Building Resilient Cloud Network Architectures

Version 1.4
Released: April 20, 2016

Securosis, L.L.C. 515 E. Carefree Highway Suite #766 Phoenix, AZ 85085 **T** 602-412-3051
info@securosis.com www.securosis.com

Author's Note

The content in this report was developed independently of any sponsors. It is based on material originally posted on the Securosis blog, but has been enhanced, reviewed, and professionally edited.

Special thanks to Chris Pepper for editing and content support.

This report is licensed by Resilient, an IBM Company.

All content was developed independently.

www.resilientsystems.com

Resilient's mission is to help organizations thrive in the face of any cyberattack or business crisis. The award-winning Incident Response Platform (IRP) empowers security teams to analyze, respond to and mitigate incidents faster, smarter and more efficiently. Part of IBM Security, the Resilient IRP also integrates security technologies into a single hub and provides easy workflow customization and process automation. With Resilient, security teams can have best-in-class response capabilities. Resilient has more than 100 global customers, including 30 of the Fortune 500 and partners in more than 20 countries. Learn more www.resilientsystems.com.

Copyright

Building Resilient Cloud Network Architectures

Table of Contents

Resilient Cloud Networks

As much as we like to believe we have evolved as a species, people continue to be scared of things they don't understand. Yes, many organizations have embraced the cloud whole hog and are rushing headlong into the cloud age. Yet it's a big world, and millions of others remain paralyzed — not really understanding cloud computing, held captive by their conviction that the cloud cannot be secure because, well, it just can't. Or it's too new. Or some for other unfounded and incorrect reason. Kind of like when folks insisted that the Earth was the center of the universe.

But for those ready, willing, and able to step forward into the future today, the cloud is waiting to break the traditional rules of how technology has been developed, deployed, scaled, and managed.

This paper builds on our recent Pragmatic Security for Cloud and Hybrid Networks research, focusing on cloud-native network architectures that provide security and availability infeasible in a traditional data center. But evolution to new cloud-native architectures will happen over the next decade, and organizations will need to support hybrid networks until they can make the full transition.

But for those ready, willing, and able to step forward into the future today, the cloud is waiting to break the traditional rules of how technology has been developed, deployed, scaled, and managed. We have been aggressive in proselytizing our belief that the move into the cloud is the single biggest disruption in technology for the next few decades. Yes, even bigger than the move from mainframes to client/server (we're old — we know). So this paper on *Building Resilient Cloud Network Architectures* will cover the basics of cloud network security, and include a few design patterns to illustrate the key concepts.

Defining Resilient

If we bust out the old dictionary to define resilient, we get:

Resilient:

- Able to become strong, healthy, or successful again after something bad happens
- Able to return to an original shape after being pulled, stretched, pressed, bent, etc.

In computing resilience is a very important concept. You want to deploy technology that cannot just become strong again, but resist attack in the first place. Recoverability is also key: if something bad happens you want to return service quickly. This brings us to a concept we've been discussing for years, and that's the need to Respond Faster and Better by focusing on efficient, effective and practiced response. Thankfully many organizations have started shifting resources from prevention to detention and response to more evenly balance efforts and funding across all aspects of handling attacks.

For cloud network architecture this resilience concept aligns perfectly with the cloud computing credo: *design for failure*. A resilient cloud network architecture both makes it harder to compromise an application, and also minimizes downtime in case of an issue by ensuring the response apparatus is in place to deal with the inevitable issues.

For cloud network architecture we always fall back on the cloud computing credo: design for failure. A resilient network architecture both makes it harder to compromise an application, and also minimizes downtime in case of an issue.

Key aspects of cloud computing which provide security and availability include:

- **Network Isolation:** Using the inherent ability of the cloud to restrict connections (via software firewalls, also known as security groups, described below), you can build a network architecture that fully isolates the different tiers of an application stack from each other, and even themselves. That prevents compromise in one application (or database) from leaking or enabling attacks on information stored in another.
- **Account Isolation:** Another important feature of the cloud is the ability to use multiple accounts per application. Each of your different environments (Dev, Test, Production, Logging, etc.) can use different accounts, which provides valuable isolation because you cannot access cloud infrastructure across accounts without explicit authorization.
- **Immutability:** An immutable server is one that is never logged into or changed in production. In cloud-native DevOps environments servers are deployed in auto-scale groups based on standard images. This prevents human error and configuration drift from creating exploitation paths for attackers. You use a new known-good state to completely replace older images in production. There is no more patching or logging into live servers.

- **Regions:** You could build multiple data centers in every corner of the world to provide redundancy, but that's not cheap and rarely feasible. To do the same thing in the cloud you can replicate an entire environment in a different region via an API call or a couple clicks in a cloud console. Regions are available all over the world, each with multiple availability zones to further minimize single points of failure. You can load balance between zones and regions, leveraging auto-scaling to keep your infrastructure running the same images in real time. We will explain this design later in this paper.

The key is that cloud computing provides architectural options which are either impossible or economically infeasible in a traditional data center, enabling greater protection and availability.

Understanding Cloud Networking

The key difference between a network in your data center and one in the cloud is that cloud customers never access the 'real' network or hardware. Cloud computing uses virtual networks to abstract the networks you see and manage from the (invisible) underlying physical resources. When your server gets IP address 10.0.1.12, that IP address does not exist on routing hardware — it's a virtual address on a virtual network. Everything is handled in software.

Cloud networks are typically managed via scripts or programs via Application Programming Interface (API) calls, rather than a graphical console or command line. That enables developers to do pretty much anything — including standing networks up and reconfiguring them — instantly via code.

Cloud networking varies across cloud providers, but differs from traditional networks in visibility, management, and velocity of change. You cannot tap into a cloud provider's virtual network, so you need to think differently to monitor your networks. Additionally, cloud networks are typically managed via scripts or programs via Application Programming Interface (API) calls, rather than a graphical console or command line. That enables developers to do pretty much anything — including standing networks up and reconfiguring them — instantly via code.

Finally, cloud networks change much faster than physical networks because cloud *environments* change faster, notably through spinning up and shutting down servers via automation. Traditional workflows to govern network change don't really map to cloud networks. It can be confusing because cloud networks *look* like traditional networks, with their own routing tables and firewalls and all. But looks are deceiving: familiar constructs carry over, but there are fundamental differences.

Cloud Network Architectures

In order to choose the right solution to address your requirements, you need to understand the types of cloud network architectures and the different technologies that enable them. There are two basic types of architectures:

- **Public Cloud Networks:** These are entirely Internet-facing. You connect to your instances (servers) via the public Internet with no special routing; every instance has a public IP address.

- **Private Cloud Networks:** Also called "virtual private clouds" or VPCs, these look like internal LANs using private IP addresses. You access these networks via some kind of non-public Internet connection — typically a VPN.
- **Hybrid Cloud Networks:** These are networks that consists of both public and private cloud networks. There is typically a connection point between the Internet-facing network (public) and the internal network (VPC).

Cloud networks are enabled and supported by the following technologies:

- **Internet Gateways:** A gateway connects your cloud network to the Internet. You don't normally manage it directly — your cloud provider does it for you because their tools move packets from 'your' internal network to the Internet.
- **Internal Gateways:** These devices connect existing datacenters to your private cloud network. You access networks via a VPN provided by the cloud provider or a direct connection, which looks suspiciously like a traditional point-to-point connection from days gone by.
- **Virtual Private Networks:** You can also set up your own overlay network to bridge private and public cloud networks within your cloud provider. This provides a private segment with access for users, developers, and administrators.

These terms will come into play when we present design patterns later in this paper.

Network Security Controls

Your network is different in the cloud, so your network security controls differ as well. But you may find some comfort in familiar categories. Cloud network security controls fall generally into five buckets:

Software firewalls combine the best of network and host firewalls, allowing you to deploy policies on individual servers (or even network interfaces) like a host firewall, but manage them together like network firewalls.

1. **Cloud Perimeter Security:** These controls generally provide coarse protection from very common network-based attacks, including Denial of Service. Your cloud provider provides and manages these controls; you have no visibility or control.
2. **Software Firewalls:** These firewalls are built into the cloud platform (they are called security groups in AWS) and protect cloud assets such as instances. They offer basic access control via ports/protocols and sources/destinations, and are designed to handle auto-scaling and cloud environments. They combine the best of network and host firewalls, allowing you to deploy policies on individual servers (or even network interfaces) like a host firewall, but manage them together like network firewalls. They will be your main tool to provide (virtual) network isolation, described above.

3. **Access Control Lists:** While a software firewall works at a per-instance (or per-object) level, ACLs restrict communications between subnets of your virtual network. Old-school networking folks will be familiar with using ACLs to control access into and out of subnets in a (virtual) cloud network.

4. **Virtual Appliances:** A number of traditional network security tools, including IDS/IPS, WAF, and NGFW, are available as virtual appliances to improve network security, but they require you to route cloud traffic through (virtual) devices.

5. **Host Security Agents:** These agents are built into immutable server images, and provide visibility and protection for each server/instance in a cloud environment.

The key to cloud networking is that you don't need to apply the same controls, or even configurations, to an entire network. You can make architectural and security decisions per project. You might decide an entirely cloud-based VPC is best for one application, and for another choose to build an overlay VPN to connect a totally different VPC to your datacenter to establish a hybrid environment. You might need to route one application's traffic through an inspection point to prevent data leakage, while for another you rely exclusively on security groups to provide full isolation between different layers of your cloud stack. The permutations are infinite, providing flexibility unavailable in a data center.

PaaS Air Gap Design Pattern

To demonstrate a more secure cloud network architecture, consider an Internet-facing application with both web server and application server tiers. Due to the application's nature, communications between the two layers takes place through message queues and notifications, so the web servers don't need to communicate directly with the app servers. The application server tier connects to the database (a Platform as a Service offering from the cloud provider), so the data layer can only be accessed via the application server tier. The application server tier also communicates with a traditional data center to access internal corporate data outside the cloud environment.

An application must be architected from the get-go to support this design. You wouldn't even try to forklift your 20-year-old legacy general ledger application into the cloud using this design. But if you are architecting a new application, or can totally re-architect an existing application for total isolation between environments, this is one way to do it.

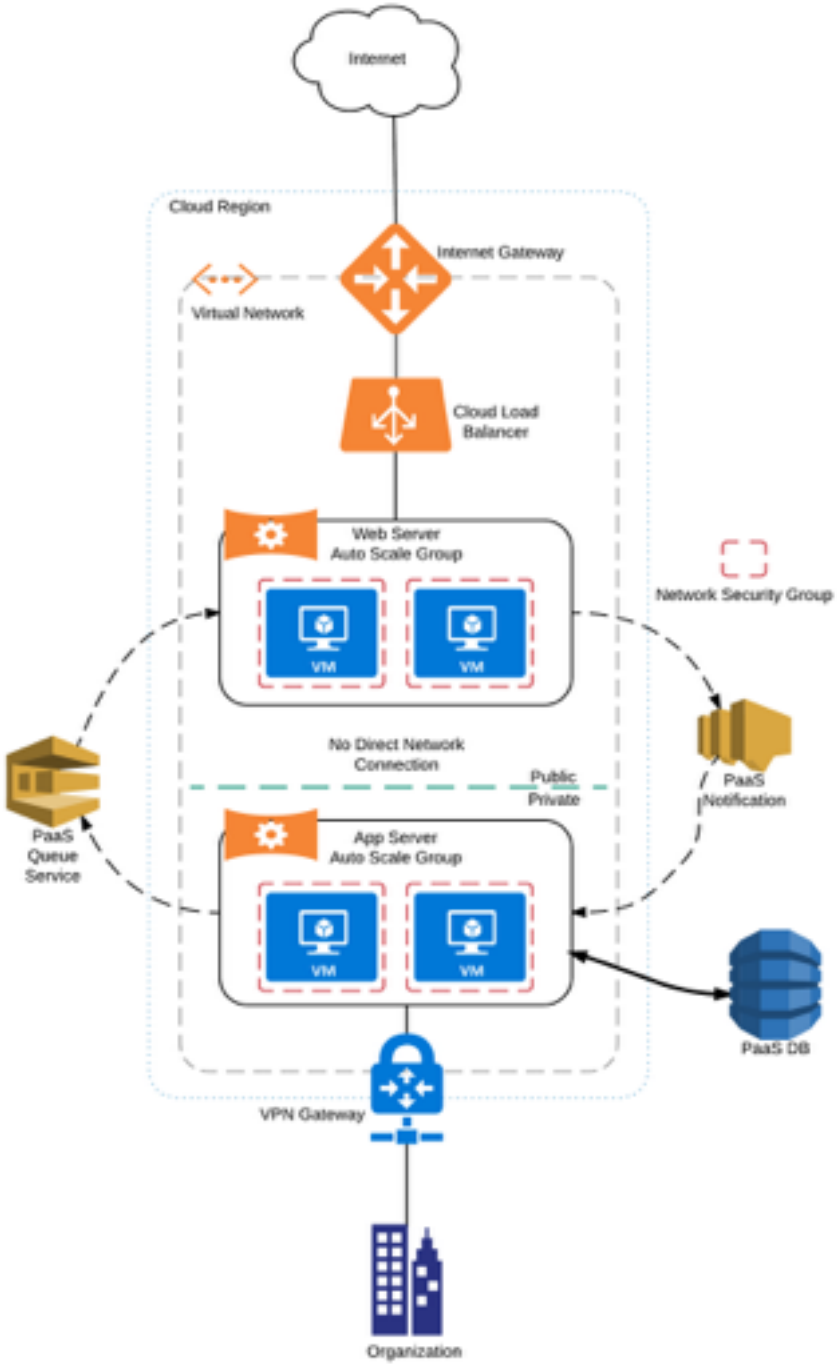
Network Security Groups

The key security control typically used in this architecture is a Network Security Group, allowing access to app servers only from web servers, and only to the specific port and protocol required. This isolation *limits blast radius*. To be clear, the NSG applies individually to each instance — not to subnets. This avoids a flat network, where all instances within a subnet have unrestricted access to all subnet peers.

PaaS Services

In this application you wouldn't need to open access from the web server NSG to the app server NSG, because the architecture doesn't require direct communication between them. Instead the cloud provider offers a message queue platform and notification service which provide asynchronous communication between the web and application tiers. So even if the web servers are compromised, the app servers are not accessible.

Further isolation is offered by a PaaS database, also provided by the cloud service provider. You can run the PaaS DB in its own private cloud network (VPC) and restrict access to specific Network Security Groups. This ensures only instances in the application tier can request information from the database service.



Download the design pattern.

Connection to the Data Center

The application may require data from an existing data center, in which case the app servers must have access to the corporate network through a VPN. All traffic to the data center is routed through this inspection and control point. We generally prefer not to route cloud traffic through inspection bottlenecks, but in this design pattern it's not a big deal, because the traffic needs to pass through a specific egress connection to the data center anyway, so you might as well inspect there.

To further protect the cloud stack you restrict all ingress traffic through that connection to the app server Security Group. This ensures that an adversary who compromises your network cannot access your whole cloud network through your data center.

Advantages of This Design

- **Isolation between Web and App Servers:** By putting the auto-scaling groups in Network Security Groups, you restrict access to only what is explicitly authorized. This is the new flavor of classic *default deny* security.

- **No Direct Connection:** In this design pattern you can block direct traffic to the application servers from anywhere but the VPN. Intra-application traffic between the web and application tiers is asynchronous via the message queue and the notification service providing isolation. This significantly reduces the attack surface of the application, putting the application tier and PaaS services out of the reach of attackers.

- **PaaS Service:** This architecture uses cloud provider services which offers strong built-in security and resilience. *Cloud providers understand that security and availability are core to their business, and do much better jobs protecting their environments than the vast majority of enterprises.*

What's next for this kind of architectural design? To advance this concept you could deploy application replicas in different zones within a region (or even across regions, which we'll show in the next design) to further restrict access to sensitive data in case one device is compromised. You also get additional resiliency in case of a zone failure since the traffic will be automatically sent to the other zone.

And if you use immutable servers within each auto-scale group, you can update/patch/reconfigure instances automatically by updating the master image and having auto-scaling replace the old instances with new ones. This limits configuration drift and can remove adversary persistence.

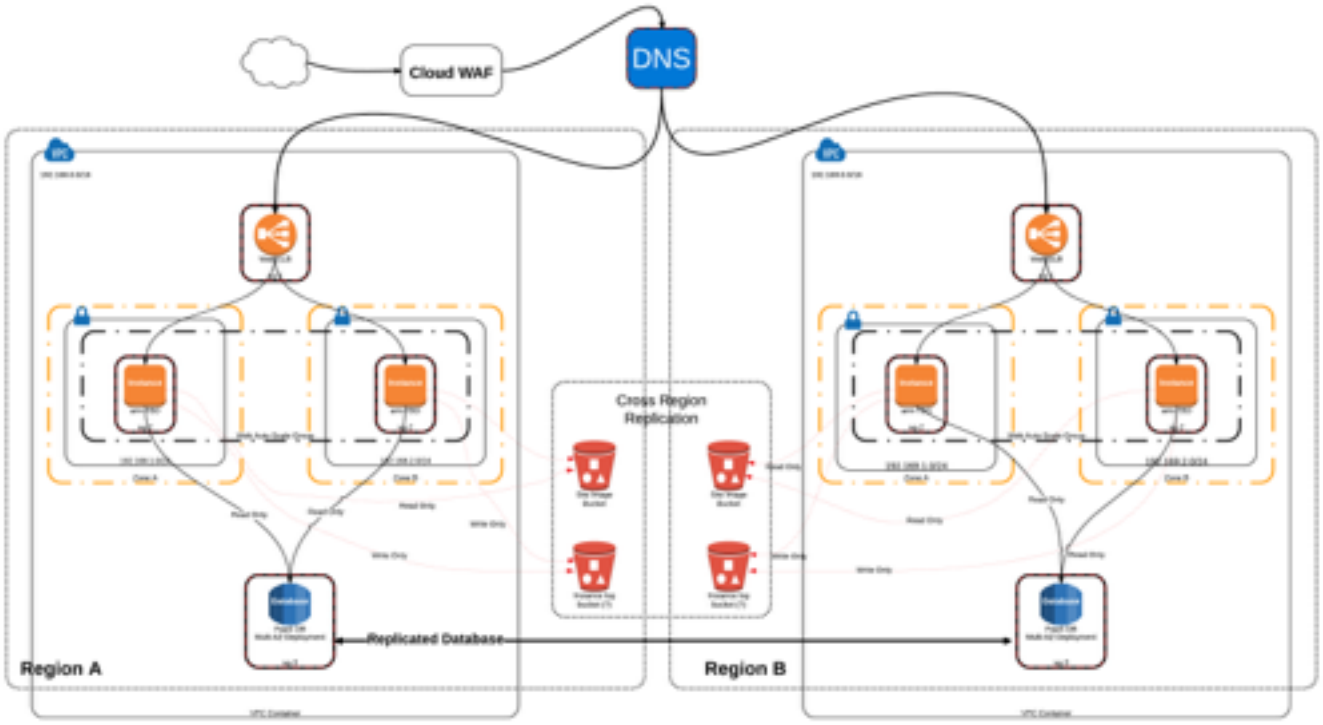
Multi-Region Website Design Pattern

This architecture was designed to deploy a website in multiple regions, with availability as close to 100% as possible. This design is far more expensive than even running in multiple zones within a single region, because you need to pay for network traffic between regions (compared to free intra-region traffic); but if uptime is essential to your business this improves resilience. We should know — Securosis uses this basic architecture for our public web presence in the cloud.

Websites are externally exposed application, so we recommend running inbound traffic through a cloud WAF to get rid of obvious attacks. Inbound sessions can be routed intelligently via DNS service to the appropriate region. You can route based on server utilization, network traffic, client location, and a variety of other criteria. This is a great example of software-defined security: programming traffic distribution within cloud stacks via code.

- **Network Security Groups:** In this design pattern you implement Network Security Groups to lock down traffic into the app servers. This isn't called out because it would greatly complicate the diagram. But Network Security Groups for web and app servers are critical to this architecture.
- **Compute Layer:** Application and web servers are in auto-scale groups within each region. The load balancer distributes sessions between zones intelligently to ensure efficient utilization.
- **Database Layer:** If this was just a multi-zone deployment you wouldn't need to worry about database replication because multi-zone replication is built into PaaS databases, but not for multiple regions. Instead *you* need to replicate your databases across regions. It is the cloud equivalent of having two separate data centers. We cannot tackle the network architecture to support database replication here because that would overcomplicate the architecture, but we need to point out another way operating across multiple regions adds complexity.

- **Static Files:** Any website includes a variety of static files, so a key aspect of this design is keeping file stores in sync between regions as well. Using Network Security Groups, you can lock access to storage buckets down to specific auto-scale groups. That's a good way to make sure you don't get malware files loaded onto your website thanks to faulty configuration. To keep files synced across the regions, your cloud provider may offer a replication service so you don't need to build it yourself.



See the full design (JPG)

Advantages of This Design

This architecture is intended to show how the cloud provides enables you to easily establish an application across multiple regions, similar to how you would deploy across data centers. So what's unique about the cloud? You can take the entire stack in Region A and copy it to Region B with a few clicks in a cloud console. Of course you'd have some networks to reconfigure, but in most ways the environment is identical. Auto-scale groups work off the same images, so the operational overhead of supporting multiple cloud regions is drastically lower than operating multiple data centers.

There are also significant availability advantages. If a region goes down, the DNS service can detect that automatically and route all new incoming sessions to the available region, which can auto-scale up to handle the increased load. Existing sessions in the failed region will be lost, so there is some collateral damage, but that happens any time you lose a data center. Those sessions can be re-established to the available region quickly. When the region recovers DNS can automatically start sending new sessions to it again, transparently to applications and users.

So what's unique about the cloud? You can take the entire stack in Region A and copy it to Region B with a few clicks in a cloud console. Services which previously required extreme planning and/or additional products to manage have become built-in platform features.

This design can also keep static files and logs in sync using cross-region replication. This capability is currently specific to Amazon Web Services, but we expect it to become a critical built-in feature of all cloud infrastructure platforms at some point. As is increasingly the case in the cloud, services which previously required extreme planning and/or additional products to manage have become built-in platform features.

Summary

The cloud provides many capabilities that enable you to deploy applications significantly more securely and reliably than rolling them out in your own datacenter. Of course there is resistance to the new way of thinking — there always is. But you can counter this resistance using information, including the suggestions in this paper, along with our other research and reference architectures, to highlight the obvious advantages of the cloud.

If you have any questions on this topic, or want to discuss your situation specifically, feel free to send us a note at info@securosis.com.

About the Analyst

Mike Rothman, Analyst and President

Mike's bold perspectives and irreverent style are invaluable as companies determine effective strategies to grapple with the dynamic security threatscape. Mike specializes in the sexy aspects of security — such as protecting networks and endpoints, security management, and compliance. Mike is one of the most sought-after speakers and commentators in the security business, and brings a deep background in information security. After 20 years in and around security, he's one of the guys who "knows where the bodies are buried" in the space.

Starting his career as a programmer and networking consultant, Mike joined META Group in 1993 and spearheaded META's initial foray into information security research. Mike left META in 1998 to found SHYM Technology, a pioneer in the PKI software market, and then held executive roles at CipherTrust and TruSecure. After getting fed up with vendor life, Mike started Security Incite in 2006 to provide a voice of reason in an over-hyped yet underwhelming security industry. After taking a short detour as Senior VP, Strategy at eIQnetworks to chase shiny objects in security and compliance management, Mike joined Securosis with a rejuvenated cynicism about the state of security and what it takes to survive as a security professional.

Mike published The Pragmatic CSO <http://www.pragmaticcso.com/> in 2007 to introduce technically oriented security professionals to the nuances of what is required to be a senior security professional. He also possesses a very expensive engineering degree in Operations Research and Industrial Engineering from Cornell University. His folks are overjoyed that he uses literally zero percent of his education on a daily basis. He can be reached at mrothman (at) securosis (dot) com.

About Securosis

Securosis, LLC is an independent research and analysis firm dedicated to thought leadership, objectivity, and transparency. Our analysts have all held executive level positions and are dedicated to providing high-value, pragmatic advisory services. Our services include:

- **Primary research publishing**: We currently release the vast majority of our research for free through our blog, and archive it in our Research Library. Most of these research documents can be sponsored for distribution on an annual basis. All published materials and presentations meet our strict objectivity requirements and conform to our Totally Transparent Research policy.
- **Research products and strategic advisory services for end users**: Securosis will be introducing a line of research products and inquiry-based subscription services designed to assist end user organizations in accelerating project and program success. Additional advisory projects are also available, including product selection assistance, technology and architecture strategy, education, security management evaluations, and risk assessment.
- **Retainer services for vendors**: Although we will accept briefings from anyone, some vendors opt for a tighter, ongoing relationship. We offer a number of flexible retainer packages. Services available as part of a retainer package include market and product analysis and strategy, technology guidance, product evaluation, and merger and acquisition assessment. Even with paid clients, we maintain our strict objectivity and confidentiality requirements. More information on our retainer services (PDF) is available.
- **External speaking and editorial**: Securosis analysts frequently speak at industry events, give online presentations, and write and speak for a variety of publications and media.
- **Other expert services**: Securosis analysts are available for other services as well, including Strategic Advisory Days, Strategy Consulting engagements, and Investor Services. These tend to be customized to meet a client's particular requirements.

Our clients range from stealth startups to some of the best known technology vendors and end users. Clients include large financial institutions, institutional investors, mid-sized enterprises, and major security vendors.

Additionally, Securosis partners with security testing labs to provide unique product evaluations that combine in-depth technical analysis with high-level product, architecture, and market analysis. For more information about Securosis, visit our website: <http://securosis.com/>.