A SYSTEMATIC LITERATURE REVIEW OF FOG COMPUTING

An Tran Thien  
Østfold University College  
Halden 1783, Norway  
an.t.thien@hiof.no  

Ricardo Colomo-Palacios  
Østfold University College  
Halden 1783, Norway  
ricardo.colomo-palacios@hiof.no

Abstract: Fog Computing is an emerging paradigm that is deemed as the next generation of Cloud Computing. In this day and age, the proliferation of data and data nodes in the Internet of Things has posed problems for the cloud architecture. Fog computing can address those issues by providing elastic resources to end users at the edge of the network. The objective of this study is to obtain an overview of reference architectures, related applications and challenges with the fog paradigm. To this end, a systematic literature review (SLR), which offers a structured approach to better comprehend fog computing is conducted. The study reveals that there is a wide range of applications that can benefit from fog computing; nevertheless, they are not fully implemented in reality. The literature only presents conceptual architectures or proposed prototypes. In addition, the study also shows that security and privacy are among the great challenges for fog computing.

1. INTRODUCTION

The Internet of Things (IoT) refers to an environment consisted of interrelated objects (the so-called ‘things’) that are uniquely identified and can transfer data through the Internet with minimum human intervention (Perera et al. 2014). The goal of IoT is to take advantage of the information collected from those devices to perform synthesis, analysis, support decision making and finally make tasks richer and easier to carry out. In a press release on IoT published by Juniper Research (2015), the number of devices connected to the Internet was placed at 13.4 billion in 2015, which already outnumbered human beings on the planet. It is estimated that the number will have been staggering by the year of 2020: 38.5 billion, or roughly 5 connected devices for every person in the world.

The IoT heavily relies on cloud computing as a promising solution due to certain limitations of smart devices: inadequate storage, restricted computation power and battery (Zheng et al. 2014). Therefore, cloud computing has been used to provide elastic resources for IoT applications in support of processing, analysis and storage of data (Botta et al. 2016). Vera-Baquero et al. (2015) presented a cloud-based system for supply chains with the aim of performing analysis of historical big-data over highly distributed environments. However, with the ever-increasing proliferation of heterogeneous physical and virtual objects, the cloud will be burdened by a vast amount of connections and data transactions. This is because data is collected and synthesized from IoT networks that consist of smart devices and is always sent back to cloud servers for them to store and process. Moreover, not all applications are suitable for cloud deployment because this computing paradigm has its own drawbacks such as high latency, lack of mobility support and geo-distribution (Bonomi, Milito, Zhu, & Addepalli, 2012). For instance, it is not acceptable for health-monitoring and emergency applications to be delayed by data transfer (Dastjerdi & Buyya, 2016). Thus a new platform called “Fog computing” has been introduced by Cisco in an attempt to address the shortcomings of the cloud architecture (Bonomi et al., 2012).

The underlying principle of fog technology is edge computing in which it allows data storage and computation close to end users at the edge of the network which, in turn, provides a new breed of applications and services (Bonomi et al., 2012). Fog computing, however, is not a replacement of the cloud technology; in fact, the two

architectures and end users together form a three layer service delivery model (S. Yi, Hao, Qin, & Li, 2015) as illustrated in Figure 1.

![Three-layer architecture of fog computing](image)

**Figure 1:** Three-layer architecture of fog computing (S. Yi et al., 2015)

In fog computing, the most fundamental components of the architecture are called *fog nodes*. They are an array of modular hardware and software elements that can be configured to execute specific functions including providing resources for services (Byers & Wetterwald, 2015). They can be resource-poor devices such as access points, set-top-boxes, routers (Willis, Dasgupta, & Banerjee, 2014), or resource-rich modules such as Cloudlet and IOx (Shanhe Yi, Li, & Li, 2015). Cloudlet is a small-scale cloud data center that is located at the edge of the Internet (“Cloudlet” 2016). Quwaider and Jararweh (2013) build a cloudlet-based system for big data collection which seeks to take advantage of the proximity of users in relation to cloudlet-based centers. In an experimental application, Jalali et al. (2016) revealed that nano servers in fog computing can help save energy of the centralized data centers in cloud computing.

Fog computing is an emerging paradigm and therefore still at its infancy. To the authors’ knowledge, there has been no systematic literature review (SLR) published on this promising platform. There is a publication by Dastjerdi et al. (2016) that offers an overview of fog computing in terms of principles, architectures and applications. This is, nevertheless, mostly a descriptive review which is mainly focused on the introduction of fog computing with one reference architecture. This paper employs a methodical, structured analysis of existing literature in order to obtain its objectives, which are to understand the usage, applications and challenges so far with fog computing. Hence, the SLR provides a more comprehensive overview of fog computing with multiple existing architectures and applications, which reveals possibilities within this new paradigm.

The rest of the paper is organized as follows: first, there is a foundational background of fog computing; second, the methodology used for this SLR is presented; third, the SLR process is described; fourth, the results of the SLR execution is outlined; and finally there is a discussion of the findings.
2. BACKGROUND

2.1 Definitions of fog computing

There are no unanimous definitions of fog computing considering it is in its immature state. This has led to several definitions of the fog, relying upon different perspectives as follows:

Definition 1: Fog computing is an extension of the cloud platform from the core to the edge of the network (Bonomi et al., 2012; Dastjerdi et al., 2016).

This simple definition does not encapsulate some key ingredients of the fog: ubiquity, enhanced network capabilities as a hosting environment and improved support for interaction between devices. Vaquero and Rodero-Merino (2014) provide another definition accordingly:

Definition 2: Fog computing is a scenario where a huge number of heterogeneous (wireless and sometimes autonomous) ubiquitous and decentralized devices communicate and potentially cooperate among them and with the network to perform storage and processing tasks without the intervention of third parties. These tasks can be for supporting basic network functions or new services and applications that run in a sandboxed environment. Users leasing part of their devices to host these services get incentives for doing so (Vaquero & Rodero-Merino, 2014).

This definition, nonetheless, is still debatable due to the fact that it fails to address the unique connection to the cloud (S. Yi et al., 2015). Thus a new definition is introduced:

Definition 3: Fog computing is a geographically distributed computing architecture with a resource pool consists of one or more ubiquitously connected heterogeneous devices (including edge devices) at the edge of network and not exclusively seamlessly backed by cloud services, to collaboratively provide elastic computation, storage and communication (and many other new services and tasks) in isolated environments to a large scale of clients in proximity (S. Yi et al., 2015).

It is believed that there will be more definitions of fog computing as it moves towards its maturity.

2.2 Characteristics of fog computing

Elastic resources including computation, storage and networking are the building blocks of both the cloud and the fog (S. Yi et al., 2015). Still, there are several unique characteristics of fog computing that make it a non-trivial extension of the cloud.

- Edge location: Due to close proximity to end users, the fog possesses the capabilities to support latency-sensitive applications that require real-time data processing.
- Location awareness: In stark contrast to the centralized cloud, the services delivered by the fog are widely distributed. The geographical distributed fog nodes have the ability to derive their locations and track end users’ devices in order to support mobility.
- Real-time interactions: Rather than bulk processing, fog applications involve real-time interactions.
- Edge analytics: In the era of big data, fog computing can support analyzing sensitive data locally instead of sending it to the cloud for analysis.
- Scalability: The cloud might become the bottleneck if the data generated by end devices are constantly transferred to it. Fog computing helps alleviate the burden of the centralized processing, thus addressing the scalability challenge stemming from the proliferation of end devices in the IoT (Dastjerdi et al., 2016).

3. SYSTEMATIC REVIEW PROTOCOL

After reviewing the literature on fog computing, it is found that there is no previously published SLR on this field. This might be due to the fact that fog computing is still an emerging paradigm. Therefore, the author
chose to conduct a SLR in an effort to attain an outline of the fog. This SLR is conducted using the reference manual adapted from Kitchenham (2007).

3.1 Research questions

This SLR aims at summarizing and obtaining the comprehension of fog computing with respect to its usage, applications and challenges. To this end, three research questions (RQs) were raised:

- RQ1: What are the current architectures/usage of fog computing?
- RQ2: Which applications can be of benefit from fog deployment?
- RQ3: What are the possible challenges with fog computing?

3.2 Search strategy

In this section, the search strategy for finding literature to answer the aforementioned RQs is presented.

3.2.1 Database

Popular literature resources are utilized for this SLR. The search process for this paper is based upon the following digital libraries:

- ACM Digital Library (http://dl.acm.org)
- IEEE Xplore Digital Library (http://ieeexplore.ieee.org)
- ScienceDirect (http://www.sciencedirect.com)
- Springer Link (http://link.springer.com)
- Wiley Online Library (http://onlinelibrary.wiley.com/)

3.2.2 Search term

The purpose of the search string is to capture all necessary keywords in the RQs together with their synonyms if applicable to obtain desired results. The list of keywords used to define the search string is: architecture, usage, application, challenge and issue. To generate the final search term, the two Boolean operators AND and OR are employed to connect the keywords together. A quotation mark is also used for exact text. As a result, the ultimate search string is:

“Fog computing” AND (architecture OR platform OR usage OR application OR challenge OR issue)

3.2.3 Search procedure

The process for choosing literature is as follows: first, the search string was executed in each selected database under the option of ‘Advanced search’; second, the titles and keywords of the results were read in order to filter out irrelevant papers; third, the inclusion and exclusion criteria were applied when analyzing the abstracts to further refine the results; finally, the remaining papers were read with full-text to ascertain their inclusion or exclusion. The entire process is summarized in Figure 2.
3.3 Study selection

After initial results were retrieved, impertinent papers were exempted by applying a set of inclusion and exclusion criteria as follows:

- Inclusion criteria:
  - Papers that explicitly discuss fog computing
  - Papers that mention and analyze an architecture for the fog
  - Papers that detail how fog computing are advantageous to an application or a domain
  - Papers that are focused on presenting present challenges with fog computing
  - Papers that are accessible
  - Papers that are from 2012 since fog computing is a relatively new paradigm

- Exclusion criteria:
  - Papers that do not have fog computing as a primary study
  - Papers that are primarily concerned with cloud computing
  - Papers that are inaccessible

3.4 Data extraction

In support of this stage, an Excel form was created as instructed in the guideline by Kitchenham (2007). The purpose of the form is to record the information attained in primary studies and to keep track of the data needed to answer the RQs. The collected information includes:

- The papers’ titles
- The source of the papers
- Is passing the title/keyword filter: This Yes/No field is used to indicate whether a paper satisfies the first filter.
- Is passing the abstract filter: This Yes/No field is used to indicate whether a paper satisfies the second filter.
- Is passing the full-text filter: This Yes/No field is used to indicate whether a paper satisfies the third filter.
- Is accessible:
- Notes: This field is used to store the detailed comments for a paper including the rationale as to why it is selected or exempted.

Those papers that have the properties for the three filter columns as Yes were then moved to a new sheet in order to extract the necessary information for addressing the RQs:
- Architectures: One or more architectures that are explained and discussed by a paper.
- Applications: One or more applications that are presented by a paper.
- Challenges: An array of issues pertinent to fog computing.

4. **SYSTEMATIC REVIEW EXECUTION**

The search procedure was performed on 28 March 2016 with a total initial number of papers being 176. 86 papers were then excluded after their titles and keywords were read. Of the 90 remaining publications, 67 were selected after the abstract were gone through. Finally, 48 papers were chosen as primary studies since 19 papers were exempted after their contents were analyzed in full length. The detailed figures for each phase are summarized in Table 1

<table>
<thead>
<tr>
<th>Library</th>
<th>Initial results without filtering</th>
<th>Title &amp; Keyword selected</th>
<th>Abstract selected</th>
<th>Full text selected</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACM</td>
<td>17</td>
<td>13</td>
<td>9</td>
<td>7</td>
</tr>
<tr>
<td>IEEE</td>
<td>71</td>
<td>64</td>
<td>50</td>
<td>33</td>
</tr>
<tr>
<td>ScienceDirect</td>
<td>30</td>
<td>3</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Springer</td>
<td>45</td>
<td>8</td>
<td>6</td>
<td>6</td>
</tr>
<tr>
<td>Wiley</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Results</td>
<td>176</td>
<td>90</td>
<td>67</td>
<td>48</td>
</tr>
</tbody>
</table>

*Table 1. Paper filtering phases*

5. **RESULTS AND FINDINGS**

5.1 Results

Of 67 publications chosen for full-text reading, 48 papers were considered as primary studies. They were then mapped into three aforementioned RQs. Table 2 shows the results of the final literature review.

<table>
<thead>
<tr>
<th>Research questions</th>
<th>Publications</th>
</tr>
</thead>
<tbody>
<tr>
<td>RQ1</td>
<td>(Aazam &amp; Huh, 2015b; Abdullahi, Arif, &amp; Hassan, 2015; Alam, Tun, &amp; Hong, 2016; Bittencourt, Lopes, Petri, &amp; Rana, 2015; Bonomi et al., 2012; Datta, Bonnet, &amp; Haerri, 2015; Dubey et al., 2015; Farris, Militano, Nitti, Atzori, &amp; Iera, 2015; Gazis et al., 2015; Giang, Blackstock, Lea, &amp; Leung, 2015; Hassan, Xiao, Wei, &amp; Chen, 2015; Hong, Lillethun, Ramachandran, Ottenwälder, &amp; Koldehofe, 2013; Ismail et al., 2015; Jayaraman et al., 2014; W. Lee, Nam, Roh, &amp; Kim, 2016; Ningning, Chao, Xingshuo, &amp; Qiang, 2016; Orsini, Bade, &amp; Lamersdorf, 2015; J. S. Preden et al., 2015; Sarkar &amp; Misra, 2016; Su, Lin, Zhou, &amp; Lu, 2015; Vallati et al., 2015; Vaquero &amp; Rodero-Merino, 2014; Xiang, Peng, Cheng, &amp; Chen, 2015)</td>
</tr>
</tbody>
</table>
The section that follows will summarize the findings to answer the RQs based on the literature.

5.2 Findings

5.2.1 RQ1: What are the current architectures/usage of fog computing?

There has been a number of proposed architectures for fog computing in recent years, which are mostly derived from the fundamental three-layer structure. It is observed that components needed and structure of the architectures are largely dependent on the purposes or the problems of the existing platform that need to be solved. The publications are grouped into six categories:

- **Resource management** (Aazam & Huh, 2015b; Chen, 2015; Gazis et al., 2015; Hong et al., 2013; Ismail et al., 2015; Ningning et al., 2016; Xiang et al., 2015)
- **Energy efficiency** (Dubey et al., 2015; Jayaraman et al., 2014; Sarkar & Misra, 2016; Vatanparvar & Al Faruque, 2015)
- **Offloading** (Alam et al., 2016; Bittencourt et al., 2015; Hassan et al., 2015; Orsini et al., 2015; J. Preden et al., 2015)
- **Data processing** (Datta et al., 2015; Giang et al., 2015)
- **Performance enhancement** (Abdullahi et al., 2015; Farris et al., 2015; Su et al., 2015)
- **Networking** (W. Lee et al., 2016; Vallati et al., 2015)

With reference to resource management, Ningning et al. (2016) constructed a system model of fog computing by incorporating the graph repartitioning theory and properties of the fog. Their future work seeks to improve the performance of the dynamic load balancing mechanism. Xiang et al. (2015) proposed a based radio access network (F-RAN) framework that aims at addressing the resource allocation platform in fog computing. For large-scale applications, Hong et al. (2013) constructed a mobile fog systems that serves two purposes: 1) provide a high-level programming model that streamlines development on a large number of heterogeneous devices across a wide area; 2) allow applications to dynamically scale based on their workload using on-demand resources.

As far as energy efficiency is concerned, Jayaraman et al. (2014) introduced the CARDAP platform (Context Aware Real-time Data Analytics Platform) for the fog, which includes energy efficient data delivery approaches in an attempt to reduce data transmission. The work by Sarkar and Misra (2016) modeled the fog computing architecture in support of the green aspect of it. Their observation showed that the service latency associated with the fog was greatly lower than that with the cloud. It was concluded that the fog is a greener computing platform.

One of the most distinguished characteristics of fog computing is the extension of the cloud to the edge of the network. Thus there has been several works seeking to implement the offloading capability. Alam et al. (2016) provided the Mobile Fog architecture to utilize mobile cloud computing and proposed the code offloading mechanism. Similar works can also be found in (Bittencourt et al., 2015; Hassan et al., 2015; Orsini et al., 2015; J. Preden et al., 2015).
Data handling is also a major concern in the big data era, and fog computing is a promising platform to resolve the information explosion in the future. The architecture by Giang et al. (2015) was expected to process the machine-to-machine (M2M) data with semantics, discovery and management of connected vehicles.

Caching is a common practice in enhancing performance of applications, which was embraced in (Abdullahi et al., 2015; Su et al., 2015). Abdullahi et al. (2015) introduced a conceptual framework taking Information Centric Networking cache at the nodes through fog computing.

5.2.2 RQ2: Which applications can be of benefit from fog deployment?

According to the literature, there is a wide variety of applications benefiting from the fog paradigm.

- **Healthcare** (Ahmad et al., 2016; Cao, Chen, et al., 2015; Cao, Hou, et al., 2015; Craciunescu et al., 2015; Fratu et al., 2015; J. S. Preden et al., 2015; Shi et al., 2015)
  
  Since the advent of the cloud, healthcare has been contemplated for the employment of this technology with an aim at improving its developments (Dzombeta, Stantchev, Colomo-Palacios, Brandis, & Haufe, 2014). The potential power of utilizing technological advancements for the health sector is further amplified with fog computing, which has gathered the most attention from the literature. Cao, Chen, et al. (2015) introduce FAST, a fog computing assisted distributed analytics system to monitor fall for stroke mitigation. They implemented fall detection algorithm and incorporated them into fog-based distributed fall detection system, U-Fall, leveraging both edge devices and data center services. Craciunescu et al. (2015) presented and implemented a fog computing system for e-Health applications. In their application, the fog nodes were installed in the home of the user in order to achieve the smallest processing time. Fratu et al. (2015) proposed the eWALL system that can offer solutions for the chronic obstructive pulmonary disease and Mild Dementia.

- **Smart grid** (Li et al., 2015; I. Stojmenovic & Wen, 2014; Tang et al., 2015)
  
  A smart grid is an electricity distribution network, with smart meters installed at different locations to measure real-time information (S. Yi et al., 2015). Fog computing can be significantly advantageous to smart grids. I. Stojmenovic (2014) studied issues such as security, demand response, privacy, fault tolerance in the context smart grids.

- **Smart vehicles** (Hou et al., 2016)
  
  Fog computing can be combined into vehicular networks. Vehicular fog computing can be categorized into two types according to S. Yi et al. (2015): infrastructure-based and autonomous. There are numerous applications for vehicular fog computing (VFC) irrespective of their types. These applications include traffic light scheduling, congestion mitigation, precaution sharing, parking facility management, traffic information sharing, etc. (S. Yi et al., 2015). Hou et al. (2016) presented their own implementation of VFC which utilized different end-user clients and edge devices to perform communication and computation.

- **Urgent computing** (Aazam & Huh, 2015a; Brzoza-Woch et al., 2015)
  
  Instant feedback and response are critical criteria in some applications such as disaster support. Brzoza-Woch et al. (2015) developed a flood decision support system embracing fog nodes in order to process acquired real data and trigger alarms in case of flood. Similarly, Aazam and Huh (2015a) presented E-HAMC (Emergency Help Alert Mobile Cloud) program, which attempts to respond promptly to a request of a user when there is an emergent situation.

- **Augmented reality** (Zao et al., 2014)
  
  Augmented reality applications require as low latency as possible since a very small delay can damage the user experience (Dastjerdi et al., 2016). Thus, fog computing has been embraced in this domain of late. Zao et al. (2014) developed Augmented Brain Computer Interaction Game based on fog computing and linked data. While a person is playing, raw streams of data by EEG sensors are generated and
categorized to detect the brain state of the player. Conventionally, brain state classification is sent to the central servers for processing, but it is a time-consuming task which can hinder the user experience with the game. The system leverages a combination of both fog and cloud servers, enabling continuous real-time processing and classification.

5.2.3 RQ3: What are the possible challenges with fog computing?

The publications reveal that there exists a number of potential limitations in the context of fog computing as follows:

- **System security**: Being close to end users, fog devices might encounter system security issues since they are usually installed out of strict surveillance and protection (Ivan Stojmenovic et al., 2015). One of the potentially malicious data hijacks is man-in-the-middle attack (K. Lee et al., 2015). In this attack, gateways serving as fog devices might be compromised or replaced by fraudulent ones (Wang et al., 2015). Encryption and decryption methods are proposed to resolve this problem.

- **Data protection**: Since fog computing is a distributed-based platform, data generated from IoT devices which are transferred to fog nodes and divided into smaller parts for storing. The integrity of the data should be guaranteed as a consequence. This can be achieved by using light-weight encryption algorithms or masking techniques (K. Lee et al., 2015).

- **Authentication and authorization**: Authentication and authorization issues were not studied in fog computing; instead, they were studied in the context of smart grids and machine-to-machine communications (Ivan Stojmenovic et al., 2015). When the connection between the fog and the cloud is fragile, the authentication challenge will arise as illustrated in the work of (Ivan Stojmenovic et al., 2015).

- **Quality of Service (QoS)**: This important metric for fog applications can be divided into four aspects, 1) connectivity, 2) reliability, 3) capacity, and 4) delay (Shanhe Yi, Li, et al., 2015).

6. DISCUSSION

Results reveal that the consideration for fog computing has grown noticeably in recent years. Still, the outcomes of those studies are largely experimental at the moment. They can come in the form of either a conceptual framework (Abdullahi et al., 2015) or models (Farris et al., 2015; Sarkar & Misra, 2016; Xiang et al., 2015) which are all theoretical. As far as experiments are concerned, there are different choices of hardware served as the fog computing layer. Dubey et al. (2015) employed an Intel® Edison which is a low-power device powered by a rechargeable lithium-ion battery while Aazam and Huh (2015a) made use of the CloudSim toolkit in their evaluation. CloudSim is a simulation toolkit that can be extended and allow simulation and modeling of cloud systems and application provisioning environments (Calheiros, Ranjan, Beloglazov, De Rose, & Buyya, 2011). Giang et al. (2015) utilized a TI SensorTag, two Raspberry Pis, a local PC server and an Amazon EC2 instance in order to simulate a fog architecture. It can be seen that these simulations are rather small-scale and solely for evaluation purposes. The challenges with fog computing setup are hardware availability and selection. Further, no emphasis has been placed in vastly simultaneous connections in the experiments. Fog computing emerged in an effort to address the exponential proliferation of connected devices in the IoT world. Therefore, this aspect should be scrutinized and tested thoroughly in future studies. However, the lack of supported infrastructure for fog computing has hindered the effort to evaluate the applications in a larger scale.

With respect to the issues of fog computing, the literature has identified several potential challenges facing the fog architecture and provided proposals accordingly. K. Lee et al. (2015) suggested adopting hybrid detection technique to detect malicious code in the fog layer. As far as unauthorized attacks are concerned, Wang et al. (2015) proposed approaches using decoy technology in an effort to detect illegitimate access by validating the authorization of accessed data and perplexing attackers. Intrusion detection is another mentioned technique which has been widely leveraged in cloud systems to tackle a number of types of attacks such as insider attack,
flood, user to root attacks, port scanning (Modi et al., 2013). In the scope of fog computing, there are potential opportunities to investigate how intrusion detection can be applied on both the centralized cloud servers and the client side. This is the study carried out by (Shanhe Yi, Qin, et al., 2015).

These proposals are, however, still theory-based solutions without specific implementation. It is thus suggested that further proposed architectures for the fog should take into account these issues such as security and privacy aspects.

7. CONCLUSION

In this paper, a SLR is presented with primary focuses on current research activities in fog computing including its architectures, applications and challenges. 48 of 176 publications were reviewed with reference to the three RQs and classified them based on their content. The results reveal that there have been several proposed architectures for fog computing in recent years which can be applied in a wide range of domains and applications. However, they are theoretical due to the fact that fog computing is still in its initial stage. This also poses a couple of issues to the fog concerning security, privacy and authentication, which needs further investigation and studies from researchers.

As future work, it would be helpful to know whether expanding the study's coverage with other academic databases can help provide insights into the advantages of fog computing to additional sectors and more possible challenges that this emerging computing paradigm might be faced with.

REFERENCES


