Chapter 17
Middleware for Preserving Privacy in Big Data

M. Thilagavathi
VIT University, India

Daphne Lopez
VIT University, India

B. Senthil Murugan
VIT University, India

ABSTRACT

With increased usage of IT solutions, a huge volume of data is generated from different sources like social networks, CRM, and healthcare applications, to name a few. The size of the data that is generated grows exponentially. As cloud computing provides an optimized, shared, and virtualized IT infrastructure, it is better to leverage the cloud services for storing and processing such Big Data. Securing the data is one of the major challenges in all the domains. Though security and privacy have been talked about for decades, there is still a growing need for high end methods for securing the rampant growth of data. The privacy of personal data, and to be more specific the health data, continues to be an important issue worldwide. Most of the health data in today's IT world is being computerized. A patient's health data may portray the different attributes such as his physical and mental health, its severity, financial status, and much more. Moreover, the medical data that are collected from the patients are being shared with other stakeholders of interest like doctors, insurance companies, pharmacies, researchers, and other health care providers. Individuals raise concern about the privacy of their health data in such a shared environment.

INTRODUCTION

Cloud computing paradigm amends the way information is processed and managed, especially when personal data processing is concerned. Cloud Computing is "a large-scale distributed computing paradigm that is driven by economies of scale, in which a pool of abstracted, virtualized, dynamically-scalable, managed computing power, storage, platforms, and services are delivered on demand to external customers over the Internet." The key characteristics of cloud computing is that the end-users can access cloud services anytime and anywhere without the need for any expert knowledge of the underlying technology, and this offers the benefit of reduction in cost as computing and storage resources are shared among multiple end-users. The services are provided on-demand based on pay-per-use business model. These new features have a direct impact on the IT budget and cost of ownership, but also bring up issues of traditional security, trust and privacy mechanisms.

Cloud computing offers three kinds of services to end-users i.e. Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS) and Software-as-a-Service (SaaS). IaaS is the delivery of computing infrastructure as a service on-demand. Infrastructure can be storage servers, applications, operating systems or any other computing resource. IaaS reduces cost as the set up and maintenance of the infrastructure is taken care of by the providers. Users pay for what they use and also help users achieve faster delivery time and service to market. Examples include Amazon Web Service, Flexiscale, OpenNebula, Nimbus, Enomaly (Bhaskar et al., 2009). PaaS provides a complete development environment where developers can develop, test, deploy or host their applications on the cloud. PaaS reduces the development time. Examples include Microsoft's Azure, Google App Engine (Bhaskar et al., 2009). SaaS provides software as a service that can be configured to suit the specific needs of the users. Example SalesForce.com (Bhaskar et al., 2009).

Clouds can be deployed in four different modes namely private cloud, community cloud, public cloud, and hybrid cloud. In a private cloud, the cloud infrastructure is used and managed within the organization thereby achieving high security. A community cloud is formed by a group of organizations that have similar requirements. With public cloud, the cloud infrastructure is made available to the public where resources can be dynamically accessed based on the requirements. A combination of two or more clouds (private, community, public) is termed as a hybrid cloud.

A huge volume of data is being generated from different sources like social networks, CRM, health care applications, sensors that collect climate information, cell phone GPS signals, etc. The size of the sensitive data that is generated thus grows exponentially and it is termed as Big Data. In general, big data refer to "datasets whose size is beyond the ability of typical database software tools to capture, store, manage, and analyze."

The aggregation and analysis of such large volumes of data can be used to map disease outbreaks, reduce frauds, improve business processes and assist in creating new innovative and wanted products. Such analysis might enable an organization to gain an insight about its customers and the market, so as to improve their businesses. However, there is also a dark side of Big Data, especially with privacy.

The extensive amount of personal information revealed through online transactions has taken the relationship between customer profiling, predicting trends and marketing to a whole other level. Big Data is capable of tracking movements, behaviors, preferences and predicting the behavior of individuals with unprecedented accuracy. The more the access a business has to Big Data the better they can target its customers with appropriate advertising and products that match specific interests. These data as collected cannot be released directly, because doing so could reveal data subjects' identities or values of sensitive attributes. Deteriorating to protect confidentiality

23 more pages are available in the full version of this document, which may be purchased using the "Add to Cart" button on the publisher's webpage: www.igi-global.com/chapter/middleware-for-preserving-privacy-in-big-data/103224

Related Content

IoT Limitations and Concerns Relative to 5G Architecture
Ohood Saud Althobaiti (2020). *Handbook of Research on Smart Technology Models for Business and Industry (pp. 45-69).*
www.irma-international.org/chapter/iot-limitations-and-concerns-relative-to-5g-architecture/259125

Fog Computing Quality of Experience: Review and Open Challenges
William Tichaona Vambe (2023). *International Journal of Fog Computing (pp. 1-16).*
www.irma-international.org/article/fog-computing-quality-of-experience/317110

An Intelligent Grid Network Based on Cloud Computing Infrastructures
Suresh Annamalai, Udendhran R.and Vimal S. (2019). *Novel Practices and Trends in Grid and Cloud Computing (pp. 59-73).*
www.irma-international.org/chapter/an-intelligent-grid-network-based-on-cloud-computing-infrastructures/230632

IoT-Based Smart Accident Detection and Alert System
C. V. Suresh Babu, Akshayah N. S., Maclin Vinola P.and R. Janapriyan (2023). *Handbook of Research on Deep Learning Techniques for Cloud-Based Industrial IoT (pp. 322-337).*
www.irma-international.org/chapter/iot-based-smart-accident-detection-and-alert-system/325950

From Cloud Computing to Fog Computing: Platforms for the Internet of Things (IoT)
Sanjay P. Ahujaand Niharika Deval (2018). *International Journal of Fog Computing (pp. 1-14).*
www.irma-international.org/article/from-cloud-computing-to-fog-computing/198409