, while ads monitor activity. Trojans install through social engineering, worms replicate, rootkits control networks, mobile malware targets devices, exploits, scareware, and botnets.

Denial-of-Service (DoS) attacks are malicious attacks that flood networks with false requests, disrupting business operations. They disrupt routine tasks and cost organisations time, money, and resources. On the other hand, DDoS attacks originate from multiple systems and are faster and more complicated to block, as they require identification and neutralisation of various methods.

Phishing is a cyberattack that uses email, SMS, phone, social media, and social engineering techniques to entice victims to share sensitive information or download malicious files. Common types include spear phishing, whaling, smishing, and vishing. Spear phishing targets specific individuals or organisations through malicious emails, while whaling targets senior executives for money or information theft. Smishing involves sending fraudulent text messages to trick individuals into sharing sensitive data, while vishing involves fake phone calls and voice messages.

Spoofing is a cybercrime technique where an attacker impersonates a known or trusted source to access a target's systems or devices. It can take various forms, including domain spoofing, email spoofing, and Address Resolution Protocol (ARP) spoofing. Domain spoofing involves impersonating a known business or person with fake websites or email domains, while email spoofing targets businesses using forged sender addresses. ARP poisoning intercepts data by tricking a device into sending messages to the hacker.

CrowdStrike's research indicates that 80% of breaches involve compromised identities and can take up to 250 days to identify. These identity-driven attacks are challenging to detect, as it is difficult to differentiate between a user's typical behaviour and that of the hacker using traditional security measures. Common identity-based attacks include Kerberoasting, Man-in-the-Middle (MITM), Pass-the-Hash Attacks, Silver Ticket Attacks, Credential Stuffing, Password Spraying, and brute force attacks. Kerberoasting attempts to crack a service account's password, while MTM involves eavesdropping on conversations to collect personal data or convince victims to act. Pass-the-Hash Attacks steal a user's credentials and use them to create a new session, while Silver Ticket Attacks make forged authentication tickets.

Code injection attacks involve an attacker injecting malicious code into a vulnerable computer or network to alter actions. There are three types: SQL Injection, Cross-Site Scripting (XSS), and Malvertising. SQL Injection uses system vulnerabilities to inject malicious SQL statements into a data-driven application, allowing hackers to alter, steal, or erase database data. Cross-site scripting inserts malicious code within legitimate websites, allowing attackers to steal sensitive information or impersonate users.

Supply chain attacks target trusted third-party vendors, injecting malicious code into applications to infect users. They compromise physical components, particularly in software supply chains, which are vulnerable due to off-the-shelf components like APIs and proprietary code.

Insider threats are internal actors with direct access to company networks, sensitive data, and intellectual property. They can be malicious or negligent. Organisations should implement comprehensive cybersecurity training programs to combat these threats that educate stakeholders on potential attacks, including those by insiders. This will help organisations stay safer and more secure.

DNS Tunneling is a cyberattack that uses DNS queries to bypass security measures and transmit data and code within a network. Infected hackers can launch malware or extract sensitive information. Attacks have increased due to their simplicity and accessibility online.

IoT-based attacks target IoT devices or networks, allowing hackers to gain control, steal data, or launch DoS or DDoS attacks. Connected devices account for nearly one-third of mobile network infections, double in 2019. IoT infections are expected to increase as connected devices grow, and 5G networks may exacerbate this issue.

3. Understand how virtual systems need to be protected due to the specific nature of attacks on them.

A Virtual Machine (VM) is a powerful computing resource that enables businesses to run an operating system that behaves like a completely separate computer in an application window on a desktop. VMs have traditionally been used for server virtualisation, which allows IT teams to consolidate their computing resources and improve efficiency (VMWare, N.D.).

In addition, virtual machines work as a process in an application window on the physical machine's operating system, similar to any other application. The essential files that make up a virtual machine include a log file, an NVRAM setting file, a virtual disk file, and a configuration file. The benefits of virtual machines include being easy to manage and maintain, running multiple operating system environments on a single physical computer, supporting legacy applications, and providing integrated disaster recovery and application provisioning options.

Moreover, there are two types of virtual machines: process VMs and system VMs. Process VMs allow a single process to run as an application on a host machine, providing a platform-independent programming environment by masking the information of the underlying hardware or operating system. System VMs are fully virtualised to substitute for a physical device, sharing a host computer's physical resources between multiple virtual machines.

Therefore, virtualisation technology offers five types of virtualisation: hardware virtualisation, software virtualisation, storage virtualisation, network virtualisation, and desktop virtualisation. Hardware virtualisation consolidates virtual versions of computers and operating systems into a single primary physical server. Software virtualisation creates a computer system with hardware that allows one or more guest operating systems to run on a physical host machine.

According to Pearce (2013), system virtualisation is a powerful platform for system building in modern architectures, using an encapsulating software layer (Hypervisor or Virtual Machine Monitor) that provides inputs, outputs, and behaviour similar to a physical device. This abstraction allows multiple virtual machines to be installed on a single hardware set, offering inherent security benefits. However, the design and

implementation of virtualisation technology have also opened up new threats and security issues.

The security implications of virtualisation are often overlooked due to its rapid growth. Threats can target various components, such as VMMs, VMs, OSs, software, and networks, which can compromise sensitive data and operations. To design secure systems in virtual environments, it is crucial to have a threat model in place. For web app security, Microsoft's threat modelling process is recommended. Additionally, taxonomies such as Lindqvist and Jonsson's incident reporting and Landwehr et al.'s foundational taxonomy are essential for developing computer security taxonomies. A new taxonomy has also been created to establish an architecture for comparing tools, evaluating their use for maintenance and change control, and categorising threats (Pearce, 2013).

Bakshi & Dujodwala (2010) highlight that organisations must consider user experience, security, and available features when developing cloud computing systems. SaaS solutions provide strong security but may have limited extensibility; PaaS platforms offer better protection but less integrated functionality. IaaS options have fewer security capabilities but offer more application-like features. Virtualisation can provide agility, flexibility, cost savings, and increased business value by allowing computing environments to be dynamically created, expanded, shrunk, or moved. Moreover, virtualisation enables the consolidation of underutilised physical servers, resulting in significant cost savings. It's worth noting that virtual worlds require substantial computing power and should have an intrusion detection system installed on the virtual switch to log network traffic and respond to potential attacks.

Hence, Shabut et al. (2016) state that the frequency and cost of cyber-attacks are rising, threatening individuals, businesses, and critical infrastructure. While traditional security tools fall short of preventing all attacks, enhancing users' ability to detect, control, and defend against such threats is imperative. It is essential to boost cyber security schemes and augment user knowledge. By leveraging cyber-attack taxonomy and classification, users can identify attacks and take preventive measures. Combining different techniques can deliver an effective protection scheme for real-time decision-making.

References:

Packtpub (2022). Available at: https://subscription.packtpub.com/book/cloud-and-networking/9781801072212/2/ch02lvl1sec03/the-evolution-of-distributed-systems.

Zettler, K. (N.D). *What is a distributed system?* [online] Atlassian. Available at: https://www.atlassian.com/microservices/microservices-architecture/distributed-architecture.

Lindsay, D., Gill, S.S., Smirnova, D. & Garraghan, P. (2021). The evolution of distributed computing systems: from fundamental to new frontiers. *Computing*. doi:https://doi.org/10.1007/s00607-020-00900-y.

GeeksforGeeks. (2022). *Evolution of Distributed Computing Systems*. [online] Available at: https://www.geeksforgeeks.org/evolution-of-distributed-computing-systems/.

Bhargava, B. & Lilien, L. (N.D.). *Vulnerabilities and Threats in Distributed Systems*. [online] Available at: https://www.cs.purdue.edu/homes/bb/bhargava-vuln-threats.pdf.

Suri, N., Beznosov, K. & Vukolić, M. (2019). *Distributed Systems Security Knowledge Area Issue*. [online] Available at: https://www.cybok.org/media/downloads/Distributed_Systems_Security_issue_1.0.pdf.

Baker, K. (2023). *Most Common Types of Cyber Attacks Today | CrowdStrike*. [online] crowdstrike.com. Available at: https://www.crowdstrike.com/cybersecurity-101/cyberattacks/most-common-types-of-cyberattacks/.

VMWare (N.D.). *Virtual Machine*. [ONLINE] Available at: https://www.vmware.com/uk/topics/glossary/content/virtual-machine.html.

Pearce, M. (2013). Virtualisation: Issues, Security Threats, and Solutions. *ACM Comput. Surv. ACM Computing Surveys*, [online] 45(17). doi:https://doi.org/10.1145/2431211.2431216.

Bakshi, A. & Dujodwala, Y. B. (2010). *Securing Cloud from DDOS Attacks Using Intrusion Detection System in Virtual Machine*. [online] Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=5437670.

Shabut, A.M., Lwin, K.T. and Hossain, M.A. (2016). Cyber attacks, countermeasures, and protection schemes — A state of the art survey. *2016 10th International Conference on Software, Knowledge, Information Management & Applications (SKIMA)*. [online] doi:https://doi.org/10.1109/skima.2016.7916194.