1. General Cloud Environments

As implied by the NIST cloud computing definition, a cloud system is a collection of network-accessible computing resources that customers (i.e., cloud subscribers) can access over a network. In general terms, a cloud system and its subscribers employ the client-server model [Com88], which means that subscribers (the clients) send messages over a network to server computers, which then perform work in response to the messages received.

Figure 1 gives a general view of a cloud and its clients: the cloud's computing resources are depicted as a grid of computer systems where clients access a cloud over network connections. As shown in the figure, new clients may arrive, existing clients may depart, and the number of clients using a cloud at any one time is variable. Similarly, a cloud maintains a pool of hardware resources that it manages to maximize service and minimize costs. To maintain highly available services despite expected component failures and service life expirations, a cloud incorporates new hardware components as needed and retires old or failing components. To provide services cost-effectively, a cloud will manage the pool of hardware resources for resource efficiency; one of the strategies that a cloud provider employs during periods of reduced subscriber demand is to power off unused components. Whether for power management, or for hardware refresh, migration of customer workloads (data storage and processing) from one physical computer to another physical computer [Chr05, Shr10, VMw11, Mic10, Red99] is a key strategy that allows a provider to refresh hardware or consolidate workloads without inconveniencing subscribers.

Figure 1: Generic Cloud and Cloud Subscriber View

From Figure 1, a small number of general statements about cloud computing (e.g., strengths and limitations, performance characteristics, etc.) can be inferred; organizations considering the use of cloud computing should consider these general statements (listed below). Many statements commonly made about clouds (e.g., that clouds scale for very large workloads or that clouds replace capital expenses with operational expenses), however, are true only for certain types of clouds. To avoid confusion, this document explicitly qualifies each such statement with the type of cloud to which it applies; i.e., each statement has a "scope." The scopes used in this document are:

<table>
<thead>
<tr>
<th>Scope Name</th>
<th>Applicability</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
**general** | Applies to all cloud deployment models.
---|---
on-site-private | Applies to private clouds implemented at a customer's premises.
outsourced-private | Applies to private clouds where the server side is outsourced to a hosting company.
on-site-community | Applies to community clouds implemented on the premises of the customers comprising a community cloud.
outsourced-community | Applies to community clouds where the server side is outsourced to a hosting company.
public | Applies to public clouds.

<table>
<thead>
<tr>
<th><strong>Table 1: Scope Modifiers for Statements Asserted about Clouds</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>Each of the scopes are explained in detail below. To add clarity, this document also uses three terms consistently:</td>
</tr>
<tr>
<td><strong>cloud subscriber</strong> or <strong>subscriber</strong>: a person or organization that is a customer of a cloud;</td>
</tr>
<tr>
<td><strong>client</strong>: a machine or software application that accesses a cloud over a network connection; and</td>
</tr>
<tr>
<td><strong>cloud provider</strong> or <strong>provider</strong>: an organization that provides cloud services.</td>
</tr>
<tr>
<td>The following statements are general in their scope, i.e., they apply regardless of the deployment model or service model:</td>
</tr>
<tr>
<td><strong>Network dependency (general)</strong>. The subscribers, being clients, need a working and secure network to access a cloud. If the network is not reliable, the cloud will not be reliable from the subscriber's point of view.</td>
</tr>
<tr>
<td><strong>Subscribers still need IT skills (general)</strong>. By operating the server computers, a provider may reduce the need for IT staff in subscriber organizations, but subscribers will still access the cloud from on-site subscriber-managed client systems that must be maintained, secure, etc.</td>
</tr>
<tr>
<td><strong>Workload locations are hidden from clients (general)</strong>. To manage a cloud's hardware resources, providers must be able to migrate subscriber workloads between machines without inconveniencing the clients, i.e., without the clients being required to track and adapt to changes and without the clients being aware.</td>
</tr>
<tr>
<td><strong>Risks from multi-tenancy (general)</strong>. The workloads of different clients may reside concurrently on the same system and local network, separated only by access policies implemented by a provider's software. A flaw in the software or flaw in the policies could compromise the security of subscribers.</td>
</tr>
<tr>
<td><strong>Data import/export, and performance limitations (general)</strong>. Because subscribers access a cloud over a network, on-demand bulk data import or export may exceed the network's ability to carry the data in a timely manner. Additionally, real-time or critical processing may be problematic because of networking limitations.</td>
</tr>
<tr>
<td>Organizations contemplating the use of cloud computing should consider these general statements and their possible consequences for an organization's mission and business model. Considering only the general statements, however, is not sufficient. Clouds will also be described by one or more of the other (i.e., not &quot;general&quot;) scopes listed in Table 1; organizations contemplating the use of cloud computing should consider the detailed statements made for the kinds of clouds they contemplate using. Each of the alternatives is broken out in a separate section below.</td>
</tr>
</tbody>
</table>

---

1 In some cases (e.g., the IaaS service model described in Section Error! Reference source not found. below) a workload may exist in a particular location for a specific time before it migrates; in other cases (e.g., for the PaaS service model described in Section Error! Reference source not found. below) a workload may exist as a fundamentally distributed entity with sequential operations performed for a subscriber potentially executing in different servers, and data existing in a geographically distributed data store.
1.1 Understanding Who Controls Resources in a Cloud

It is sometimes asserted that when compared to traditional on-premise computing, cloud computing requires subscribers to give up (to providers) two important capabilities:

control: the ability to decide, with high confidence, who and what is allowed to access subscriber data and programs, and the ability to perform actions (such as erasing data or disconnecting a network) with high confidence both that the actions have been taken and that no additional actions were taken that would subvert the subscriber's intent (e.g., a subscriber request to erase a data object should not be subverted by the silent generation of a copy).

visibility: the ability to monitor, with high confidence, the status of a subscriber's data and programs and how subscriber data and programs are being accessed by others.

The extent, however, to which subscribers may need to relinquish control or visibility depends on a number of factors including physical possession and the ability to configure (with high confidence) protective access boundary mechanisms around a subscriber’s computing resources.

This document uses the concept of access boundaries to organize and characterize the different cloud deployment models. Figure 2 illustrates a key concept from computer security relating to boundaries and control, the security perimeter [TIS94, Gas88]. As shown in the figure, a security perimeter is a barrier to access: entities that are inside the perimeter may freely access resources inside the perimeter, however entities that are located outside the perimeter may only access the resources inside if allowed by a boundary controller that enforces a policy over access.

Typical boundary controllers include firewalls [TIS94, Che94], guards [Eps99], and Virtual Private Networks [Ros99]. By implementing a security perimeter around its important resources, an organization can achieve both a measure of control over the use of those resources and a means for monitoring access to them. Furthermore, via reconfiguration, an organization can adapt a security perimeter to changing needs (e.g., blocking or allowing protocols or data formats based on changing business circumstances). The various cloud deployment models in the NIST cloud definition have implications for the locations of subscriber-controlled security perimeters and hence for the level of control that subscribers can exercise over resources that they entrust to a cloud.

---

2 When uncontrolled paths to computing resources exist, a security perimeter is weakened or may not even exist. Pervasive wireless communications, e.g., are a threat to security perimeters since there may be no reliable way to interpose a boundary controller between external entities and internal entities. Similarly, many organizations use mobile devices that are sometimes connected within an organization's security perimeter, and sometimes exposed directly, e.g., when on travel.
The NIST cloud definition lists four deployment models: private, community, public, and hybrid. The private deployment model, however, admits of two variants that should be discussed separately because they affect the security perimeter: on-site, and outsourced. The hybrid deployment model is a combination of the others and therefore a hybrid deployment may be subject to the implications of all of its building blocks as well as unique implications that arise when multiple systems are composed into more complex integrated systems.

1.2 The On-site Private Cloud Scenario

Figure 3 presents a simple view of an on-site private cloud. As shown in the figure, the security perimeter extends around both the subscriber's on-site resources and the private cloud's resources. The private cloud may be centralized at a single subscriber site or may be distributed over several subscriber sites. The security perimeter will exist only if the subscriber implements it. If implemented, the security perimeter will not guarantee control over the private cloud's resources, but its existence gives an opportunity for a subscriber to exercise control over resources entrusted to the on-site private cloud.

Although the general implications remain true with an on-site private cloud, the on-site-private scenario allows for additional and more detailed implications that organizations considering the use of an on-site private cloud should consider:

**Network dependency (on-site-private).** Depending on the configuration (e.g., single physical site, protected cloud network), the network dependency for an on-site private cloud may be limited to dependence on networking resources over which a subscriber has control.
(e.g., local area networking). In this scenario, larger-scale network problems, such as Internet congestion or communications with remote Internet Domain Name Servers (DNS) [Moc87-1, Moc87-2] may be avoided. Additionally, if the security perimeter has been implemented for high security levels, application of security mechanisms such as multi-factor authentication and end-to-end encryption within the secure enclave, while still prudent policies, may not be necessary in all cases.

If a subscriber organization spans multiple physical sites and wishes different sites to access the same private cloud, however, the subscriber must either provision a controlled inter-site communications media, such as an encrypted leased line, or must use cryptography (e.g., with a VPN) over less controlled communications media such as the public Internet. Both of these options introduce risks to a private cloud's networking availability and security because performance dependencies are established to resources that exist off of the subscriber's site and that are not directly under the subscriber's control, and because any failure to implement and configure cryptographic mechanisms could allow outsiders access.

**Subscribers still need IT skills (on-site-private).** Subscriber organizations will need the traditional IT skills required to manage user devices that access the private cloud, and will require cloud IT skills as well. Early in the rollout of an on-site private cloud, subscriber organizations may wish to maintain parallel cloud and non-cloud operations for an evaluation period. During any such evaluation period, traditional IT skills will be required. Even after an evaluation period, however, traditional IT staff may be needed (perhaps at reduced levels) to manage legacy licensing agreements, special hardware or system requirements, unique security needs for special projects, and legacy investments in equipment and training.

In addition, new skills in working with a cloud may be required. For example, an organization that performs compute-intensive jobs may need to eventually reorganize those jobs so that they can run using a higher level of parallelism on the cloud's resources [Dea04]; an organization that processes large data sets in the cloud will need to develop skills with cloud-based storage [Cha06, Ghe03, Ama06, SNI10, Msf11].

**Workload locations are hidden from clients (on-site-private).** As in the general case, to manage a cloud's hardware resources, a private cloud must be able to migrate workloads between machines without inconveniencing clients, i.e., without the clients being aware. With an on-site private cloud, however, a subscriber organization chooses the physical infrastructure in which the private cloud operates, and hence determines the possible geographical locations of workloads. While individual clients still may not know where their workloads physically exist within the subscriber organization's infrastructure at any given time, the subscriber organization has both visibility and control over where workloads are allowed to reside.

**Risks from multi-tenancy (on-site-private).** As in the general case, the workloads of different clients may reside concurrently on the same systems and local networks, separated only by access policies implemented by a cloud provider's software. A flaw in the software or the policies could compromise the security of a subscriber organization by exposing client workloads to one another contrary to the subscriber's security policy. An on-site private cloud mitigates these risks somewhat by restricting the number of possible attackers; all of the clients would typically be members of the subscriber organization or authorized guests or partners, but the on-site private cloud is still vulnerable to attack conducted by authorized but also malicious insiders. Different organizational functions, such as payroll, storage of sensitive information, or the generation of intellectual property

---

3 Note: not a comprehensive list of cloud storage systems.
may be merged as a consequence of such security failures, which can provide access to
users who are not authorized to access specific classes of data and who then may disclose
data from the on-site private cloud. Such failures may also allow low-priority clients to
interrupt higher-priority clients.

**Data import/export, and performance limitations (on-site-private).** As with the general case,
on-demand bulk data import/export is limited by the on-site private cloud's network
capacity, and real-time or critical processing may be problematic because of networking
limitations. In the on-site private cloud scenario, however, these limits may be adjusted,
although not eliminated, by provisioning high-performance and/or high-reliability
networking within the subscriber's infrastructure. Particularly if a subscriber has only
one site that requires access to the on-site private cloud, a subscriber may be able to
provision local networks that provide higher performance than can practically be
achieved via wide area networks.

**Potentially strong security from external threats (on-site-private).** In an on-site private cloud,
a subscriber has the option of implementing an appropriately strong security perimeter to
protect private cloud resources against external threats to the same level of security as can
be achieved for non-cloud resources. For low-impact data and processing, the security
perimeter may consist of commercial firewall rule sets and VPNs. For higher-impact
data, security perimeters can be constructed via more restrictive firewall policies [Zwi00,
Ran99], multi-factor authentication [SP-800-63], encryption [Sch94, Ros99], and even
physical isolation.

**Significant-to-high up-front costs to migrate into the cloud (on-site-private).** An on-site
private cloud requires that cloud management software be installed on computer systems
within a subscriber organization. If the cloud is intended to support process-intensive or
data-intensive workloads, the software will need to be installed on numerous commodity
systems or on a more limited number of high-performance systems. Installing cloud
software and managing the installations will incur significant up-front costs, even if the
cloud software itself is free, and even if much of the hardware already exists within a
subscriber organization. Three potential approaches to accomplish this are:

- **New Data Center:** The most direct approach is for a subscriber to provision a data center
  in which to deploy the cloud software. In this case, the on-site private cloud incurs up-
  front costs that are similar to those of a typical data center and the subscriber can
  provision the data center for anticipated workloads.

- **Converted Data Center:** As an alternative to provisioning a new data center, a subscriber
  may convert part or all of an existing data center to support the on-site private cloud. This
  approach, however, may not be compatible with running parallel cloud and non-cloud
  systems during the initial evaluation period.

- **Scavenged Resources:** Another alternative approach, supported by [Nur-08, Nur-08-2],
is for cloud software to be installed primarily on computers that already exist within an
organization. In this scenario, cloud systems share hardware resources with other uses of
the hardware and essentially can harvest cycles that might otherwise be wasted. This
approach offers the advantage that cloud services can be made available on an
experimental basis without a large hardware investment, however the resources available
to such a configuration will be limited to the previously-surplus resources in the
organization's infrastructure (unless the former uses of the hardware are reduced in favor
of the cloud). Additional limitations are that: (1) hardware resources must be
incorporated into the on-site private cloud from wherever they exist in a subscriber
organization's infrastructure (via networking) rather than being co-located for efficiency,
and (2) the available hardware may not be homogeneous and thus may be somewhat
more difficult to administer.
**Limited resources (on-site-private).** An on-site private cloud, at any specific time, has a fixed computing and storage capacity that has been sized to correspond to anticipated workloads and cost restrictions. If an organization is large enough and supports a sufficient diversity of workloads, an on-site private cloud may be able to provide elasticity to clients within the subscriber organization while smaller on-site private clouds will exhibit maximum capacity limits similar to those of traditional data centers.

### 1.3 The Outsourced Private Cloud Scenario

Figure 4 depicts an outsourced private cloud. As shown in the figure, an outsourced private cloud has two security perimeters, one implemented by a cloud subscriber (on the right) and one implemented by a provider⁴ (left). The two security perimeters are joined by a protected communications link. As is apparent from the figure, the security of data and processing conducted in the outsourced private cloud depends on the strength and availability of both security perimeters and of the protected communication link. The provider thus accepts a responsibility to enforce the provider-implemented security perimeter and to prevent mingling of private cloud resources with other cloud resources that are outside the provider-controlled security perimeter. The suitability of various mechanisms for achieving an appropriate strength of separation between private cloud resources and other cloud resources depends on the subscriber's security requirements. A number of possible mechanisms could be used with various tradeoffs between separation strength and cost/convenience (e.g., VLAN, VPN, separate network segments or clusters). This scenario should, however, exclude separation mechanisms that are identical to the normal mechanisms that separate customers in a public cloud. If those mechanisms alone were used, this scenario would essentially become the public cloud scenario.

---

⁴ But perhaps configured by the subscriber.
should consider:

**Network Dependency (outsourced-private).** In the outsourced private scenario, subscribers may have an option to provision unique protected and reliable communication links with the provider. Although network dependence does not appear to be avoidable, in this scenario the impact of the network dependency may be ameliorated at a negotiated price.

**Workload locations are hidden from clients (outsourced-private).** As in the general case, to manage a cloud's hardware resources, an outsourced private cloud must be able to migrate workloads between machines without inconveniencing the clients, i.e., without the clients being aware of the migrations. The outsourced private cloud scenario, however, provides an opportunity for the subscriber's organization to have some visibility and control regarding workload locations. Assuming that the provider faithfully implements the security perimeter agreed upon with the subscriber, the subscriber organization workloads move only within the agreed-upon security perimeter. Depending on the mechanisms chosen to implement the perimeter, the subscriber may know the physical location (e.g., cluster, network segments) of the resources devoted to the outsourced private cloud even though the clients are unaware.

**Risks from multi-tenancy (outsourced-private).** The implications are the same as those for an on-site private cloud.

**Data import/export, and performance limitations (outsourced-private).** As with the general case, on-demand bulk data import/export is limited by the network capacity between a provider and subscriber, and real-time or critical processing may be problematic because of networking limitations. In the outsourced private cloud scenario, however, these limits may be adjusted, although not eliminated, by provisioning high-performance and/or high-reliability networking between the provider and subscriber. This provisioning, however, would require a special contract and incur significant costs.

**Potentially strong security from external threats (outsourced-private).** As with the on-site private cloud scenario, a variety of techniques exist to harden a security perimeter. The main difference with the outsourced private cloud is that the techniques need to be applied both to a subscriber's perimeter and provider's perimeter, and that the communications link needs to be protected.

**Modest-to-significant up-front costs to migrate into the cloud (outsourced-private).** Unlike the case of an on-site private cloud, where physical computing resources need to be provisioned or scavenged by a subscriber before the cloud can start operating, in the outsourced private cloud scenario, the resources are provisioned by the provider, and the main startup costs for the subscriber relate to: (1) negotiating the terms of the service level agreement (e.g., agreeing on suitable protection mechanisms), (2) possibly upgrading the subscriber's network to connect to the outsourced private cloud, (3) switching from traditional applications to cloud-hosted applications, (4) porting existing non-cloud operations to the cloud, and (5) training. Although these costs may be significant, they do not include server-side equipment and its supporting infrastructure.

**Extensive resources available (outsourced-private).** Unlike the case of an on-site private cloud, in which the resources must be provisioned by a subscriber up front, in the case of the outsourced private cloud, a subscriber can rent resources in any quantity offered by the provider. Provisioning and operating computing equipment at scale is a core competency of providers. Hence it is likely that a provider can provision relatively large private clouds as needed. As with the on-site private cloud, an outsourced private cloud has a fixed capacity at any given time, and providing elasticity for clients is achievable only if the cloud is large enough and there is sufficient diversity of workloads. As with an on-site private cloud, an outsourced private cloud will exhibit maximum capacity limits similar to those of traditional data centers.
1.4 The On-site Community Cloud Scenario

Figure 5 depicts an on-site community cloud. The community depicted in the figure is made up of a set of participant organizations. Each participant organization may provide cloud services, consume cloud services, or both. It is necessary for at least one community member to provide cloud services for a community cloud to be functional. The figure depicts members that provide cloud services (and possibly consume them also) on the left and those that consume-only on the right. Assuming that each organization implements a security perimeter, the participant organizations are connected via links between the boundary controllers that allow access through their security perimeters. Optionally, organizations may implement extra security perimeters to isolate the local cloud resources from other local resources. Many network configurations are possible; the figure shows the extra security perimeter as being inside an organization's "non-cloud" security perimeter although it could be located outside as well. The boundary controllers in any configuration should grant appropriate access to the cloud resources both to local clients and to clients of other participant organizations. Importantly, providing access to local cloud resources should not grant access to non-cloud resources unless that is a specific policy goal.

In Figure 5 it is easy to see that the access policy of a community cloud may be complex: if there are $N$ community members, a decision must be made, either implicitly or explicitly, on how to share a member's local cloud resources with each of the other members. A number of policy specification techniques (e.g., discretionary access control using standards such as XACML [Mos05], role-based access control [Fer92], and attribute-based access control [Kar09]) might be used to express sharing policies. Additionally, identity management [Oid11, Rag08, Oix10] is important in this scenario since clients from multiple participant organizations access a common pool of resources.

As with the on-site private cloud and the outsourced private cloud, although the general statements remain true for the on-site community scenario, the on-site community cloud scenario also allows for a more detailed understanding of some of the general statements as well as additional statements that organizations considering the use of an on-site community cloud should consider:
As with the on-site private scenario, where the organization spans multiple sites, the subscribers in an on-site community cloud need to either provision controlled inter-site communication links or use cryptography over a less controlled communications media (such as the public Internet). The reliability and security of the community cloud will depend on the reliability and security of the communication links. In the on-site community case, in addition, care should be taken to understand the actual dependencies between member organizations since there are multiple organizations participating and any subset of them could suffer a cloud infrastructure failure (e.g., going offline). Additionally, local clouds will probably need to be taken offline for maintenance at various times and therefore communication in advance among the community members is essential to achieving a clear understanding of the service levels that they offer to one another and require from one another.

Subscribers still need IT skills (on-site-community). In an onsite community cloud, there are potentially two classes of participant organizations: those who provide cloud services to the community, and those to consume cloud resources only. For the participant organizations that provide cloud resources, the IT skills required are similar to those required for the on-site private cloud scenario except that the overall cloud configuration may be more complex and hence require a higher skill level. If any participant organizations are consumers only, the IT skills required are similar to those of the general case except that if there are multiple participant organizations providing cloud services, the configuration from the consuming side may be more complex, e.g., forcing clients to maintain multiple authentication credentials or to commit to an identity management framework.

Identity and access control configurations among the participant organizations may be complex; organizations considering a community cloud should ensure that the IT staff
from the participant organizations negotiate and clearly document the access policies that are planned within the community cloud.

**Workload locations are hidden from clients (on-site-community).** As with the outsourced private cloud scenario, assuming that participant organizations faithfully implement their security perimeters and have policies to keep workloads onsite, workloads should remain within participant organizations. Variations on this scenario are possible, however. For example, a participant organization providing cloud services to the community cloud may wish to employ an outsourced private cloud as a part of its implementation strategy. An organization that is concerned with knowing workload locations should discuss potential outsourcing configurations prior to joining a community cloud, and should ensure that the outsourcing policies are clearly documented for the participant organizations.

**Risks from multi-tenancy (on-site-community).** As with the on-site private scenario, the on-site community scenario somewhat mitigates any multi-tenancy risks by restricting the number of possible attackers. In the on-site community scenario, however, the cloud encompasses more organizations and hence may restrict the set of potential attackers less than in the case of on-site private scenario.

**Data import/export, and performance limitations (on-site-community).** The communication links between the various participant organizations in a community cloud can be provisioned to various levels of performance, security and reliability, based on the needs of the participant organizations. The network-based limitations are thus similar to those of the outsourced-private cloud scenario.

**Potentially strong security from external threats (on-site-community).** The security of a community cloud from external threats depends on the security of all the security perimeters of the participant organizations and the strength of the communications links. These dependencies are essentially similar to those of the outsourced private cloud scenario, but with possibly more links and security perimeters.

**Highly-variable up-front costs to migrate into the cloud (on-site-community).** The up-front costs of an on-site community cloud for a participant organization depend greatly on whether the organization plans to consume cloud services only or also to provide cloud services. For the consume-only scenario, the up-front costs appear to be similar to those for an outsourced private cloud (i.e., modest-to-significant). For a participant organization that intends to provide cloud services within the community cloud, the costs appear to be similar to those for the on-site private cloud scenario (i.e., significant-to-high).

**Limited resources (on-site community).** As with the on-site private cloud scenario, resources for an on-site community cloud must be provisioned or scavenged locally. Therefore the resource limitations appear to be similar to those of the on-site private cloud, i.e., relatively limited.

### 1.5 The Outsourced Community Cloud Scenario

Figure 6 depicts an outsourced community cloud. The community depicted in the figure is made up of a set of participant organizations that consume cloud services. This scenario is very similar to the outsourced private cloud scenario: server-side responsibilities are managed by a cloud provider that implements a security perimeter and that prevents mingling of community cloud resources with other cloud resources that are outside the provider-controlled security perimeter. A significant difference is that the cloud provider may need to enforce a sharing policy among participant organizations in the community cloud.
Although general statements remain true for the outsourced community cloud scenario, the outsourced community cloud scenario also allows for a more detailed view of some of the general statements.

**Network dependency (outsourced-community).** As can be seen from Figure 6, the network dependency of the outsourced community cloud is similar to that of the outsourced private cloud. The primary difference is that multiple protected communications links are likely from the community members to the provider's facility.

**Workload locations are hidden from clients (outsourced-community).** The implications appear to be the same as for the outsourced private cloud scenario.

**Risks from multi-tenancy (outsourced-community).** The implications appear to be the same as for the on-site community cloud scenario.

**Data import/export, and performance limitations (outsourced-community).** The implications appear to be the same as for the outsourced private cloud scenario.

**Potentially strong security from external threats (outsourced-community).** The implications appear to be the same as for the on-site community cloud scenario.

**Modest-to-significant up-front costs to migrate into the cloud (outsourced-community).** The implications appear to be the same as for the outsourced private cloud scenario.

**Extensive resources available (outsourced-community).** The implications appear to be the same as for the outsourced private cloud scenario.

### 1.6 The Public Cloud Scenario

Figure 7 depicts a public cloud. This diagram is essentially similar to Figure 1 except that a subscriber facility implementing a security perimeter is shown. In the case of a public cloud, however, more statements can be made based on the diagram than could from Figure 1. For example, in the public setting, the provider's computing and storage resources are potentially large; the communication links can be assumed to be implemented over the public Internet; and
the cloud serves a diverse pool of clients (and possibly attackers).

As with the other scenarios, although general statements remain true for the public cloud scenario, the public cloud scenario also allows for a more detailed view of some of the general statements.

**Network dependency (public).** In the public scenario, subscribers connect to providers via the public Internet. The dependability of connections thus depends on the Internet's infrastructure of Domain Name System (DNS) servers, the router infrastructure, and the inter-router links. The reliability of connections can thus be affected by misconfiguration [Opp03] or failure of these components as well as network congestion or attack. Additionally, subscribers require a connection via an Internet Service Provider, often designated the "last mile." This connection must also be functional for the cloud to be on-line.

**Workload locations are hidden from clients (public).** In the public scenario, a provider may migrate a subscriber's workload, whether processing or data, at any time. One of the central arguments for cost efficiency in public cloud computing is that data centers (and hence workloads) can be located where costs are low. Generally, workloads in a public cloud may be relocated anywhere at any time unless the provider has offered (optional) location restriction policies and the subscriber has configured their account to request specific location restrictions. Generally, location restrictions in a public cloud are somewhat coarse grained (e.g., the east coast of the US). The confidence that restrictions are actually enforced rests upon protection of subscriber credentials (e.g., that the account has not been hijacked) and the faithfulness with which the provider implements the advertised policies. Generally, subscribers are not in a position to verify that location restrictions have been enforced.

**Risks from multi-tenancy (public).** In a public cloud, a single machine may be shared by the workloads of any combination of subscribers. In practice, this means that a subscriber's workload may be co-resident with the workloads of competitors or adversaries. As summarized in the general case, this introduces both reliability and security risk, and a
failure or attack could be perpetrated by any subscriber. Scaling to larger sets of subscribers and resources is one of the important strategies for public clouds to achieve low costs and elasticity; if this scaling is achieved, however, it also implies a large collection of potential attackers.

**Limited visibility and control over data regarding security (public).** The details of provider system operation are usually considered proprietary information and are not divulged to subscribers. In many cases, the software employed by a provider is also proprietary and not available for examination by subscribers. Consequently, subscribers do not (at the time of this writing) have a decisive way to monitor or authorize access to their resources in the cloud. Although providers may make strong efforts to carry out the requests of subscribers and some may provide monitoring services, subscribers must either trust that the provider is performing operations with fidelity or, if the provider has contracted with a third party auditing organization, trust that the auditing is accurate and timely. As an example of this limitation, a subscriber cannot currently verify that data has been completely deleted from a provider's systems.

**Low up-front costs to migrate into the cloud (public).** The implications appear to be the same as for the outsourced private cloud scenario.

**Elasticity: illusion of unlimited resource availability (public).** Public clouds are generally unrestricted in their location or size. Additionally, they can generally use multi-tenancy without being limited by static security perimeters, which allows a potentially high degree of flexibility in the movement of subscriber workloads to correspond with available resources. As a consequence, public clouds have unique advantages in achieving elasticity, or the illusion (to subscribers) of unlimited resource availability.

**Restrictive default service level agreements (public).** The default service level agreements of public clouds specify limited promises that providers make to subscribers and outline limitations of remedies for subscribers and subscriber obligations. Although marketing literature may make broad claims about cloud system reliability, security, etc., the terms of the service level agreements define the actual (legal) obligations of providers. Section Error! Reference source not found. describes these terms in greater detail.

### 1.7 The Hybrid Cloud Scenario

As given by the cloud definition in Section Error! Reference source not found., a hybrid cloud is composed of two or more private, community, or public clouds. As presented in this section, both the private and the community deployment models have two significant variations: on-site and outsourced. The variations are significant because they have different performance, reliability, and security properties, among others. A hybrid cloud, consequently, is a composition of clouds where each constituent cloud is one of the five variants. There are many conceivable configurations of hybrid clouds and it is not realistic to enumerate them, however the space of possibilities, and the potential challenges, can be illustrated.

Figure 8 depicts how a hybrid cloud could be comprised of a number of clouds representing all of the deployment model variants. The figure depicts access points into the constituent clouds as well as (full) connectivity between them. Both the access points and the connectivity could be implemented in a wide variety of ways, e.g., with access limited based on policies applied by individual constituent clouds. Additionally, global issues such as identity management and shared standards for authentication and information protection within the hybrid cloud are not shown. A further complication not shown is that a hybrid cloud may change over time with constituent clouds joining and leaving.
As depicted in Figure 8, a hybrid cloud can be extremely complex. However, many less complex and highly useful hybrid cloud configurations are possible. For example, "cloud bursting" is an often-discussed concept in which a subscriber uses a private cloud for routine workloads but optionally accesses one or more external clouds during periods of high demand. Using one type of cloud to provide backup resources to another [SN109] is another hybrid possibility as well as using one cloud for disaster recovery [SN109] for a second. For new software developed specifically to run on cloud platforms (e.g., [Msf11-2, Goo11, Sal11]), multi-cloud configurations are possible and even likely. For example, web request handling platform clouds (see Section Error! Reference source not found.) can be very efficient for making web applications continuously available at low cost while on-site or community infrastructure clouds may be more suitable for performing necessary background work to support the applications. Different cloud deployment variants may also be appropriate for particular organizational functions or roles; for example, an organization may elect to process sensitive data such as payroll information in an outsourced private cloud but use a public cloud for new software development and testing activities.