CHAPTER 7

A New Security Architecture to Improve Business Agility

An organization's ability to learn, and translate that learning into action rapidly, is the ultimate competitive advantage.

—Jack Welch

Some *Star Trek* episodes feature suspense-filled battles in which adversaries use sophisticated phase-shifting weapons that can be rapidly adjusted until they find a way to penetrate static force-field defenses. For a beleaguered starship, the only effective response is to use similarly adaptable and fast-changing shields.

As information security professionals, we also need extremely agile defenses that quickly adapt to meet new demands. Attackers are continually adapting, and defenders also need to continually adapt. But rapidly evolving threats are only part of the challenge. We also need to continually adapt our defenses to a rapidly changing technology landscape.

As information risk and security groups consider the future, it's clear that we need to radically change our approach in order to face the challenges ahead and support the Protect to Enable mission.

One problem in recent years has been that most of the protection offered by the industry has not kept up with the attackers. Because these tools have failed to prevent harm, many companies have defaulted to a detect-and-respond approach. This means they continue to expose themselves to high risks and higher long term costs since they are reactively responding to attacks that have already breached the organization's defenses.

We also need to consider whether our existing control architecture improves or impedes business agility and velocity. It's important to recognize that controls can place a "drag coefficient" on the business. By hindering users, they can stifle business velocity and innovation. Users react to this control friction by circumventing the controls whenever possible; as a result, the controls can actually introduce new risks, as discussed in Chapter 2.

M.W. Harkins, *Managing Risk and Information Security*,
DOI 10.1007/978-1-4842-1455-8_7

As we move forward, we will need an agile security architecture that quickly and automatically learns and adapts to new challenges as they emerge. A learning system is harder to defeat because it can more quickly predict and thus prevent new attacks. The pace of change is so rapid that we cannot predict all the challenges we will face, and manual or semi-manual processes will not be enough to keep up. We will need solutions that can learn to manage what we don't know.

The right control architecture will enable flexibility that helps the business move more quickly, allowing us to rapidly adopt new technologies and emerging usage models while continuing to provide security in the ever-evolving threat landscape.

A few years ago, after intense brainstorming sessions, the information risk and security team I led at Intel devised a new security architecture for the company. This architecture represented our implementation of the Protect to Enable strategy, using the technologies that were current at that time. With the benefit of hindsight, I believe that we got many things right—but there were also some omissions because we didn't have a full understanding of the controls that would be needed.

In this chapter, I'll provide a high-level overview of a new security architecture and describe how it meets some key challenges. Some of this overview is based on the work at Intel a few years ago, but I have added a new perspective on controls that I have realized is lacking in the industry. An important aspect of this new perspective is the concept of control friction. As I'll explain later in the chapter, I've developed a simple framework called the 9 Box of Controls, which takes control friction into account when assessing the value of security controls.

I believe that the architecture includes some novel approaches that may be valuable to many organizations facing these universal challenges. My conversations with peers at other companies have validated this view. Many of them are considering similar strategies and in some cases have begun implementing them.

Any future security architecture must provide better prevention, and it must also be more flexible, dynamic, and granular than traditional enterprise security models. This will helps us all accommodate future evolving usage models. We can provide users with different levels of access depending on factors such as the devices they are using and their location. To achieve this, the architecture dynamically adjusts a user's access privileges as the level of risk changes. For example, an employee should have more limited access to our systems when using a less-secure device than when using a more hardened or perhaps fully managed enterprise-class system.

The new architecture greatly improves threat management. As new attacks appear, we need to be able to recognize good from bad in milliseconds, so that we can stop the bad and allow the good. For any attack that gets past the preventive controls, we need to learn as much as we can without compromising the user's computing performance or privacy. This information enables us to investigate what occurred, so we can quickly take action to mitigate the risk and also learn how to prevent similar attacks in the future. A control architecture should assume that attempts at compromise are inevitable, but we should also understand that it's possible to achieve real prevention for 99% or more of malicious code. We can apply artificial intelligence and machine learning to analyze the features of files, executables, and binaries to stop malicious code prior to execution. For the remaining attacks, representing less than 1% of malware, we need to focus heavily on survivability.

The 9 Box of Controls, Business Trends, and Architecture Requirements

Before diving into the specifics of the architecture, I'll explain the 9 Box of Controls. Then I'll recap some of the key business and technology trends, focusing on how they drive the need for specific capabilities in security technology.

9 Box of Controls

There are three primary types of security controls: prevention, detection, and response. Prevention occurs when an action or control prevents a risk before it affects users or the environment. Detection is identifying the presence of something malicious that has already entered the environment. Response is a reaction. From a risk perspective, prevention focuses on minimizing vulnerability and the potential for harm, while detection and response focus on minimizing damage.

There are also three primary levels of control automation: automated, semi-automated, and manual. Automated control occurs entirely through machines. Semi-automated involves some level of human intervention. Manual controls are managed entirely by hand.

The combinations of these control types and automation levels comprise the cells of the 9 Box, as shown in Figure 7-1. Risk increases as we move from prevention to detection to response. Cost increases as we move from automated to semi-automated to manual controls.

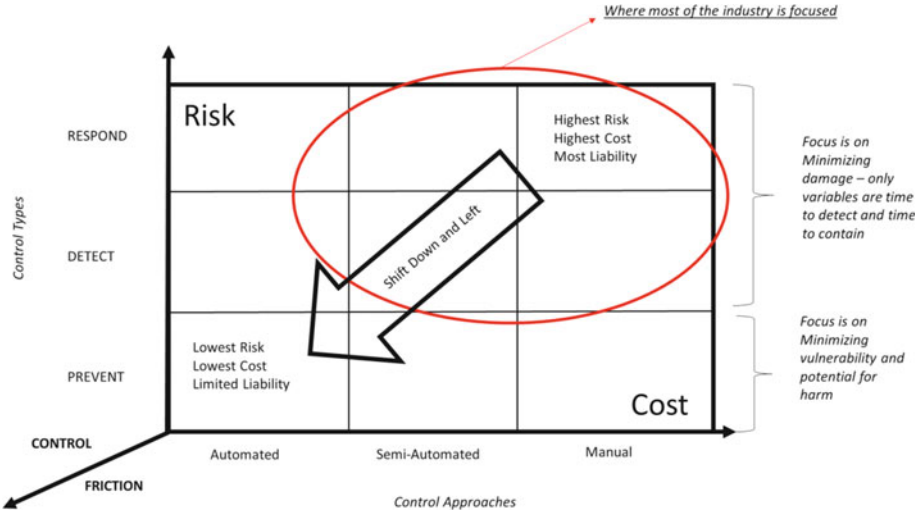

Figure 7-1. *The 9 Box of Controls*

However, there is a third dimension to the 9 Box: control friction. As we know, friction is the force that causes a moving object to slow down when it is in contact with another object. Similarly, controls can impose a "drag coefficient" on business velocity—they can slow the user or a business process. However, friction is not a fundamental, immutable force like gravity or electromagnetism. Instead, we have the ability to determine how much control friction we apply. Apply too much control friction, and business users will go around IT and its security controls. This adds cost: IT is no longer managing the technology; data and business silos are created, and the organization loses its volume purchasing power. It also adds risk: because the security team lacks visibility into the technology, it cannot prevent compromises, detection is difficult, and in many cases response after the fact becomes the only option. If a business adheres to high-friction controls, the effect can be to generate systemic business risk. High-friction controls can hinder business velocity; the organization can lose time to market and the ability to innovate, and over the long term it may even lose market leadership.

IT Consumerization

As I discussed in Chapter 5, consumerization is a major IT theme with ever-broadening impact. It includes several trends, including the adoption of new applications and support for consumer devices.

Many highly mobile employees want to use their own consumer devices, such as smartphones, wearables, and tablets, for work. This increases productivity by enabling employees to collaborate and access information from anywhere, at any time. To support this, organizations provide access to corporate e-mail and other applications from employee-owned smartphones and tablets.

Some people believe that in the future, all devices will be consumer-owned, and that enterprises will no longer purchase devices for their users. I believe this might be the case in some work environments, but I doubt that it will suit all organizations. For a company providing call center services, with most employees working from home, it might make sense that employees exclusively use their own personal systems for work. But this strategy could be more risky for a financial services company whose employees handle highly sensitive information that is subject to extensive regulatory requirements.

Nevertheless, the consumerization trend continues to grow at almost all organizations. Accordingly, we'll need to provide employees with a level of access to resources from an expanding continuum of client devices, some of which may have much weaker security controls than today's enterprise clients (see sidebar).

CONSUMERIZING ENTERPRISE IT AND "ENTERPRISING" THE CONSUMER

Discussions of IT consumerization tend to draw a clear line between business devices that can be managed and trusted, and personal consumer devices that are essentially unmanaged and untrusted.

However, not all consumer devices are created equal. From a security standpoint, it may be more valuable to think about a device's capabilities than to categorize it based solely on whether it's marketed as an enterprise device or a personal device. The security of a device depends on the inherent features of the hardware, operating system, and applications, and on whether it enables us to add further security and manageability capabilities that mitigate the risks of enterprise use.

As the variety of consumer devices, such as smartphones and wearables, continues to expand, users may choose from dozens of models with different levels of security capabilities. Greater security and manageability means that IT can place greater trust in the device and provide a correspondingly greater level of access to enterprise resources.

Extending this idea further, the information security group could evaluate the security of available consumer devices and provide guidance about the level of enterprise access that users will be allowed with each device. Users may prefer to buy a more secure device because it will provide them more access. With greater access, they can use the device for more of their daily work activities. This ability in turn enables them to be more productive.

At the same time, employees increasingly expect to have available to them at work the types of consumer services and cloud applications that they use in their personal lives. These include social computing applications such as blogs and wikis, video-sharing sites, and file-sharing services.

We need a security architecture that enables us to more quickly support new devices and provide access to a greater range of applications and data, without increasing risk. We need to be able to dynamically adjust the levels of access we provide and the monitoring we perform, depending on the security controls of the client device.

New Business Needs

Nearly all companies now rely on a growing network of business partners, and conduct many of their interactions with those partners online. Many organizations are also expanding into new markets through both organic growth and acquisitions. Because of these business trends, most organizations will need to provide access to a broader range of users, many of whom are not employees. Many organizations also need to be able to smoothly integrate acquired companies and provide them with access to resources. In general, we need to quickly provide new users access while minimizing risk and providing selective, controlled access only to the resources they need.

Cloud Computing

Most organizations are already using cloud services in some form to achieve benefits such as greater agility and lower cost. Some are also implementing a private cloud based on virtualized infrastructure while using external cloud services for noncritical applications. In the future, I expect greater use of hybrid clouds that use both internal and external resources, especially for organizations that are anchored to legacy environments. Organizations able to let go of their legacy environments will predominantly use the cloud, with limited internal infrastructure.

This trend means that IT services at many organizations will be provided by a mixture of traditional and cloud-based internal and external services. During a typical day, employees may access a variety of different services, some of which are internal and some external. Ultimately, they should be able to easily move between these services without needing to log in multiple times or even know where the services are located.

Securing access to cloud-based services presents challenges that aren't easily addressed using conventional security controls. In cloud environments, systems and their data are virtualized and may migrate dynamically to different network locations. This makes it difficult to effectively restrict access using traditional security controls such as firewalls, which rely on fixed locations of systems and a more static nature of the data. We need much more granular and dynamic controls that are linked to the resources themselves rather than just their network location.

Changing Threat Landscape

The threat landscape is evolving rapidly. Increasingly, attackers have taken a stealthy approach, creating malware that quietly gains access and attempts to remain undetected in order to maintain access over time. This has been possible because the security solutions deployed on endpoints in most organizations today do not adequately prevent malicious code from executing. As the number of threats increases and new types of malware emerge, we need to focus on the 9 Box of Controls and seek new prevention methods reduce risk, reduce cost, and reduce control friction.

Traditional enterprise security architectures have relied largely on protective controls such as firewalls located at the network perimeter and signature-based antivirus at the end points. At the same time, our focus has shifted to providing controlled access to a broader range of users and devices, rather than simply preventing access. Combine this with a continually changing threat landscape, and we can assume that attempts to compromise the environment are inevitable. Although existing perimeter controls such as firewalls will continue to have some value, we need tools that can dramatically increase the ability to prevent attackers from gaining access to the environment, but in way that does not introduce a cost burden or a high degree of control friction.

Privacy and Regulatory Requirements

The growing emphasis on privacy requirements and the increasingly complex regulatory environment have many implications for the way we manage information. Some regulations create the need for more control over where information is stored and require specific levels of protection and tracking. Our architecture must provide this assurance, allowing us to build a high-security environment and access controls appropriate for the protection of highly regulated information. In addition, the security controls themselves must not introduce privacy risks.

New Architecture

To meet these rapidly changing requirements, we need a highly flexible and dynamic architecture. The architecture should enable us to more quickly adopt new devices, use models, and capabilities; provide security across an increasingly complex environment; and adapt to a changing threat landscape.

Key goals include helping increase employee productivity while supporting new business requirements and technology trends, including IT consumerization, cloud computing, and access by a broader range of users. At the same time, the architecture should be designed to prevent risk, reduce our attack surface, and improve survivability—even as the threat landscape grows in complexity and maliciousness.

The architecture moves away from the traditional enterprise trust model, which is binary and static. With this traditional model, a user is in general either granted or denied access to resources; once granted, the level of access remains constant. The new architecture replaces this with a dynamic, multi-tiered trust model that exercises more fine-grained control over identity and access control, including access to specific resources. This means that for an individual user, the level of access provided may vary dynamically over time, depending on a variety of factors—such as whether the user is accessing the network from a highly secure managed device with advanced anti-malware capabilities or an untrusted and perhaps unmanaged device.

The architecture's flexibility allows us to take advantage of trust based on real proof that malware execution is being prevented. Increasingly, devices may include some level of hardware-enforced security designed to ensure the integrity of the applications and data on the device. The architecture takes this into account when determining whether to allow access to specific resources—a more-trusted platform can be allowed greater access than a less-trusted one. The architecture is based on four cornerstones:

- **Trust calculation**: This unique element of the architecture handles user identity and access management, dynamically determining whether a user should be granted access to specific resources and, if so, what type of access should be granted. The calculation is based on factors such as the user's client device and location, the type of resources requested, and the security controls that are available.

- **Security zones**: The infrastructure is divided into multiple security zones that provide different levels of protection. These range from trusted network zones containing critical data, with tightly controlled access, to untrusted zones containing less-valuable data and allowing broader access. Communication between zones is controlled and monitored; this helps ensure users can only access the resources for which they have been authorized and prevents compromises from spreading across multiple zones.
- **Balanced controls**: To increase flexibility and the ability to recover from a successful attack, the model emphasizes the need for preventative controls but also to balance them with detection and response.
- **User and data perimeters**: Recognizing that protecting the enterprise network boundary is no longer adequate, we need to treat users and data as additional security perimeters and protect them accordingly. This means an increased focus on the endpoint device and prevention of malicious code, in addition to increasing user awareness and building data protection into the information assets.

I'll describe each of the four cornerstones in more detail.

Trust Calculation

The trust calculation plays an essential role in providing the flexibility required to support a rapidly expanding number of devices and usage models. The calculation enables us to dynamically adjust users' levels of access, depending on factors such as the devices and networks they are currently using.

It calculates trust in the interaction between the person or device requesting access (source) and the information requested (destination). The calculation consists of a source score and a destination score, taking into account the controls available to mitigate risk. As shown in Figure 7-2, the result of this calculation determines whether the user is allowed access and the type of access provided.

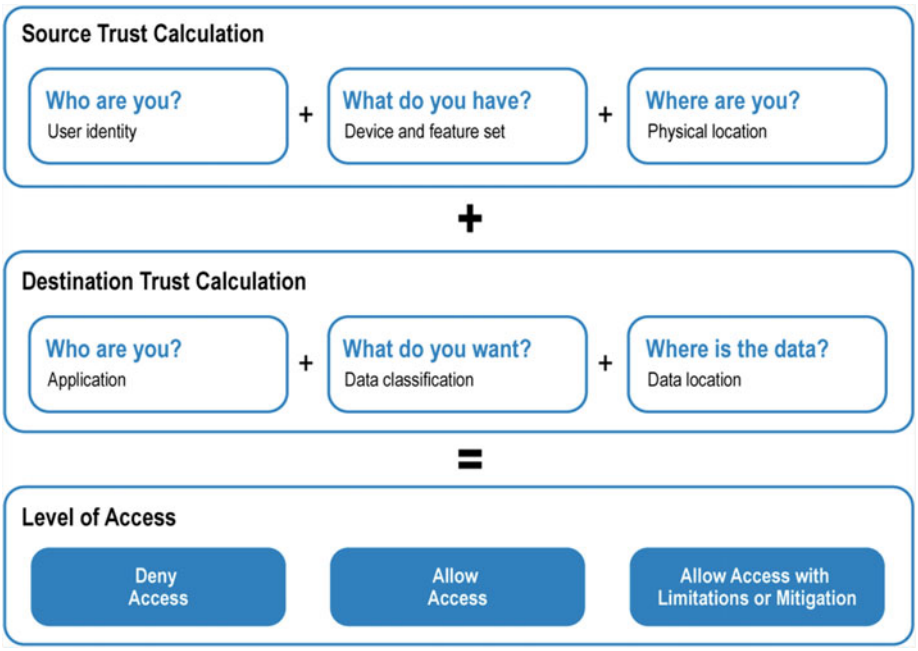

Figure 7-2. *Trust calculation. Source: Intel Corporation, 2012*

Source Score

Trust in the source, or requestor, is calculated based on the following factors:

- *Who*: The identity of the user or service requesting access and our confidence level in the authentication mechanism used; how confident are we that users are who they say they are?
- *What*: The device type, its control capabilities, our ability to validate those controls, and the extent to which IT manages the device.
- *Where*: The user's or service's location. For example, a user who is inside the enterprise network is more trusted than the same user connecting through a public network. There may also be other considerations, such as the geographical region where the user is located.

Destination Score

This is calculated based on the same three factors, but these are considered from the perspective of the destination (the information the source is trying to access):

- *Who*: The application that stores the requested data. Some applications can enforce greater controls, such as enterprise rights management (ERM), and therefore provide a higher level of trust.
- *What*: The sensitivity of the information being requested and other considerations, such as our ability to recover it if compromise occurs.
- *Where*: The security zone in which the data resides.

Available Controls

The trust calculation also takes into account the security controls available for the zone. If the only controls available are controls that simply block or allow access, we might deny access due to lack of other options. However, if we have extensive preventative controls with highly granular levels of access, detailed logs, and highly tuned security monitoring—as well as the ability to recover from or correct problems—then we can allow access without creating additional risk.

Calculating Trust

The trust calculation adds the source score and the destination score to arrive at an initial trust level. The available controls are then considered to make a final decision about whether access is allowed and, if so, how. This calculation is performed by a logical entity called a *policy decision point* (PDP), which is part of the authentication infrastructure and makes access control decisions based on a set of policies.

Based on the results of this calculation, the PDP makes a decision, allocating a trust level that determines whether the user can access the requested resource and the type of access that is allowed. Broadly, the decision will fall into one of the following categories:

- Allow access
- Deny access
- Allow access with limitations or mitigation

This trust calculation therefore allows us to dynamically apply granular control over access to specific resources. For example, employees using IT-managed devices with additional hardware features such as a trusted platform module (TPM), global positioning system (GPS), and full disk encryption would be allowed access to more resources than when using devices that lack those features.

Employees directly connected to the network would typically get greater access than when using a public network. If we are unable to verify the location of a high-security device such as a managed PC, we would allow less access.

The trust calculation also can be used for more fine-grained distinctions between different device models. For example, we could provide different levels of access based on manageability, hardware-enabled authentication and encryption, and installed applications.

We anticipate situations in which the trust level is not adequate to allow any access, but there is still a business requirement to allow a connection or transaction to occur. In these conditions, the result of the trust calculation could be a decision to allow access with limitations or with compensating controls that mitigate the risk. For example, a user might be allowed read-only access or might be permitted access only if additional monitoring controls are in place.

Today, the trust calculation makes decisions based on information gathered from components at multiple levels of the infrastructure, such as network gateways, access points, and user devices. Once the trust calculation mechanism is in place, we can extend it to include information from a broader range of sources.

The trust calculation can be used to determine access to internal systems by business partners as well as employees. Let's say we're collaborating with another company on the design of a new product. An engineer at that company wants access to a specific document. We can add a variety of criteria to the trust calculation for deciding whether to grant access. Did the engineer's request originate within the business partner's enterprise network? Is it consistent with the type of request that we'd expect from an engineer? If so, we have a higher level of trust in the requestor.

If we cannot establish an adequate level of trust in the user's device, but other factors provide enough confidence to grant access, we might provide one-time access for a specific job. We could do this by allowing a document to be downloaded, but only within a container that ensures the document is completely removed from the user's device once the job is completed.

Longer term, the trust calculation could become a mechanism that is used to determine access to both internal and external resources, including cloud-based applications.

Security Zones

The architecture divides the IT environment into multiple security zones. These range from untrusted zones that provide access to less valuable data and less important systems to trusted zones containing critical data and resources.

Because the higher-trust zones contain more valuable assets, they are protected with a greater depth and range of controls, and we restrict access to fewer types of devices and applications, as shown in Figure 7-3. However, devices allowed access to higher-trust zones also have more power; they may be able to perform actions that are not allowed within lower-trust zones, such as creating or modifying enterprise data.

Figure 7-3. *As the value of an asset increases, the depth and span of controls increase, while the number of allowed devices, applications, and locations decrease. Source: Intel Corporation, 2012*

Aligning the infrastructure in this fashion provides an excellent way to right-size security controls so that security resources are utilized effectively. It also helps improve the user experience by enabling employees to choose from a wider range of devices, such as smartphones, for lower-risk activities. However, all devices should have, at a minimum, advanced endpoint capabilities that prevent more than 99% of malware from executing.

Access to zones is determined by the results of the trust calculation and is controlled by *policy enforcement points* (PEPs). PEPs may include a range of controls, including firewalls, application proxies, intrusion detection and prevention systems, authentication systems, and logging systems.

Communication between zones is tightly restricted, monitored, and controlled. We separate zones by locating them on different physical or virtual LANs; PEPs control communication between zones. This means that if one zone is compromised, we can prevent the problem from spreading to other zones or increase our chances of detection if it does spread. In addition, we can use PEP controls, such as application proxies, to provide devices and applications in lower-trust zones with limited, controlled access to specific resources in higher-trust zones when required.

The architecture includes three primary categories of security zones: untrusted, selective, and trusted. Within the zones, there are multiple subzones.

Untrusted Zones

These zones host data and services (or the interfaces to them) that can be exposed to untrusted entities. This allows us to provide widespread access to a limited set of resources from non-managed consumer devices, without increasing the risk to higher-value resources located in other zones. Untrusted zones might provide access to enterprise resources, such as corporate e-mail and calendars, or they might simply provide Internet access.

These zones are regarded as “shark tanks,” with a high risk of attack and compromise. Therefore, detective and corrective controls are needed to mitigate this risk. These controls might include a high level of monitoring to detect suspect activity and correction capabilities such as dynamic removal of user privilege.

We anticipate a need to provide controlled access from these zones to resources in higher-trust zones. For example, an employee using an untrusted device might be allowed limited, read-only access to customer data located in a trusted zone; or their device might need access to a directory server in a trusted zone to send e-mail. We expect to provide this controlled access using application proxies. These proxies act as secure intermediaries, evaluating the request from the device, gathering the information from the resource in a trusted zone, and passing it to the device.

Selective Zones

Selective zones provide more protection than untrusted zones. Examples of services in these zones include applications and data accessed by contractors, business partners, and employees, using client devices that are managed or otherwise provide a level of trust. Selective zones do not contain critical data or high-value intellectual property. Several selective subzones provide access to different services or users.

Trusted Zones

Trusted zones host critical services, data, and infrastructure. They are highly secured and locked down. Examples of services within these zones are administrative access to data center servers and network infrastructure, factory networks and devices, enterprise resource planning (ERP) applications, and design engineering systems containing intellectual property. Accordingly, we might only allow direct access to these resources from trusted systems located within the enterprise network, and all access would be monitored closely to detect anomalous behavior.

Many organizations have implemented secure high-trust zones as part of their transition to an enterprise private cloud. Implementing these zones is a key step in allowing several types of applications to be moved onto virtualized cloud infrastructure, including applications requiring high security. The security features in these trusted zones include application hardening and increased monitoring.

NEW SECURITY ARCHITECTURE IN ACTION: A DAY IN THE LIFE OF AN EMPLOYEE

This example (illustrated in Figure 7-4) describes how the new security architecture could enable an organization's sales force to access the information they need in the course of a day. At the same time, the architecture protects security by dynamically adjusting the level of access provided, based on the user's device, its location, and its capabilities for preventing malicious code, and by monitoring for anomalous behavior.

The employee travels to a customer site. The employee is using a personal smartphone with limited security features and so is allowed access only to services in untrusted zones. From here, the employee can view limited customer information, including recent orders, extracted from an enterprise resource planning (ERP)

system in a trusted zone—but only through an application proxy server, which protects the trusted zone by acting as an intermediary, evaluating information requests, accessing the ERP system, and relaying the information to the user.

If a smartphone requests an abnormally large number of customer records—an indication that it may have been stolen—further access from the smartphone is blocked. To help understand the reason for the anomalous access, there is increased monitoring of the employee's attempts to access the system from any device.

The employee reaches the customer site and logs into the enterprise network from a company-owned mobile business PC. Because this device is more trusted, the employee now has access to additional capabilities available in selective zones, such as the ability to view pricing and create orders that are relayed by an application proxy to the ERP system in a trusted zone.

The employee returns to the company's office and connects to the corporate network. Now the employee is using a trusted device from a trusted location and has direct access to the ERP system in a trusted zone.

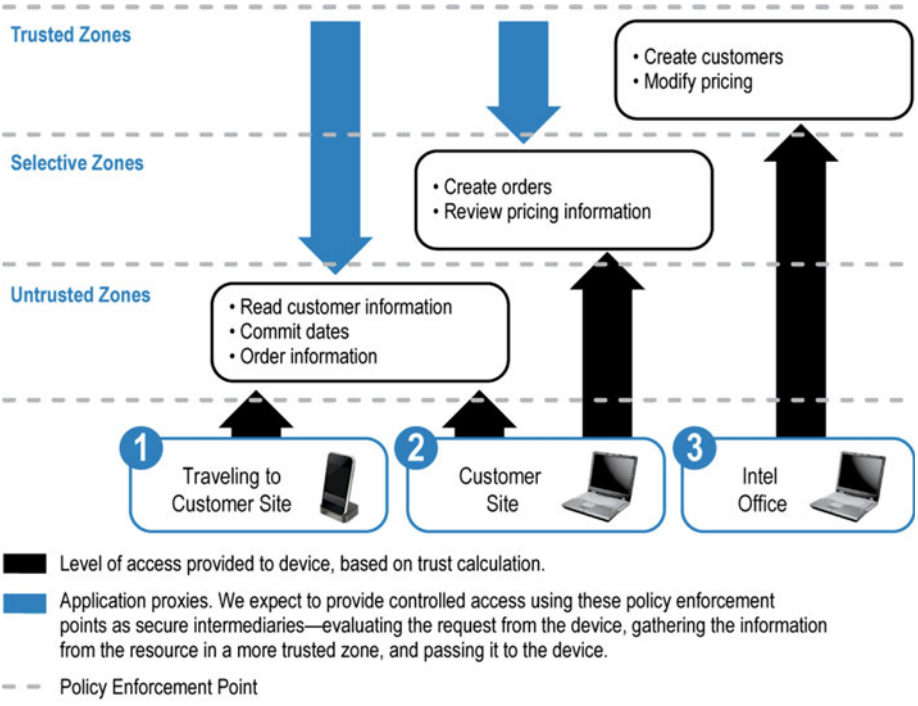

Figure 7-4. *The new security architecture dynamically adjusts the user's access to information, based on factors such as the user's device and location. Source: Intel Corporation, 2012*

Balanced Controls

Over the past decade, enterprise security has focused heavily on controls such as firewalls, signature based antivirus, and intrusion detection systems such as behavior-based anomaly detection tools. As we have seen so often in the past few years, this approach is not working. At many companies, the default belief is now that prevention is not possible and we can only correct problems after they have occurred.

However, the new security model requires that we understand the implications of the 9 Box. Preventative controls should not only stop malicious code from executing but do so in a way that lowers our overall cost of controls and with low friction. More effective prevention will reduce the alert fatigue within due to the "whack-a–mole" effect associated with over-reliance on detective (monitoring) and response controls. Detection capabilities will also be more effective because effective prevention reduces the "noise" in the environment. Over the long term, this approach will free up resources that can then be applied to other corrective controls.

By using the 9 Box to guide the control philosophy, and demanding solutions that continually shift down and to the left (reducing risk, cost, and control friction), we will be able to change the risk dynamics in the industry.

USING SECURITY ANALYTICS TO DETECT SUSPICIOUS BEHAVIOR

Almost all organizations have experienced security issues involving both external attackers and insiders, including attempts to steal intellectual property. Investigations have identified markers and indicators that are frequently associated with these events. If we can spot these indicators sooner, we can respond and mitigate the threats more quickly.

Security analytics technology can be used to detect suspicious behavior as the environment becomes more complex and attackers become more adept at concealing compromises. The technology automates the process of analyzing large volumes of data to detect and monitor anomalous activity, allowing companies to detect problems that they might otherwise miss. These capabilities are similar to those already implemented by financial institutions to prevent fraudulent credit-card transactions, and by online consumer services to prevent theft of user data.

On a large scale, logging data generated by servers and sensors across the network can be collected into a database for analysis. Security business intelligence can also be applied at the level of individual users and devices, as long as we are careful to protect users' privacy.

The balance between preventative, detective, and corrective controls will vary, depending on the security zone. In high-trust zones, we implement extensive monitoring to detect possible attempts to steal data or compromise critical systems. Redundancy within each type of control can be used to provide additional protection.

The following includes possible examples of using detective and preventative controls:

- An employee attempts to send a confidential document to an external e-mail address. Monitoring software detects the attempt, prevents the document from being sent outside the firewall, and asks the employee if he or she really intended to do this. If the employee confirms that this was intended, the document may be transmitted, or if the document is highly sensitive, a redacted version may be sent.
- Inappropriate use of a document protected with enterprise rights management technology results in revocation of access to the document.
- The system allows access to specific documents but tracks the activity. A user can download a few documents without causing concerns. However, if the user attempts to download hundreds of documents, the system slows down the speed of delivery (for instance, only allowing ten to be checked out at a time) and alerts the user's manager. If the manager approves, the user is given faster access.
- The detection of an infected system places the system on a remediation network, isolating the system and restricting access to enterprise information and applications. The system may retain some ability to access corporate assets, but all activity is closely logged to enable incident response if necessary.
- When a system is found to be compromised, we examine all its recent activities and interactions with other systems. Additional monitoring of those systems is automatically enabled.

USING MACHINE LEARNING TO IMPROVE ANTI-MALWARE TECHNOLOGY

Traditional antivirus software relies on recognizing characteristic signatures to identify specific malware families. But today, adversaries have access to off-the-shelf malware toolkits that make it easy to create custom malware variants that signature-based antivirus products cannot recognize. This custom malware sails past traditional antivirus products as if they didn't exist.

Machine learning technology provides a solution to the problem. Rather than relying on humans to identify malware signatures, machine learning technology can automatically analyze and classify hundreds of thousands of characteristics per file, breaking them down to an atomic level to discern whether an object is "good" or "bad" in real time.

The process works like this. A machine learning platform continuously collects vast amounts of data from many sources. It analyzes the data and extracts DNA-level features that the machine learning platform itself determines are unique characteristics of good and bad files. Most of these characteristics are so microscopic that human malware researchers and reverse engineers don't understand their importance. The software constantly adjusts to the real-time threatscape, thus learning to make higher-fidelity decisions. For each file, the platform assigns a threat score that is used to automate policy-based protection decisions—ignore, alert, block, or terminate file/process execution. A mathematical model encapsulating the platform's intelligence is then periodically extracted and incorporated into an anti-malware solution that is installed on endpoints. Using this solution, it's possible to stop more than 99% of malware before execution.

Users, Data, and the Internet of Things: The New Perimeters

The concept of balanced controls also extends to the protection of users and data. Traditional network security boundaries are dissolving with the proliferation of new devices and users' expectations that they should be able to access information from anywhere at any time. Users are under direct assault from a barrage of attacks designed to trick them into taking actions that can compromise the information on their devices or on enterprise systems. These trends mean that we need to think more broadly about how we protect information, as well as the users of this information.

While we continue to implement enterprise network controls, such as perimeter defenses and the detective controls described earlier, we need to supplement these controls with a focus on the users and on the primary assets we are trying to protect such as intellectual property. The new architecture therefore expands our defenses to two additional perimeters: the data itself and the users who have access to the data.

Data Perimeter

Important data should be protected at all times: when it is created, stored, and transmitted. This becomes increasingly challenging as we move data to more and more devices and let more people access it. How do we protect information when it's located outside the physical perimeter on a personal device?

One approach is to use technologies that closely integrate protection with high-value data so that the data remains protected as it moves to different devices and locations. Technologies such as enterprise rights management and data leak prevention can be used to watermark and tag information so that we can track and manage its use. With enterprise rights management, the creator of a document can define exactly who has access rights throughout the life of the document and can revoke access at any point. Data loss prevention is used to tag documents, track their movements, and prevent transfer outside the organization if necessary.

User Perimeter

As I described in Chapter 5, people are part of the security perimeter, and we need to treat them as such. Users can become security risks for a variety of reasons. They are targeted more frequently in social engineering attacks, and they are more vulnerable to these attacks because their personal information is often readily available on social networking sites. They may also click malicious links in e-mail, download malware, or store data on portable devices that then are lost. A combination of training, incentives, and other activities can help instill information security and privacy protection into the corporate culture and successfully encourages employees to own responsibility for protecting enterprise and personal information.

Internet of Things

The Internet of Things can be viewed as an extension of the user and data perimeters into new connected devices and systems such as cars, wearables, and smart buildings. IoT devices should be included in the security architecture; for example, the trust calculation could be applied to access from IoT devices, so that the security of the device is a factor in determining the level of access provided. For machine-to-machine communications, each communicating machine can be considered conceptually as analogous to a user; the security architecture focuses on preventing, detecting, and responding to behavior that it identifies as anomalous.

Conclusion

This chapter describes a new control architecture designed to support the Protect to Enable mission. With this approach we can lower risks, lower costs, and lower control friction. It will also allow for faster adoption of new services and capabilities because it helps prevent risk and improve survivability. I believe that this architecture can be used to meet a broad range of evolving requirements, including new usage models and threats. The architecture's flexibility and granular trust model should also make it easier for the security team to identify and contain anomalous activity that signals potential insider threats. By publishing information about the architecture, I hope to encourage others to take advantage of these ideas. I also hope that making this information available will stimulate more discussion and ideas, and that others will build on these concepts to create further innovations that benefit all of us.