VDC Research
Insights for the Connected World

The Evolution of PKI for IoT

Establishing Trust in an Untrusted Environment

Exclusive License to Distribute: Entrust Datacard™

By Steve Hoffenberg, Director, with Chris Rommel, Executive Vice President

CONNECTIVITY IMPLIES VULNERABILITY

In years past, device functionality was enough to sell most embedded products without much concern for cybersecurity. Of course there were exceptions, such as in critical infrastructure, aviation, and military, for which security was always of importance. But today's environment has evolved on several fronts. First, organizations across nearly all markets are demanding Internet connectivity to monitor and control devices as well as to aggregate and analyze data. Second, the magnitude of security threats has exploded, driven by highly sophisticated hackers including organized criminal gangs seeking financial returns, creating a constantly evolving threat landscape. Third, the increasingly complex nature of connected systems makes them ever more challenging to protect. The more complex a system, the more potential vulnerabilities it may contain. And fourth, the data generated by connected devices represent an asset that is becoming increasingly valuable for organizations to derive insights about their operations. In addition, an increasing portion of connected systems involve devices that can potentially damage people or physical property if compromised.

In some markets, such as industrial automation, device makers and systems integrators need to meet the occasionally contradictory security requirements of both IT and operations departments. And in markets involving sensitive personal or financial data, such as medicine and banking, government regulations mandating device and data security may change periodically, requiring security modifications to existing systems in the field. Even those who have successfully handled device security in the past may be faced with new threats and vulnerabilities introduced through cloud-based device control and data storage. Embedded devices are no longer standalone entities, they are elements of connected systems, the security of which may be only partially under the control of the device maker or user organizations.

"No device or system connected to the public Internet should be considered impenetrable, because impenetrability is impossible to prove."

No device or system connected to the public Internet should be considered impenetrable, because impenetrability is impossible to prove. While one can hope that no attack will succeed, it is best to assume that some will. Long gone are the days when systems administrators assumed that if they could control the perimeter of a network, they could prevent all unauthorized access. Today, insider threats pull security boundaries inward, while cloud services project them outwards. The perimeter itself can be both flexible and indistinct. Perimeter security remains necessary, but it is not sufficient.

DEFENSE IN DEPTH

Security policies are necessary to enforce levels of protection for organizations that use, sell and operationalize IoT devices. Collectively, such policies and the measures to enforce them help those organizations mitigate the operational and financial risks that arise when connecting their systems to other IoT systems and platforms.

"Security policies and measures help organizations mitigate the operational and financial risks that arise when connecting their systems to the IoT."

But if one cannot ensure impenetrability, how should user organizations protect their systems? Many organizations periodically conduct risk assessments, which factor in both the business value of assets and the technical difficulties (including the costs) of protecting those assets. Once the risks have been understood and evaluated, the organizations can develop appropriate multifaceted "defense in depth" strategies, essentially reducing risk at every level of a system, from the operational systems, to the users and their devices, to the local area network and any cloud services. The

technical controls of defense in depth utilize a layered approach, such that if one line of defense is breached, subsequent lines of defense will continue to hinder efforts to further penetrate the system and mitigate the damage an attacker could inflict by disrupting operations or exfiltrating data. Each layer of defense further reduces risk.

In an IT network, the layered approach typically includes measures to block or limit unauthorized access to system resources (including devices and data files) at a variety of levels, such as:

- user authentication (in some cases multifactor) that prevent unauthorized users from accessing the system, and are configured to permit or deny authorized users' permissions to access and utilize specific system resources
- firewalls at the interfaces between local network segments and the Internet to disallow entry into the network from unauthorized sources
- network-based intrusion detection and/or intrusion prevention systems to detect and/or block anomalous behavior on the network
- segmentation of the network such that an unauthorized incursion into one segment will not have access to resources on other segments
- "honeypot" devices to lure attackers so their malicious behavior can be analyzed
- host-based intrusion detection and/or intrusion prevention software to detect and/or block anomalous behavior on the individual devices
- anti-virus and anti-malware software to detect known malicious software on the devices
- encryption of stored data (data at rest)
- encryption of data during communications (data in motion)
- a unique permanent identity for each endpoint device
- a protected hardware region in each endpoint, to provide a root-of-trust, a secure storage enclave, and/or an execution environment for cryptographic functions

In an Internet of Things network, these measures can be more challenging to apply, due to:

- the plethora of operating systems that can be found on IoT devices
- the limited processing power and memory capabilities of many IoT devices
- the extensive range of communications protocols that may be utilized by IoT devices
- the large number of IoT devices that may be part of a single system
- the potential for software of unknown provenance on the IoT devices
- the lack of a human-oriented user interface on many IoT devices
- the longer lifecycle of many IoT devices compared to IT devices
- the lack of available software patches for security vulnerabilities, and/or the lack of applying available patches on many IoT devices
- the unfamiliar nature of many IoT devices to IT systems administrators
- the safety implications of IoT devices that have direct physical impact on their environments (A wayward IoT device could, for example, shut down a power plant, disable traffic lights, or inject a patient with a lethal dose of medication.)
- the lack of standardization when configuring and deploying IoT devices.

In addition to the defensive measures commonly used on IT networks and their endpoints, an IoT network can include security measures tailored to their device's embedded systems, such as:

- an automated process for securely provisioning new devices on the network
- an application whitelist that only permits specific known software to run
- embedded hypervisors to run multiple instances of operating systems, each responsible for hosting different system functions
- sandboxes to isolate individual applications
- a secure process for automated reception and installation of over-the-air (OTA) firmware updates.

THE ROLE(S) OF PKI IN IOT

Many of the measures to secure IoT systems rely on authenticating the identity of users and devices, to subsequently determine whether they are authorized to perform certain actions or access specific system resources. Without the ability to trust that people and things are who or what they say they are, the system cannot function securely. The ability to implement system controls for authorization, encryption, and digital signatures relies on the cornerstone of establishing and authenticating the identity of people, systems, and things. Each of the things needs its own identity so it can be tracked as it moves through a complex supply chain, and to enable the varied and disparate constituents of the IoT ecosystem to securely handle data and commands to and from that specific device.

With respect to identity, the people challenges for IoT aren't necessarily more problematic than for IT systems, but the Internet of Things presents two particular problems: the Internet, and the Things.

> *"With respect to identity, the Internet of Things presents two particular problems: the Internet, and the Things."*

On the Thing side, most IoT devices run unattended, such that no human user is directly operating or monitoring them. The sheer number (in the billions) of devices that will be deployed means that the methods of initially provisioning device identities and later authenticating them both must be highly automated. It is economically unworkable to have humans perform these tasks at IoT scale.

The IoT platforms which receive data from these devices must be able to verify the identity of the devices sending the data, and that the data hasn't been improperly manipulated in transit.

On the Internet side, the IoT is still untamed, with more than a thousand IoT platforms available to manage devices, aggregate and analyze data, display dashboards, trigger alerts, and produce big data insights. Most IoT platforms offer security features, but those features vary from one platform to the next, and they may not be as strong as those that an organization would implement on its own network.

When IoT devices communicate with a platform to send data or receive commands, they must be able to verify that they are communicating with the correct service, as opposed to malicious systems that may be trying to spoof the identity of the service. And device owners must be certain that their data—which is now beyond their local network management and firewalls—is secured against unauthorized access from anywhere on the Internet.

In addition, an IoT system may need to communicate with more than one service, including device makers' platforms for predictive maintenance, repair services, or firmware updates; platforms for cellular or LPWAN connectivity management; and platforms of other business partners. In each case, the IoT system must be able to trust that it is communicating with the correct party. This is especially true when the platform will be providing firmware updates, as the integrity of the firmware and the identity of its development organization (independent of the identity of the sending platform) must be verifiable.

Fortunately, solutions leveraging Public Key Infrastructure (PKI) can help with all these issues.

PKI Functions

The core function of PKI services is to generate and issue pairs of keys (one private, one public) for asymmetric cryptography. As the names imply, a private key is kept private by the entity to which it is issued, while a public key can be publicly distributed. The private and public keys are mathematically related, but with today's computing technology it is considered "computationally infeasible" to derive the private key from the public key.

These asymmetric key pairs are most commonly used in two ways:

1) A digital message (or data) is encrypted using the private key, creating a digital signature such that any entity with the corresponding public key can verify that it was created by a holder of the private key.
2) A digital message (or data) is encrypted using the public key such that only a holder of the corresponding private key can decrypt it, thus keeping it secret from anyone or anything else.

Separately or together, these functions provide the ability to uniquely identify a sender or recipient and enable the exchange of encrypted data. Both sides of a communication link (e.g. two IoT devices, or an IoT device and a cloud platform) can usually both send and receive, thus each side needs its own key pair so that they can mutually authenticate each other prior to sending actual message data. (In practice, asymmetric cryptography is computationally intensive, so it is most often used to securely establish initial communications through which symmetric cryptographic session keys are distributed. As such, the remainder of the communication session proceeds using less computationally intensive symmetric encryption. Each time communication is re-established, asymmetric cryptography is again used to distribute new symmetric session keys. Typically, this nuance between asymmetric and symmetric keys/cryptography is wrapped inside of protocol implementations that alleviate the system developer from engaging at this level.)

PKI services may also generate truly random numbers used as symmetric encryption keys or in "digital birth certificates" that uniquely identify each unit of a product.

For most PKI services, key generation and storage take place in Hardware Security Module (HSM) appliances, which are specially designed to perform those functions with the utmost security.

While a large variety of use cases exist for PKI in the enterprise IT environment, in the Internet at large, PKI is primarily used for Secure Socket Layer (SSL) or Transport Layer Security (TLS) protocols to secure communications between browsers and websites, and for Secure Shell (SSH) protocol to authenticate users and host computers. In the IoT, PKI is also used for creating and authenticating device identity, securing firmware updates, tracking data sources, encrypting communications, granting access and usage rights with IoT cloud platforms, and other capabilities.

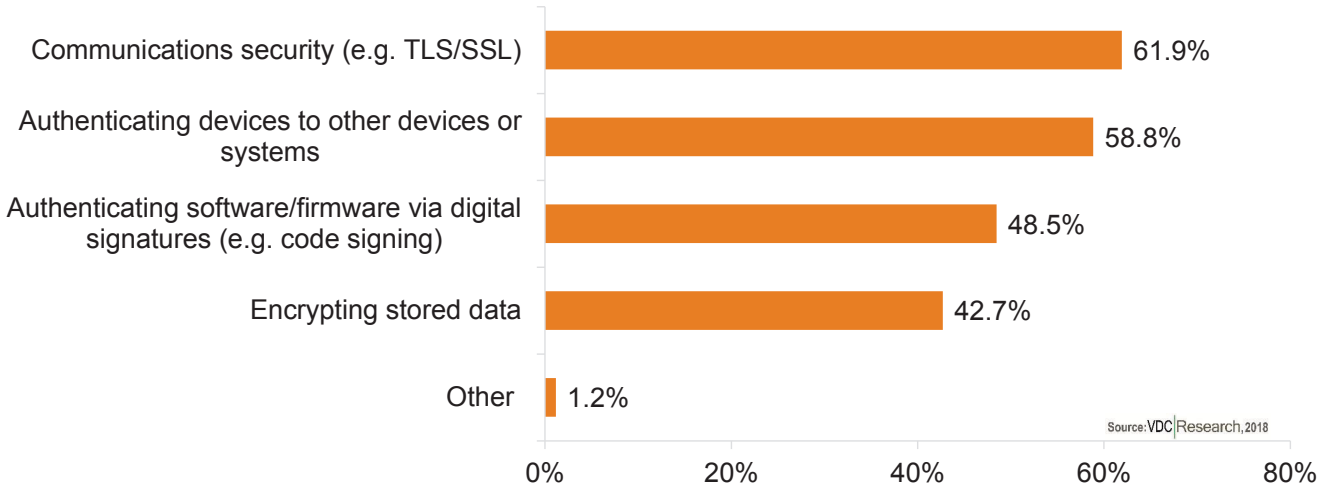
Typically, a public key is delivered in the form of a digital certificate, such as an X.509 certificate, which includes the key validity period, the asymmetric cryptographic algorithm used, the identity of the issuing organization, and other relevant information. The certificate is digitally signed by a trusted certificate authority (CA), which encrypts it using the CA's own private key. By using the CA's public key, the recipient can verify that the digital certificate is from that CA.

Similarly, a device-maker's private key can be used to digitally sign software or firmware, so that recipients can confirm it's genuine prior to installation. The digital signature also includes a hash value which recipients can use to verify that the software or firmware has not been modified since it was signed by the device maker.

"In a recent VDC Research survey of embedded systems engineers, 49% of respondents said their most recent IoT project used PKI keys and/or digital certificates."

In a recent VDC Research worldwide survey of embedded systems engineers, 49% of respondents said their most recent IoT project used PKI keys and/or digital certificates. They used these technologies for a variety of applications, as shown in the figure below.

Uses of PKI Keys and Digital Certificates in Most Recent IoT Project (Percent of Respondents)



Note: Percentages sum to more than 100% due to multiple responses.

Practical and Emerging Challenges for Securing the IoT

Traditionally, most PKI vendors have provided solutions primarily for broader key and certificate needs in website/browser and general enterprise/IT applications, with only a relatively small portion of their key/certificate business derived from IoT, often applying general purpose PKI solutions in IoT use cases. Nevertheless, typical solutions for enterprise PKI can be challenging to implement for organizations rolling out IoT systems today, for several reasons.

- **Security immaturity for IoT** – Many first-mover industries expecting to benefit from IoT lack adequate knowledge of embedded security practices within their organizations. The operational staff may have been accustomed to securing their systems by walling them off from external connections. To add Internet connectivity requires assistance from experienced security professionals to assess the risks and levels of controls needed to protect the system from IoT threats. However, the shortage of skilled labor for security positions limits the ability to build these teams internally, and the operational staff often has little experience working with the organization's own IT security people.
- **Device heterogeneity** – Most enterprise/IT systems are comprised of standard Windows PCs and perhaps Apple Mac laptops, along with Android and iOS mobile phones and/or tablets. In the IoT, devices may be running any of 70 or more operating systems (including some that are open source) on a wide variety of processor chip hardware, with different memory capacities, peripheral capabilities, and communications protocols. Many of the devices have limited processing and memory that precludes them from running elaborate security measures. No single embedded security solution is universally applicable, and each of these IoT device configurations may require customized security software or hardware. In addition, security monitoring for the network to which they are connected needs to have the ability to characterize and distinguish abnormal from normal behavior and data for each type of device.
- **Edge computing** – Gateways and other intermediary computing devices add a network architectural tier not present in most enterprise IT systems. Conventional security solutions may not be designed to handle the security requirements of these important system elements. At the same time, edge computing systems offer the opportunity to add an extra layer to system defenses, and can be especially useful for securing legacy components not originally engineered with protection against threats from the Internet.
- **Real-time requirements** – In many IoT systems, especially those in industrial and critical in-frastructure environments, protecting physical assets and the humans operating them may require systems to react to safety hazards extremely quickly (i.e. in milliseconds). Cybersecurity measures cannot add excess latency that would make response times unsafe. This often precludes cloud-based security solutions from directly serving gatekeeping roles on a local system, although such cloud-based solutions can be particularly useful for subsequent analysis of security incidents.
- **Safety vs. Security** – Beyond real-time response issues, some security controls may be at odds with safety requirements for IoT systems. The strictest security protocols might warrant cutting off all external communication to a system—or even shutting down the entire system—when a breach is detected, but doing so could risk the safety of the system or its users. In a mining operation, for example, workers may need communications and functioning equipment to get out of the mine. Or in an oil refinery or other chemical process facility, unexpected cut off of communications or control systems could risk processes running out of control. For reasons such as these, most industrial IoT systems are designed to "fail open" (continuing operations), rather than "fail closed" (cutting off access or shutting down) as would be the case for many enterprise/IT systems to ensure data and asset security. Typically, such IoT systems include intrusion detection systems that alert operators in the event of a suspected security breach, but not intrusion prevention systems that directly implement system responses.

IoT technology and the cyber threats that attack it are constantly evolving. Any long term security strategy must include the ability for security controls to evolve along with the technology and the threats. Throughout those changes, device identity and authorization will remain crucial elements of building and maintaining security for IoT systems.

Next Generation Solutions for IoT Security

In recent years, and in effort to meet organizations' needs, vendors have been developing highly automated solutions specifically for IoT security. These solutions provide toolkits and dashboards that hide some of the complexity and ease deployment of digital identities.

"To meet organizations' needs, in recent years vendors have been developing highly automated solutions specifically for IoT security."

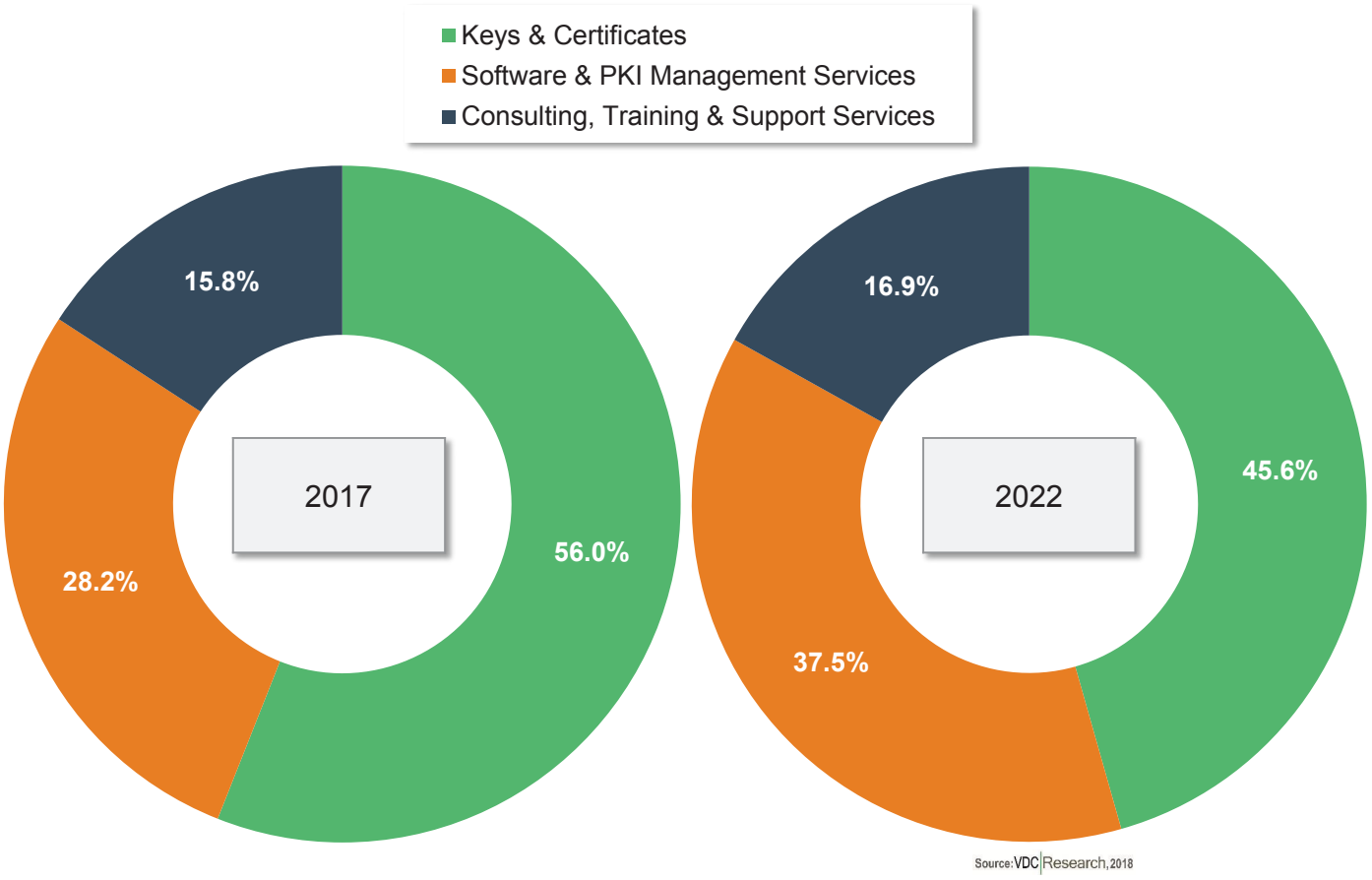
In theory, basic IoT customer needs for device identity, encryption keys, and digital certificates can be fulfilled by PKI services originally designed for enterprise/IT applications. However, such services do not scale well to the high volume and operational requirements of the IoT. It is simply too labor-intensive to perform these tasks manually. Further, organizations are increasingly seeking vendors that offer managed services. Although these organizations still need to have internal resources devoted to security, they don't want to rely on hiring their own entire team of IoT security specialists.

IoT-oriented offerings may contain a range of turnkey managed services, including any of the following:

- **Identity provisioning on the manufacturing line** – At the factory where IoT devices are built, a PKI provider can set up a physically secure area containing one or more Hardware Security Module (HSM) appliances or a virtual pipeline from an HSM-in-the-cloud, to generate a unique identity for each unit produced, then deposit that identity into a secure enclave in the device hardware. In addition to serving the need to rapidly provide identities to the devices, such a service can help control inventory and ensure the integrity of the supply chain, i.e. that all units being deployed are genuine (non-counterfeit), and that a contract manufacturer isn't building extra "gray market" units that are being sold through unofficial channels.
- **Key and certificate management** – software and storage solutions to help device makers and their customers securely keep track of large numbers of encryption keys and digital certificates, including which private keys correspond to which devices, as well as handling expirations of temporary keys, provisioning of new keys, and revocations of compromised keys. These solutions can reside in an edge gateway or server, in an on-premises datacenter, or as a cloud service.
- **Onboarding services** – assigning newly provisioned devices to IoT platforms and provisioning privileges to access other local or remote system services in accordance with organization policies.
- **Digital signatures** – computing a hash value for a firmware package or data message, then encrypting the hash value with an organization's private key. This enables recipients of the software to verify both the source and integrity of the firmware or data.
- **Web portal** – web-based interface for administrators to monitor and control an organization's keys, certificates, and other PKI services.
- **Professional consulting and training services** – PKI providers can help organizations design and implement security for their IoT devices and systems, and can train employees about the proper use of keys and certificates to maintain security of IoT devices and data.

Additional PKI-related services such as these reduce the time and effort that device makers, installers, and operators must spend integrating and managing their devices in customer environments. And the services can evolve along with the threats. The benefits of these services are sufficiently strong that the market for PKI in IoT is shifting more toward these value-added solutions. As shown in the exhibit below, VDC forecasts the revenue share for software and services to increase by approximately 10% from 2017 to 2020, such that the sale of keys and certificates will no longer represent the majority of PKI revenue.

Worldwide Revenue for PKI for IoT Keys, Certificates and Related Software and Services 2017 & 2022, Share by Product Category (Percent of Revenue)



PKI Options and Alternatives

Depending on its needs, an organization could utilize PKI services for all of its key and certificate needs, or just a portion, or none at all. For example:

- A company could look at IoT specific security solutions that leverage PKI, but have been designed specifically for the technical and organizational challenges of operational environments. If the trust models require interoperability with IT systems, the organization should make sure these solutions support this integration.
- A company might use a Certificate Authority to issue its keys and certificates, then write its own software to manage and utilize those keys. This can be a viable approach if the company has software development resources with sufficient experience in cryptography and key management.
- It could effectively act as its own Certificate Authority, generating and issuing its own keys. Again this could be a viable approach, but it requires the company to have even greater expertise in cryptography. In addition, third parties may or may not trust the certificates unless the company is a major multinational organization that is already known and trusted.
- It could join a Pretty Good Privacy (PGP) "web of trust," essentially replacing the role of Certificate Authority with designated other parties that agree to trust each other's certificates. This can work if the company only plans to exchange keys and certificates with other members of the web of trust. Other entities have no basis on which to trust such keys and certificates, and this method would be extremely difficult to scale to IoT volumes.

Conclusion

Overall, nearly all IoT device makers and users leave the generation of keys and certificates to recognized Certificate Authorities. And value-added PKI services simplify the processes of securely deploying and managing devices, saving time and money in the long run. Using these services enables IoT organizations to focus their attention on what they do best, building and using great IoT products.

"Using PKI services enables IoT organizations to focus their attention on what they do best, building and using great IoT products."

ABOUT THE AUTHORS

Steve Hoffenberg

Steve Hoffenberg is a leading industry analyst and market research professional for Internet of Things technology. He has more than two decades of experience in market research and product management for technology products and services. Prior to joining VDC, he spent 10 years as Director of Consumer Imaging and Consumer Electronics Research at the firm Lyra Research, where he led industry advisory services providing extensive market research on consumer technology trends, user adoption, market sizing, marketing strategy, and competitive analysis for major consumer electronics manufacturers. Previously, he worked in product management for electronic design companies that developed and licensed embedded digital imaging and audio products. Steve holds an M.S. degree from the Rochester Institute of Technology and a B.A. degree from the University of Vermont.

Contact Steve:
shoffenberg@vdcresearch.com

Chris Rommel

Chris Rommel is responsible for syndicated research and consulting engagements focused on development and deployment solutions for intelligent systems. He has helped a wide variety of clients respond to and capitalize on the leading trends impacting next-generation device markets, such as security, the Internet of Things, and M2M connectivity, as well as the growing need for system-level lifecycle management solutions. Chris has also led a range of proprietary consulting projects, including competitive analyses, strategic marketing initiative support, ecosystem development strategies, and vertical market opportunity assessments. Chris holds a B.A. in Business Economics and a B.A. in Public and Private Sector Organization from Brown University.

ABOUT VDC RESEARCH

Founded in 1971, VDC Research provides in-depth insights to technology vendors, end users, and investors across the globe. As a market research and consulting firm, VDC's coverage of AutoID, enterprise mobility, industrial automation, and IoT and embedded technologies is among the most advanced in the industry, helping our clients make critical decisions with confidence. Offering syndicated reports and custom consultation, our methodologies consistently provide accurate forecasts and unmatched thought leadership for deeply technical markets.

Located in Natick, Massachusetts, VDC prides itself on its close personal relationships with clients, delivering an attention to detail and a unique perspective that is second to none.

P 508-653-9000
info@vdcresearch.com