Extending the Cloud with Fog: Security Challenges & Opportunities

Completed Research Paper

Dr. Jordan Shropshire
University of South Alabama
jshropshire@southalabama.edu

Abstract

Fog is an emerging computing paradigm which provides storage, processing, and communication services closer to the end user. It reduces latency, provides location awareness, and supports high-density wireless networks. Fog does not replace cloud computing. It extends the cloud to the edges of the network. If properly integrated, the resulting infrastructure would provide reduced latency, geographic awareness, improved data streaming, and access to commodity resource pools. This paper provides a high-level overview of fog computing. It then describes several architectural revisions necessary to support a combined fog-cloud platform. The two primary features are a unified virtualization layer consisting of cloud and fog compute nodes and a management backplane to facilitate communication between the data center and the network edges. Assuming these modifications, the combined architecture is then subjected to a systematic security review. This analysis focuses on implications to confidentiality, integrity, and availability. The paper also includes recommendations for overcoming these security challenges through deliberate, security-conscious designs.

Introduction

Cloud computing has emerged as a mainstream model for meeting the computing needs of enterprises and end users. The cloud ecosystem is premised on the concept of resource pooling. Converging infrastructure allows for increased reliability, superior economies of scale, and real-time resource provisioning (Carroll et al. 2011). Clouds provide elastic computing in that they provide on-demand services which scale to meet peek loads. Deploying software in cloud environments also increases agility, as the process of re-provisioning technological infrastructure is greatly simplified. Further, clouds provide location independence because they are accessible to any client with a web browser. Clouds are built upon multiple layers of hardware and software. They consist of commodity servers, hypervisors, and guest operating systems and guest applications (Zaborovskiy et al. 2013). In most cases, they are housed within large data centers or co-location facilities, located in close proximity to core fiber networks. Because data centers are heavy energy consumers, they are often developed in regions with access to relatively cheap power.

Although centralized computing systems offer a myriad of benefits, they are saddled with a few drawbacks. A primary limitation is delay – the lag between client request and cloud response. This can be explained by the fact that data centers are often located well away from major cities and population areas. The physical distance between data centers and end users and has an impact on latency. This is problematic for applications which lean heavily on streaming data and offline processing and storage. Examples of such devices are autonomous systems, sensor networks, mobile devices, and clients with thin-layer operating systems. Collectively, these devices compromise the Internet of Things (IoT) (Mingozzi et al. 2013). To meet the latency requirements of modern applications, a new paradigm was proposed. This computing model, called fog, was designed primarily to reduce delay (Bonomi et al. 2012). Fog meets enhanced network performance requirements by locating data, compute, and networking

capabilities closer to end nodes. It also provides location awareness, enhanced mobility features, and support for real-time processing (Hong et al. 2013; Somorovsky et al. 2011; Zhu et al. 2013). In contrast with centralized clouds, fog nodes are geographically distributed. They are deployed near wireless access points in areas which sustain the heaviest usage. Fog devices may take the form of stand-alone servers or as network devices with onboard computing capabilities.

Fog does not replace cloud computing. Rather, fog computing extends the cloud to the edges of the network. The concepts of cloud and fog computing can be integrated into a single platform to achieve the best of both worlds: reduced latency, geographic awareness, improved data streaming, and access to commodity resource pools (Madsen et al. 2013). The resulting environment supports advanced applications and services. Integrating cloud and fog architectures is not a simple matter. A number of physical and logical changes must be made. For instance, it will be necessary to extend the virtualization layer beyond the cloud to include fog nodes. Furthermore, management networks will have to span the distance between data centers and network edges. Having previously enjoyed physical isolation, sensitive cloud data will be forced to pass over public infrastructure. These changes carry significant security implications (Dhillon et al. 1996; Mingozzi et al. 2013). The purpose of this paper is to perform a systematic security review of the proposed infrastructure. This analysis focuses on elements of the CIA triad: confidentiality, integrity, and availability. The goal is to identify potential weaknesses before any software is developed.

The remainder of this paper is organized as follows. The following section provides background information. It describes the element of the CIA triad and then introduces the fog computing paradigm, explaining its uses and benefits. The third section discusses the integration of the fog and cloud platforms into a unified architecture. Based on this merger, the fourth section describes a systematic security analysis. This analysis focuses on potential implications to confidentiality, integrity, and availability. The fifth section suggests a path toward overcoming these security risks. Finally, recommendations and concluding thoughts are shared.

Background

CIA Triad

The elements of the CIA triad – confidentiality, integrity, and availability – make up a widely-used benchmark for evaluating information system security (Dhillon et al. 2000). Information security audits are often centered on the implications to these three criteria.

Confidentiality provides assurance that information is only shared with authorized individuals or entities. It also prevents disclosure to persons unauthorized to access it. The science of ensuring data confidentiality is cryptography, the process of encryption and decryption (Angell 1993; Morsy et al. 2010). Encryption is an accepted and effective way of protecting data in transit. Increasingly, data at rest (in storage) is also encrypted as a safeguard against leaks. Breaches of confidentiality occur when data is not handled in a manner adequate to safeguard the confidentiality of information. These disclosures may occur when passwords or identifiers are shared with untrusted persons or when an encryption key is cracked. The terms confidentiality and privacy are often used interchangeably.

Integrity is the assurance that information is trustworthy, complete, correct and authentic. Integrity controls ensure that information can be relied upon to be sufficiently accurate for a specific purpose (Morsy et al. 2010). This includes ensuring the integrity of the source of information. Integrity controls also protect against improper or unauthorized modification of information. Integrity protection controls may be divided into two groups: preventive mechanisms (such as access controls) and detective mechanisms which catch unauthorized modifications. The latter group often works as a fallback, catching alterations when preventative mechanisms have failed. Examples of such controls include separation of processes, providing least privileges to all constituents, and rotation of duties among peer devices (Dhillon et al. 1996).

Availability is the assurance that systems responsible for information service provisioning will be responsive, accessible, and meet expected standards (Biennier et al. 2005). Attacks against availability often involve denial of service attacks. Natural and manmade disasters may also affect availability.

Controls for availability usually center on resource allocation, distribution of systems, and backup systems. Availability controls are also addressed in disaster recovery or business continuity plans.

Fog computing

Fog computing offers a highly virtualized platform that provides compute, storage, and networking services between end devices and data centers (Hong et al. 2013). As with the cloud, fog is predicated on the availability of compute, storage, and connectivity resources. These resources must be located within close physical proximity to users to alleviate problems associated with cloud computing. Fog nodes may take the form of servers or networking equipment with additional computational resources. They may even be integrated into wireless access points. Fog nodes will typically be located at the edge of the network, within close proximity to end users. Figure 1 (below) presents a potential cloud-fog architecture.

If implemented as expected, fog will provide a number of non-trivial benefits and offer myriad opportunities for new applications (Zhu et al. 2013). Some defining characteristics of fog are edge location, awareness, and low latency. These features would support streaming video, gaming, distributed computing, monitoring, and control applications. Further, fog will be geographically distributed. This means that moving vehicles, robots, and autonomous systems will be able to receive high quality streaming content even as they pass between proxies and access points, because fog nodes will be positioned along roadways, highways, and cellular phone towers. Fog will also support large-scale sensor networks. These sensors may be dispersed in remote areas for environmental monitoring or they may be used for controlling industrial systems such as power grids, water treatment facilities, and factories (Hong et al. 2013). Fog will provide enhanced support for mobile devices. It is expected that fog applications will be optimized for direct communication with mobile devices. These applications will decouple host identity from geographic location. Further, fog will support near-real-time interaction. Instead of waiting for batch processing in a data center, fog nodes will provide compute services in close proximity to end devices. To sum, the addition of the fog platform provides location awareness, geographic distribution, and reduced lag.

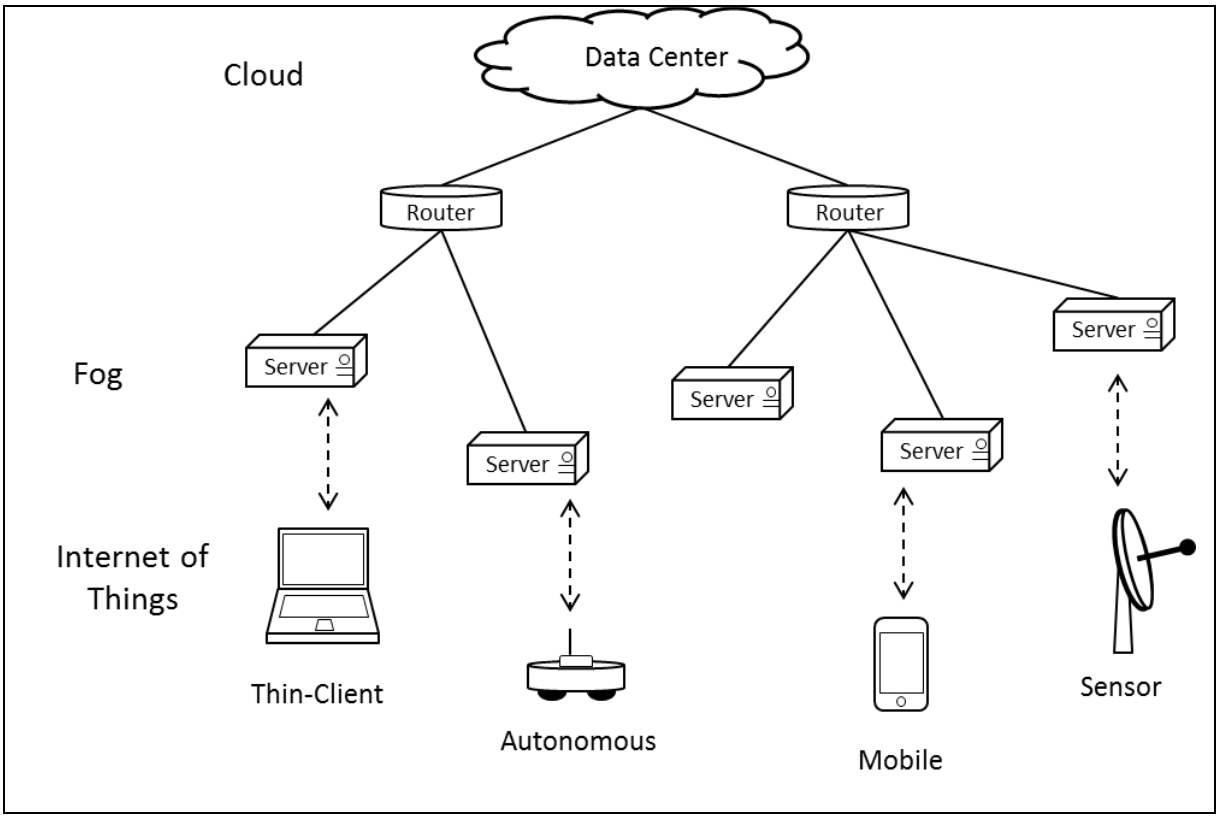

Figure 1: Fog Computing Architecture

Integrating Fog and Cloud

If properly integrated, cloud and fog platforms could support a wide variety of optimized services. These services would be poised to take advantage of the complimentary benefits in both structures. To create a unified platform, the cloud virtualization layer must expand beyond the data center to reach fog nodes. To facilitate this expansion, a management network which supports virtualization must also extend to the edges. These changes are depicted in Figure 2 (below) and are discussed herein.

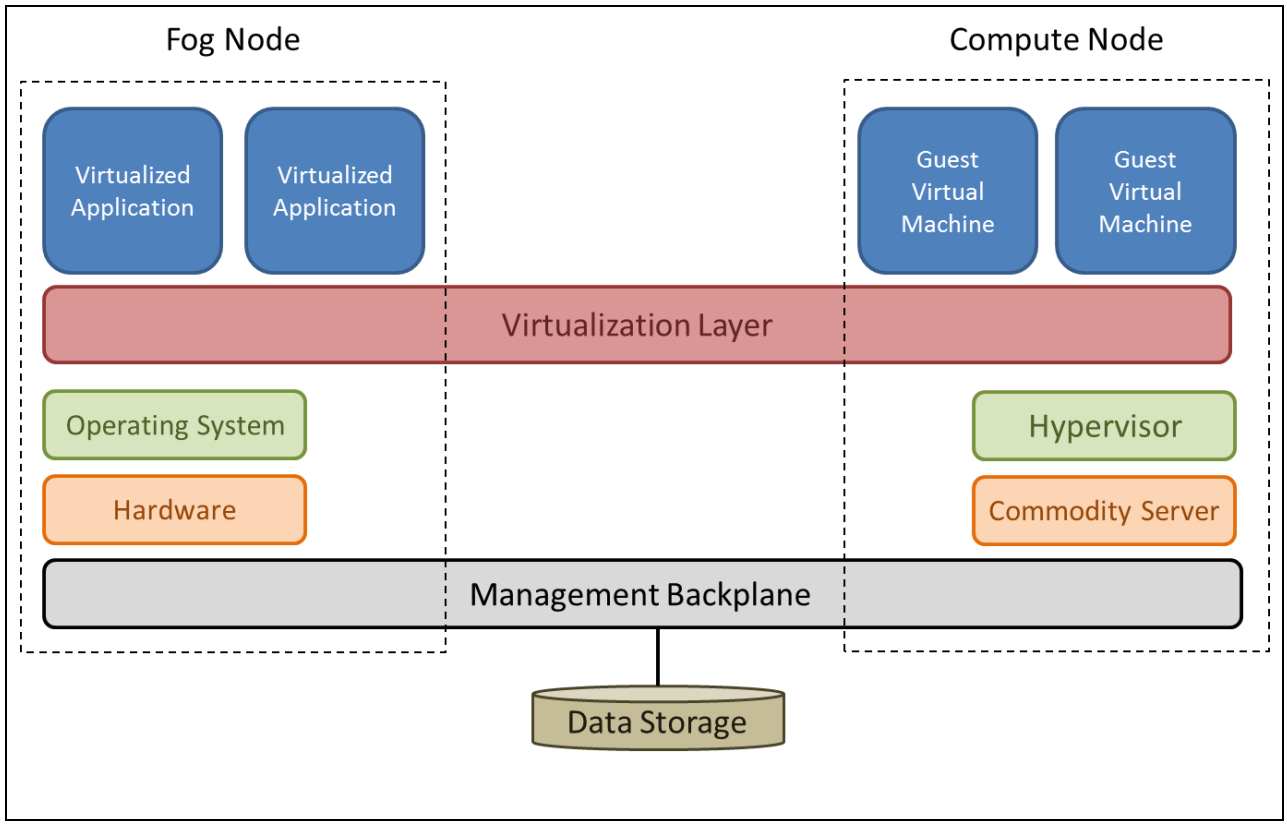

Figure 2: Cloud-Fog Integration

Presently, cloud computing relies on a virtualized environment for software hosting (Mingozzi et al. 2013). Applications and operating systems are deployed in isolated software containers called virtual machines. Virtual machines interact with their hosts' resources via hypervisors. Hypervisors provide processing, memory, connectivity and storage as commodity resources. In this way, a virtual machine is not tied to a specific node. Further, a single compute node is capable of supporting multiple guest virtual machines. Virtualization has proven to be a very effective tool for service hosting in the cloud. It is assumed that the backend software for a distributed application will be deployed in a cloud as a virtual machine. To simplify to the deployment of frontend software, the virtualization layer will need to extend beyond the cloud to incorporate fog nodes. This will result in a single logical substrate. It will free designers from the specification of many technical details, allowing them to develop front-end components which take advantage of reduced latency and location awareness. The concept of a unified virtualization layer does not necessarily imply centralized management of fog and cloud installations. Adoption of open virtualization formats and development of standard administrative processes will allow for decentralized management.

To facilitate communication between virtual machines on geographically-distributed nodes, it will be necessary to support some form of network virtualization. Virtualized networks combine network hardware and software functionality into a single, software-based entity for exchanging packets (Zaborovskiy et al. 2013). External virtual networks combine server network interfaces and routers, switches, firewalls, and load balancers into a single virtual unit. The resulting platform provides network functionality to virtual machines without forcing them to interpret underlying infrastructure. In virtual networks, virtual machines are equipped with virtual network interface cards, which are associated to physical network cards via virtual switches. With external virtual networks, linking virtual machines on physically-removed compute nodes is as simple as linking virtual machines on the same server. Because virtual networks simplify the connectivity process for those outside of networking, it is assumed that some form of this concept will be featured in collaborative cloud-fog architectures. In order to support network virtualization, data centers rely on isolated managed networks. These networks are physically separated from front-end and storage networks. Besides enabling network virtualization, they also support the

control and administrative functions needed to coordinate activity within data centers. In the future, it will be necessary to extend the management network beyond the cloud. This will facilitate a virtualized network which links fog nodes with cloud servers and simplifies the development of distributed applications.

Security Implications

Having reviewed the elements of the CIA triad, introduced the concept of fog computing, and explored the cloud-fog computing concept, this now provides an analysis of the emerging architecture. The unified computing structure will results in a number of key service improvements. However, it will also challenge prevailing assumptions regarding infrastructure assurance. Although many technical details have not yet been established, it is possible to predict attack vectors based on the known weaknesses in cloud platforms and the expected functionality of fog computing. This section provides a high-level overview of potential vulnerabilities to system confidentiality, integrity, and availability:

Confidentiality

The cloud-fog platform offers a number of advances over present architectures. Chief among these is location-awareness. The proposed system provides the ability to design applications and services which take end user position into account. Although this is a benefit, it also presents a major liability. At the application layer, there is no precedent for separating user identity from user location (Somorovsky et al. 2011). This would be difficult to accomplish at the transport layer without new protocols. Thus, it is expected that communication streams between fog nodes and clouds would contain both elements in a single transmission. If this data were intercepted, location-based attacks on end nodes would be possible. This problem is exacerbated by the fact that intercommunications will occur over independent intermediary networks. These links will be targeted using an array of network exploits. The primary concern is packet sniffing – collecting packet streams, reassembling data, and parsing user information (identity, location). This poses a significant risk to data confidentiality. Further, there is little that can be done to secure these public network links. To sum, the combination of richer endpoint data and the possibility of interception give pause for concern.

Integrity

The proposed architecture presents two areas of weakness with respect to integrity. The first concerns the relative weakness of in-place authentication practices. It will be necessary for fog and cloud components to identify themselves before conducting certain transactions (e.g. accepting data for backend processing or releasing the identities of wirelessly-connected clients). Clouds already have rudimentary systems for authenticating compute nodes to administrative servers. These authentication techniques are adequate for data centers because they fall under a single management domain and exist in closed ecosystems. However, they would present a non-trivial weakness in open environments. Lacking a robust deterrent, attackers could exploit the identity management system and masquerade as legitimate compute nodes. This attack vector has already been used in the cloud domain (Carroll et al. 2011). With a dense network of edge servers falling under different management domains, cloud and fog nodes would have little shared information which they could use to prove their identities to each other. Attackers could take advantage of this by assuming the identity of fog nodes and attempting to authenticate themselves to cloud computing systems. If successful, this would provide access to backend processes and vast data stores. Alternatively, attackers could direct their attention to end users. Taking on the role of a fog node, attackers could offer wireless connectivity and seemingly legitimate services to clients. This would afford an opportunity to steal clients' login credentials. It would also be possible to manipulate autonomous entities which rely on distributed software for coordination and control.

The second integrity-related weakness concerns the insecure management backplane. In order to facilitate a unified logical substrate, it would be necessary to expand the management backplane beyond the data center. Clouds typically incorporate management networks to support the administration, control, and monitoring of hardware, hypervisors, and virtualized software (Kanuparthi et al. 2013). These networks are physically isolated from front-end traffic. With the extension of the management backplane comes the loss of physical isolation. This will expose management traffic to a range of threats with little in the way of security. Aware of their lax security, hackers will probe these management

networks for exploitation opportunities. They will use a combination of traditional attack methods and techniques developed specifically for the cloud-fog architecture. For instance, if an encryption key is cracked it would be possible to alter the content of management packet payloads. This would represent a major blow to system integrity. Given that many virtualization platforms rely on unsophisticated authentication techniques, it may also be possible to inject malicious data into communication streams. If successful, these attacks would alter the behavior of the cloud-fog ecosystem. For instance, clouds make workload distribution decisions based on meta-data reported by compute nodes. This data includes metrics of server performance, workload, and resource availability. By injecting malicious data into the packet stream, attackers could make it appear as though large swaths of infrastructure are unavailable.

Availability

With respect to availability, the cloud-fog architecture has two weak points. The first is the reliance on distributed images. The edge virtualization environment will likely be integrated with the cloud such that a single logical layer spans the platforms. In this configuration, software instances running on fog nodes will either be retained in the cloud or they may be neutrally located. In either case, compute nodes will end up streaming virtualized images over public networks. This represents the first category of vulnerability: reliance on distributed software. Virtualization is not only sensitive to packet loss, but to delay as well. Interruptions as short as a few milliseconds will halt or even corrupt streaming software, making hosted services unavailable to end users (Kaufman 2009). Relatively simplistic attacks would be successful. For instance, sending a burst of malformed packets to an intermediary network device could cause enough delay to force a restart. Such an attack would not need to be sustained for an extended period of time. It would only need to last long enough to trigger an interruption. It could then resume only when the targeted instance begins to recover. Short bursts would escape anomaly detection while denying services to end users.

The second availability weakness concerns the limited capacity of fog nodes. Compared with clouds, they could be overwhelmed by relatively small denial of service (DOS) attacks (Hong et al. 2013). DOS-based attacks could be application or packet-based, focusing on exhausting memory, processing capabilities, or overwhelming network interfaces. Although the temporary loss of a single fog node would not be disastrous, a geographically-coordinated attack could have serious implications. For instance, attackers could target a contiguous series of edge nodes along a highway. This would impair the functionality of automobiles which rely on streaming content for navigation or even autonomous control. Further, a concentrated series of small-scale DOS attacks could disrupt the sensor networks which support public infrastructure. For instance, incapacitating fog nodes in power plants, water treatment facilities, or airports could cause a ripple effect throughout a large geographic region. It is expected that fog networks in shopping areas, industrial parks, and business districts would be targeted for extortion under the threat of localized DOS attacks. Such attacks would be harder to detect and mitigate than DOS attempts on traditional infrastructure. Because fog is designed for direct connectivity with end users, there are no upstream peering points from which it is possible to observe traffic patterns.

Towards a Secure Architecture

This research holds that the path to a secure computing system begins with a high-level focus on design. Towards an integrated cloud-fog environment, two architectural modifications have been projected. The first change focuses on the development of an integrated virtualization layer. This layer would span from the data center to the network edges, incorporating cloud and fog compute resources in the new logical substrate. The second change is the extension of the management backplane to the network edges. This would allow for simplified communication among distributed virtual machines, because connectivity would be abstracted in the virtual layer. Based on these changes and on the inherent weaknesses in fog and cloud, several vulnerabilities were anticipated. Potential threats associated with these vulnerabilities have also been identified. The purpose of this section is to identify solutions for mitigating these vulnerabilities. Table 1 (below) summarizes these recommendations.

As a highly virtualized platform, Fog will rely on remotely located virtual machines. These images will be executed at the network edges, even though they reside in centralized storage. Remotely-located will be sensitive to interruption. Even just a few missing or delayed packets could corrupt an image. If attackers can determine the location and network path to the remote storage location, they cloud interrupt runtime

execution using a combination of network attacks. To circumvent this risk, the proposed architecture should make use of decentralized storage arrays. The physical locations of these storage facilities should be geographically distributed throughout the network edges. Decentralized storage would reduce the reliance on a particular data storage center. Without a specific network path to overwhelm, the process of interrupting VM operations would become increasingly difficult. Alternatively, attackers could focus on overwhelming the fog nodes themselves. Denial of service attempts which successfully target a geographic cluster of fog nodes would have widely-felt repercussions. In the cloud, the typical response is to muster additional resources to bolster contested services. Within the fog paradigm, the best approach may be to direct service requests to responsive nodes in close geographic proximity. This will require the development of a protocol which accounts for the availability of specific services before routing traffic to different nodes.

If the management backplane is extended beyond the data center, it will be subjected to a number of network-based attacks. Attackers are expected to attempt to infiltrate the administrative processes supporting virtualization by manipulating network activity. The management backplane was originally designed for exclusive usage in data centers. The assumption of physical isolation negated development of advanced security features. Thus, a number of critical functions performed over the management network may be vulnerable. One solution is to systematically identify the weakest processes and harden them at the application level. This may be the only feasible approach, as it would be impossible to guarantee packet confidentiality and integrity over public networks. However, it will still be necessary to increase the strength of current authentication procedures. It is possible that attackers could crack an authentication key and masquerade as fog nodes. To diminish the likelihood of this threat, more sophisticated authentication procedures should be codified into existing communication schemes.

Vulnerability	Threat	Recommendation
Reliance on distributed images	Runtime execution errors	• Store images at midpoints between cloud and fog clusters. • These locations should be selected with geographic proximity in mind.
Isolated edge nodes	Concentrated denial of service attacks	• Develop techniques for routing user traffic to edge nodes with spare capacity
Open management backplane	Process exploitation via open network	• Redesign administrative processes under the assumption of a low-trust backplane.
Weak authentication procedures	Brute force attacks	• Integrate asymmetric techniques for verifying node identities
Intermediary networks	Identity-related attacks	• Implement robust encryption, VPNs, and leased lines. • Reduce sensitivity to delay and packet loss

Table1: Overcoming Weaknesses

Conclusion

The fog computing paradigm will support the next generation of applications and services. Enterprises with a competitive edge in cloud computing should consider integrating this platform into their existing infrastructure. It is expected that implementation costs will be less for organizations that already possess the in-house talent required to support a cloud computing system. While cloud computing offered the advantage of cost savings, fog enables a new breed of applications. First-mover advantage is therefore essential. Once a new application or service achieves critical mass, it will be difficult to lure users onto other platforms.

When considering a strategy which includes early adoption of this emerging capability, potential costs and benefits should be carefully weighed. The benefits of fog computing are: reduced response time, geographic proximity data, and support for the internet of things. It is expected that fog nodes will be

owned and administered by a service provider, with organizations paying for the privilege to host their applications (similar to SaaS cloud computing). The benefits of the integrated cloud-fog architecture are: support for distributed applications, best-of-breed approaches to resource allocation, and a platform for the next generation of wireless applications. It should be noted that the integration cannot begin until edge hardware is in place. The costs to the organization are: modest reorganization of computing infrastructure, significant logical redesign, and the need to hire/train/equip IT professionals to implement, administer, and secure the new infrastructure. The fog computing paradigm will likely see deployment over the next 3-5 years. This provides ample time for the development of a fog strategy for attaining competitive advantage. It also allows for contingency planning for scenarios such as data theft/compromise. For their money, organizations get a powerful platform that enables a number of features for which the cloud is currently lacking.

Although many of the implementation details surrounding fog have yet to be established, this research makes reasonable projections and assumes conservative security implications. The goal is not to provide technical guidance, but to lay the foundation for considering security before software is published and distributed. In the future, more specific analyses should be conducted in order to identify specific weaknesses in fog computing.

REFERENCES

Angell, I. 1993. "Computer security in these uncertain times: the need for a new approach," in *Proceedings of The Tenth World Conference on Computer Security, Audit and Control, COMPSEC*: London, UK, pp. 382-388.

Biennier, F., and Favrel, J. 2005. "Collaborative business and data privacy: Toward a cyber-control?," *Computers in Industry* (56:4), pp 361-370.

Bonomi, F., Milito, R., Zhu, J., and Addepalli, S. 2012. "Fog computing and its role in the internet of things," in *MCC workshop on Mobile Cloud Computing (MCC '12)*: Helsinki, FI, pp. 13-16.

Carroll, M., van der Merwe, A., and Kotze, P. 2011. "Secure cloud computing: Benefits, risks and controls," in *Information Security South Africa (ISSA)*: Johannesburg, ZA, pp. 1-9.

Dhillon, G., and Backhouse, J. 1996. "Risks in the use of information technology within organizations," *International Journal of Information Management: The Journal for Information Professionals* (16:1), pp 65-74.

Dhillon, G., and Backhouse, J. 2000. "Technical opinion: Information system security management in the new millennium," *Communications of the ACM* (43:7), pp 125-128.

Hong, K., Lillethun, D., Ramachandran, U., Ottenwalder, B., and Koldehofe, B. 2013. "Mobile fog: a programming model for large-scale applications on the internet of things," in *Proceedings of the second ACM SIGCOMM workshop on Mobile cloud computing (MCC '13)*, pp. 15-20.

Kanuparthi, A., Karri, R., and Addepalli, S. 2013. "Hardware and embedded security in the context of internet of things," in *Proceedings of the 2013 ACM workshop on Security, privacy & dependability for cyber vehicles*: Berlin, DE, pp. 61-64.

Kaufman, L. 2009. "Data security in the world of cloud computing," *IEEE Security & Privacy* (7:4), pp 61-64.

Madsen, H., Albeanu, G., Burtschy, B., and Popentiu-Vladicescu, F. Year. "Reliability in the utility computing era: Towards reliable Fog computing," 2013 20th International Conference on Systems, Signals and Image Processing (IWSSIP), Bucharest, RO, 2013, pp. 43 - 46.

Mingozzi, E., Tanganelli, G., Vallati, C., and Di Gregorio, V. 2013. "An open framework for accessing Things as a service," in *2013 16th International Symposium on Wireless Personal Multimedia Communications (WPMC)*: Atlantic City, NJ, pp. 1 - 5.

Morsy, A., Grundy, J., and Mueller, I. 2010. "An analysis of the cloud computing security problem," in *17th Asia-Pacific Software Engineering Conference (APSEC 2010) Cloud Workshop*: Sydney, AU, pp. 8-15.

Somorovsky, J., Heiderich, M., Jensen, M., Schwenk, J., Gruschka, N., and Iacono, L. 2011. "All your clouds are belong to us: Security analysis of cloud management interfaces," in *Proceedings of the 3rd ACM Workshop on Cloud Computing (CCSW '11) Security Workshop*: Chicago, Ill.

Zaborovskiy, V., Lukashin, A., Popov, S., and Vostrov, A. 2013. "Adage mobile services for ITS infrastructure," in *2013 13th International Conference on ITS Telecommunications (ITST)*: Tampere, FI, pp. 127-132.

Zhu, J., Chan, d., Prabhu, M., Natarajan, P., Hu, H., and Bonomi, F. 2013. "Improving web sites performance using edge servers in fog computing architecture," in *2013 IEEE 7th International Symposium on Service Oriented System Engineering (SOSE)*: Redwood City, CA, pp. 320-323.