J Supercomput (2013) 65:61–70
DOI 10.1007/s11227-011-0709-y

# Toward designing a secure biosurveillance cloud

**Tai-hoon Kim · Sabah Mohammed**

Published online: 14 October 2011


**Abstract** Biosurveillance is very complex, and it complements traditional public health surveillance to provide both early warning of infectious disease events and leads to situational awareness as well as to signaling any potential threat for using biological agents as weapons of mass destruction. Biosurveillance requires close cooperation and rapid information-sharing among many healthcare partners including primary care units and the biosurveillance hubs. Achieving improvements in this direction has become a bipartisan top priority for governments and institutions. Currently there are many national and international centers envisioned as clouds for intelligence on biological threats, however security obstacles have hindered their progress. This article investigates the requirements for a biosurveillance secure cloud. The investigation identifies the major security components needed to build a trusted environment for cloud based biosurveillance system through the integration of the public health enterprise private cloud with public clouds based on the Distributed OSGi framework along with a distributed authentication service. The trusted environment allows biosurveillance to be conducted over primary care private clouds including patient information from the electronic medical records.

**Keywords** Cloud computing security · Biosurveillance · Distributed OSGi

T.-h. Kim (✉)
GVSA and University of Tasmania, Burnie, Australia
e-mail: taihoonn@empal.com

S. Mohammed
Department of Computer Science, Lakehead University, 955 Oliver Road, Thunder Bay, Ontario P7B 5E1, Canada
e-mail: Sabah.mohammed@lakeheadu.ca

## 1 Introduction

There are many challenges and expected shifts in healthcare, including: the use of electronic health records; the increased use of health surveillance systems; higher patient privacy and the availability of clinical data standards that are accepted by the norm. When it comes to disease surveillance (also known as syndromic surveillance), health information technology (HIT) can play a crucial role in quickly detecting initial disease outbreak, epidemic, or bioterror attack before confirmed diagnosis can be made. As we move into the 21st century, new technological tools and methods are becoming available; the epidemiology of diseases is changing; new diseases are emerging and old ones re-emerging. Moreover, budgets are shrinking and workers are expected to respond to the new challenges and to the increasing societal expectations with little in the way new resources. To address all of these projected shifts and challenges many research groups, venders and institutions have established cloud or fusion hubs for biosurveillance capable of delivering effective dynamic resourcing to the enterprise cloud by offering reliable, mature, and a high performance mechanism combined with some sort of security and management. These resources, offered as a Platform as a Service (PAAS), helps to build variety of cloud networks easily without having to buy and set up an infrastructure. The sophistication of the hubs varies from extremely complex (e.g. US National Biosurveillance Integration Center's (NBIC),[1] Europe Global Health Security Action Group (GHSAG)[2]) to moderately complex (e.g. Oracle Fusion Platform,[3] caGrid 1.0,[4] eDSS [9], BioSTORM [20], AEGIS [22], Panorama [11], eTriage [14], ESSENCE [13], RODS [7] (http://openrods.sourceforge.net/), and BioSense [3]). Based on such infrastructures more and more data are being captured around healthcare processes in the form of Electronic Medical Records (EMR), health insurance claims, medical imaging databases, disease registries, spontaneous reporting sites, and clinical trials. As these data get collected, government regulations (e.g. US CCHIT ONC-ATCB certification[5] and HITSP IS 02)[6] are requiring healthcare providers to not only store it in an electronic format but also to securely use it in meaningful ways. Actually, several notable reports (e.g. Naylor [18], Walker [24], Campbell [5]) highlighted the post-SARS disease surveillance needs and recommended to invest in a "*seamless public health system that will allow public health professionals to coordinate activities in a carefully planned infrastructure*". The most important features of such a system are to be able to scale across data sources and analytical methods as well as be extended to meet the needs of different end-users. The new trend of cloud computing seems to qualify for such infrastructure that is able to gather health information from different streams (including EHRs systems) and be able to provide surveillance

[1] http://www.dhs.gov/xnews/testimony/testimony_1191608625983.shtm.

[2] http://ec.europa.eu/health/archive/ph_threats/com/preparedness/docs/ev_ghsag_2006.pdf.

[3] http://www.oracle.com/us/products/applications/fusion/index.html.

[4] https://cabig.nci.nih.gov/workspaces/Architecture/caGrid1-0.

[5] http://www.cchit.org/.

[6] http://www.hitsp.org/.

information to local public health officials in real time as well as allowing the public health official to push requests for additional required data elements (patient medications, allergies, tests, etc.) down to the physician to test a hypothesis or to alert the region for any disease outbreak. While low start-up deployment costs, scalability, ease of access, and immediate time to market are undeniable benefits of cloud computing, there are number of security concerns that are holding healthcare enterprises back from migrating their infrastructures to the cloud. The obvious reality is that the cloud is still evolving, and solution providers battling for customer attention are developing and deploying more sophisticated security measures to ensure best security and privacy practices and certifications (including previledges like access controls, emergency access, automatic log-off, audit log, data integrity, authentication, general encryption, and encryption when exchanging electronic health information). The security and privacy criteria must be met by all certified EHRs, regardless of whether they are housed on a local cloud or are based on the web or public cloud.

## 2 Biosurveillance cloud security

There are many security risks that are a potential threat to institutions moving to a cloud environment [9]:

- user access to data and information
- location of the data
- the encryption used at every level
- recovery measures in the event of a security breach.

However, such security risks involve only three types of concern: Confidentiality, Trust and Availability (CTA) [2]. Figure 1 lists these risks under the three security concerns.

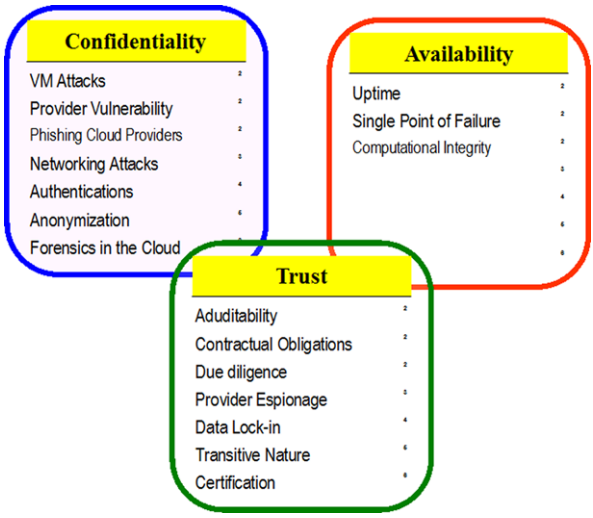


**Fig. 1** Cloud security concerns

Confidentiality concerns involve computer and network attacks that will be made possible by moving to the cloud. Cloud providers respond to these concerns by arguing that their security measures and processes are more mature and tested than those of the average company [6]. The availability concerns center on critical applications and data being available. But new solutions like the Nasuni Filer[7] may help ease some of those concerns. Nasuni can replace or augment the institution current file backup and disaster-recovery schemes. Related to the third party trust and data control, the legal implications of data and applications being held by a third party are complex. Such complexity prompts some institutions to build private/local clouds for their highly secure data and yet retain some of the advantages of cloud [8]. To ensure applicability of the CTA factors, the service provider should offer tested encryption schema, stringent access controls and scheduled data backups [12]. In this direction, a large group of the computing industry established in 1999 a Trusted Computing Platform Alliance (TCPA)—which has been named later as the Trusted Computing Group (TCG)[8]—that focused on building confidence and trust of computing platform in e-business transactions. The distinguishing feature of TCG technology is arguably the incorporation of "roots of trust and confidentiality" into computer platforms. The TCG advocates for the development of trusted computing (TC) systems that integrate data security mechanism into the platform core operations, rather than implementing it by using add-on applications. In this concept, TC systems would cryptographically seal off the parts of the computer that deal with data and applications and give decryption keys only to programs and information that the technology judges to be trusted. The TCG made this mechanism as their core criterion to define the technology specification [23]. Actually, trusted computing is a broad term that refers to technologies and proposals for resolving computer security problems through hardware enhancements and associated software modifications. According to Microsoft[9] trusted computing is defined by four technologies, all of which require the use of new or improved hardware at the personal computer (PC) level:

- Memory curtaining—prevents programs from inappropriately reading from or writing to each other's memory.
- Secure input/output (I/O)—addresses threats from spyware such as keyloggers and programs that capture the contents of a display.
- Sealed storage—allows computers to securely store encryption keys and other critical data.
- Remote attestation—detects unauthorized changes to software by generating encrypted certificates for all applications on a PC.

In order to be effective, these measures must be supported by advances and refinements in the software and operating systems (OSs) that PCs use. The trusted computing features just described will add new capabilities to the PC. To be used, they must be supported by software; in the absence of trusted computing software drivers, the trusted computing PC is just an ordinary PC, which remains capable of

[7] http://www.nasuni.com/product/product-overview/.

[8] http://en.wikipedia.org/wiki/Trusted_Computing_Group.

[9] http://searchsecurity.techtarget.com/sDefinition/0,,sid14_gci1232073,00.html.

running all existing PC software. To put it in another way, having TC features with any biosurveillance system will add to the system's limitations. This only means that the compelling case for the security of biosurveillance systems has yet to be made and the only source of security that a biosurveillance system customer can have is through a transparent SLA agreement and protocol with the cloud provider. The SLA is the only legal agreement between the service provider and client which provides guidelines on the following issues [1]:

- Services to be delivered, performance
- Tracking and Reporting
- Problem Management
- Legal Compliance
- Resolution of Disputes Customer Duties
- Security responsibility
- Confidential Information Termination.

## 3 Identifying components for a secure biosurveillance system

As public health and medicine proceed in our information age, the use of existing electronic data for public health surveillance will not appear to be an untested experiment for long. The challenge is to allow these systems to flower without burdening them with unrealistic expectations, centralized control, and unbalanced funding [17]. To help syndromic surveillance systems reach their full potential, we need to identify security components and guidance to the designers and developers that will ensure data flow and interoperability to respond when the alarm sounds. First we need to identify the basic cloud services that can affect security issues. Figure 2 illustrates the main components of the standardized cloud reference model [25]. It is clear that the most important layers are the Front End Client/Browser Services and the cloud specific infrastructure that constitutes the heart of cloud vulnerabilities. It is within these two layers that one can start incorporating the major security components.

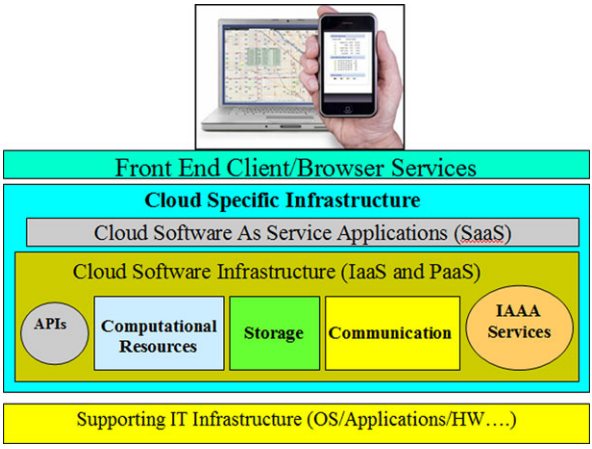


**Fig. 2** The standard cloud reference model

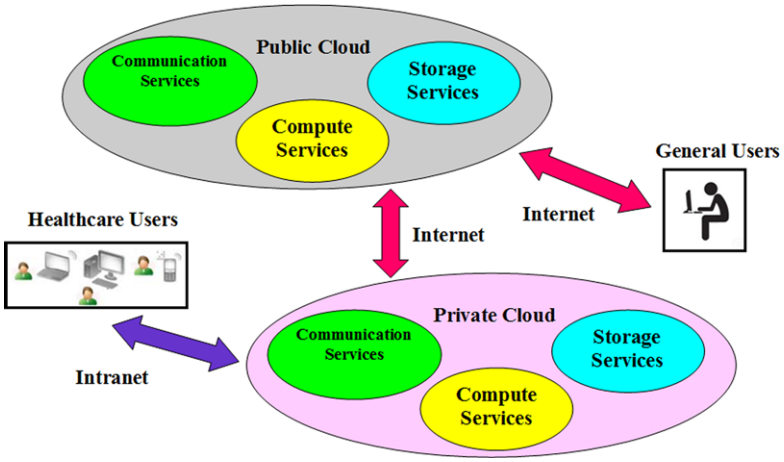


**Fig. 3** A biosureveillance cloud configuration

For biosurveillance we also need to understand that healthcare data are very sensitive and most healthcare enterprises prefer to keep this information on a private cloud. A private cloud maintains all corporate data in resources under the control of the legal and contractual umbrella of the organization. This eliminates the regulatory, legal and security concerns associated with information being processed on third party computing resources. However, a private biosurveillance cloud needs to collaborate with the other public clouds for data exchange and other services (Fig. 3).

According to Michael Brock and Andrzej Goscinski [4] the flow of communication messages between the private and public clouds can be represented in a general framework as is illustrated in Fig. 4.

There are many options for a gateway server like the OracleAS[10] and the IBM WebSphere-RPSS.[11] However, the client front end needs to be able to communicate with the gateway server. Actually, since a cloud aims at providing computing resources as a service in a scalable way, it should be possible to instantly deploy or un-deploy applications within a large set of nodes including the client front end. For this purpose we need a generic deploy a generic framework that serves cloud scalability. In this direction we have very limited options and the best one seems to be based on the Distributed OSGi[12] framework. The only restriction for the use of the Distributed OSGI framework is that every client and server machines needs to run Java Virtual Machine (JVM). Actually scalability of the cloud based on the Distributed OSGi framework stems from the notion of bundles. Bundles provide flexibility and enable scalability. There are three types of such bundle:

- **Single bundle**. It loads from local file system and installed in local OSGi framework.

[10] http://www.oracle.com/technetwork/middleware/ias/overview/index.html.

[11] http://www.ibm.com/developerworks/library/it-expertprog_tpss/index.html.

[12] http://cxf.apache.org/distributed-osgi.html.

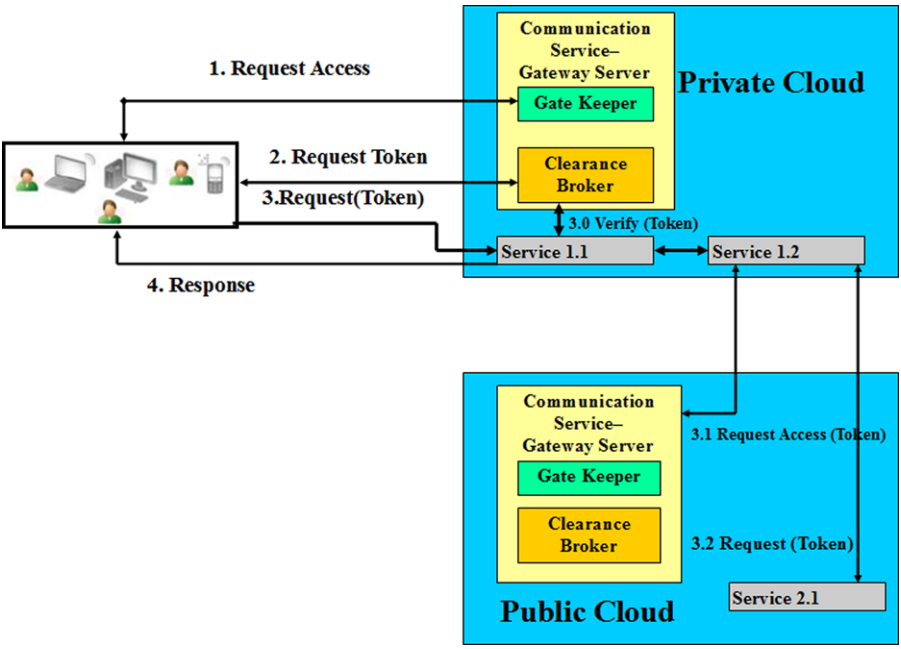


**Fig. 4** Biosureveillance private and public clouds communication

- **Multiple bundle**. It loads not only from local file system but from remote bundle library server. It works as the event receiver and event sender between local OSGi framework and remote OSGi framework.
- **Remote bundle**. Remote bundle can be downloaded from remote bundle library server to local file system. Then installed in a remote OSGi framework. Remote OSGi framework will maintain the dependency between remote bundles.

Biosurveillance management is based on deploying OSGi bunldes.The OSGi bundles collect data and send report to the distributed OSGi administration server for analysis and final reporting (Fig. 5). The gateway server starts and initializes machine instances for biosurveillance services and the service registry. At the same time, the cloud administration servers register their service with the service registry. The service consumer queries service registry for a listing of available biosurveillance services. Finally the service consumer sends a request to view or update health information as well as to receive an appropriate response. *Dotted lines* indicate interactions transparent to the service consumer.

This type of configuration based on the Distributed OSGi framework has been reported in our earlier research article [15]. However, since healthcare users and other general users need to access the various types of cloud including the private cloud, the biosurveillance framework requires a component for access control management that has the flexibility and scalability based upon the design principle of third party certificate authority and needs auditing which includes stored data. Table 1 provide a list of some of the notable third party certification authorities.

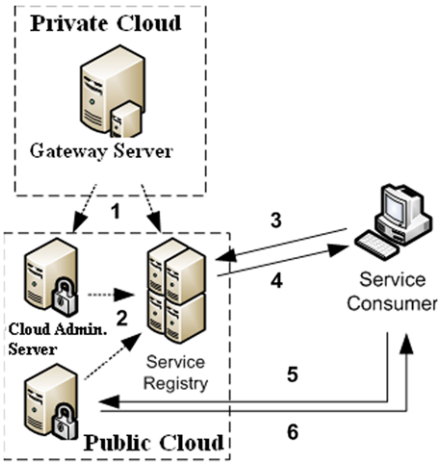


**Fig. 5** A distributed OSGi biosurveillance system

**Table 1** Third party certificate authorities

| |
|---|
| **VeriSign**—http://www.verisign.com/ |
| **Thawte Digital Certificates**—http://www.thawte.com/ |
| **GlobalSign**—http://www.globalsign.net |
| **Entrust.net**—http://www.entrust.net/ |
| **Digicert**—http://www.digicert.com/ |
| **GeoTrust**—http://www.geotrust.com |
| **Comodo**—http://www.comodogroup.com |
| **Trustwave**—https://www.trustwave.com/ |
| **Rapid SSL**—http://www.rapidssl.com |
| **CAcert**—http://www.cacert.org/ |
| **Network Solutions: SSL Certificates**—http://www.networksolutions.com/SSL-certificates/index.jsp |
| **SSL.com**—http://www.ssl.com |
| **LiteSSL**—http://www.litessl.com |
| **Enterprise SSL**—http://www.enterprisessl.com |
| **The USERTRUST Network**—http://www.usertrust.com/ |
| **Digital Signature Trust Co.**—http://www.digsigtrust.com/ |
| **ProntoSSL**—http://www.prontossl.com |
| **Pink Roccade PKI**—http://www.pki.pinkroccade.com/ |
| **Ebizid**—http://www.ebizid.com |
| **SimpleAuthority certification authority**—http://simpleauthority.com |
| **QualitySSL**—http://www.qualityssl.com |
| **Digi-Sign**—http://www.digi-sign.com |
| **Secure SSL**—http://www.securessl.co.uk |
| **SSL Certificate Management Site**—http://ssl4net.com/ |

The access control management component needs to work with the distributed nature of the cloud. Actually the traditional authentication methods are not appropriate

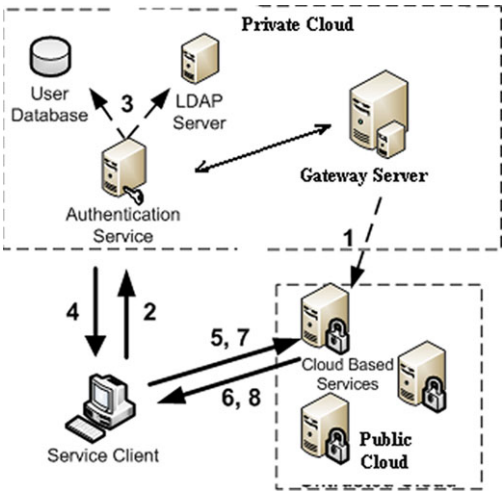


**Fig. 6** The RBSSO distributed authentication protocol

as they would require duplication of authentication mechanism and user databases or the creation of a single point failure resulting in a bottleneck of the system. In [16] we reported a solution for a single-sign authentication service that works for the Distributed OSGi framework, in which users first authenticate with a trusted party to receive an authentication token that enables access to services that trust the same party. Several technologies currently exist which enable single-sign on capabilities, such as Kerberos [19], SAML [21], and X.509 [10]. However, the nature of the cloud and the architecture of HCX make traditional solutions complicated as new machine instances are spawned and destroyed automatically based on demand and have no persistent memory to store public/private key pairs or certificates. The new distributed authentication service is called RBSSO (Roll Based Single-Sign On). RBSSO is loosely based on X.509 single-sign on and aims to minimize the number of request on an authentication server, support a large number of authentication methods, supports sessions spanning multiple services and be relatively easy to implement and understand. Figure 6 illustrates the RBSSO protocol and the full details of the protocol can be found in [16].

## 4 Conclusions

In this paper we utilize cloud computing as a low-cost computational platform to address critical challenges and limitations in today's surveillance systems and to introduce design concepts to facilitate the security in such systems. We identified several security components to build a trusted environment for cloud based biosurveillance system by integrating the legacy health enterprises private clouds with the public clouds based on the Distributed OSGi framework along with a distributed authentication service. The integral trusted environment will enable biosurveillance to perform over primary care private clouds and hence provide wider and more effective public health alerts and warnings. We are currently evaluating our security components including authentication, confidentiality and integrity within a primary health-

care scenario that utilizes electronic healthcare records as the main source of data surveillance.

## References


1. Balachandra RK, Ramakrishna PV, Rakshit A (2009) Cloud security issues. In: 2009 IEEE international conference on services computing, 26 October 2009, pp 517–520
2. Bertino E et al (2006) Secure knowledge management: confidentiality, trust, and privacy. IEEE Trans Syst Man Cybern, Part A, Syst Humans 36(3)
3. Bradley CA et al (2005) BioSense: implementation of a national early event detection and situational awareness system. Morb Mortal Wkly Rep, CDC Surveill Summ 54(Suppl):11–20
4. Brock M, Goscinski A (2010) Toward a framework for cloud security. In: Algorithms and architectures for parallel processing. LNCS, vol 6082/2010. Springer, Berlin, pp 254–263
5. Campbell A (2006) The SARS Commission. Final Report, Toronto, Ontario Ministry of Health and Long-Term Care. Available Online: www.health.gov.on.ca/
6. Chow R et al (2009) Controlling data in the cloud: outsourcing computation without outsourcing control. In: ACM CCSW'09 conference, Chicago, Illinois, USA, November 13, 2009
7. Espino JU et al (2004) The RODS open source project: removing a barrier to syndromic surveillance. Stud Health Technol Inf 107(Pt 2):1192–1196
8. Hang C, Can C (2010) Research and application of distributed OSGi for cloud computing. In: Int conference on computational intelligence and software engineering CiSE 2010
9. Heiser J, Nicolett M (2008) Assessing the security risks of cloud computing. Gartner Association, ID Number: G00157782
10. Housley R et al (2002) Internet X.509 public key infrastructure certificate and Certificate Revocation List (CRL) Profile. RFC3280
11. Panorama Project (2007) IBM Pan-Canadian public health communicable disease surveillance and management project
12. Kaufman L (2009) Data security in the world of cloud computing. IEEE Secur Priv 7(4):61–64
13. Lombardo L, Burkom H, Pavlin J (2004) ESSENCE II and the framework for evaluating syndromic surveillance systems. CDC 53(Suppl):159–165
14. McDonald L et al (2007) Evaluation of a systematic emergency department chief complaint system for near real-time public health surveillance. J Adv Dis Surveill 2:206
15. Mohammed S, Servos D, Fiaidhi J (2010) HCX: a distributed OSGi based web interaction system for sharing health records in the cloud. In: 2010 IEEEWICACM international conference on web intelligence and intelligent agent technology
16. Mohammed S, Servos D, Fiaidhi J (2011) Developing a secure distributed OSGI cloud computing infrastructure for sharing health records. In: AIS 2011—international conference on autonomous and intelligent systems, Burnaby, BC, Canada, 22–24 June 2011
17. Mostashari F, Hartman J (2003) Syndromic surveillance. J Urban Health Bull NY Acad Med 80(2), Suppl (1)
18. Naylor D (2003) National Advisory Committee on SARS and public health learning from SARS. Health Canada, Ottawa
19. Neuman BC et al (1994) Kerberos: an authentication service for computer networks. IEEE Commun Mag 32:33–38
20. Nyulas C et al (2008) An ontology-driven framework for deploying JADE agent systems. In: IEEE/WIC/ACM int conf on web intelligence and intelligent agent technology (WI-IAT'08)
21. OASIS Open (2009) Assertions and protocols for the OASIS security assertion markup language (SAML) V2.0—Errata Composite, December 2009
22. Reis B et al (2007) AEGIS: a robust and scalable real-time public health surveillance system. J Am Med Inform Assoc 14(5):581–588
23. Shen Z, Tong Q (2010) The security of cloud computing system enabled by trusted computing technology. In: 2nd International conference on signal processing systems (ICSPS)
24. Walker D (2003) Ontario expert panel on SARS. Ministry of Health, Toronto
25. Youseff L, Butrico M, Da Silva D (2008) Towards a unified ontology of cloud computing. In: IEEE Proc grid computing environments workshop (GCE)