Independent market research and competitive analysis of next-generation business and technology solutions for service providers and vendors

HEAVY READING
WHITE PAPER

Evolving the Mobile Security Architecture Toward 5G

A Heavy Reading white paper produced for Cisco Systems Inc.

AUTHOR: PATRICK DONEGAN, CONTRIBUTING ANALYST, HEAVY READING

INTRODUCTION

While the mobile threat landscape is becoming more challenging, only a subset of leading mobile operators is keeping up with the software programmability that's needed to help their mobile security architecture meet those threats. In one sense, 5G will provide better-than-ever continuity between the former and new generations of technology. But the growing consensus that 5G must be based on an automated, shared network functions virtualization infrastructure (NFVI) means that from 5G onward, mobile operators "run out of road" regarding the option of implementing software programmability on a limited scale.

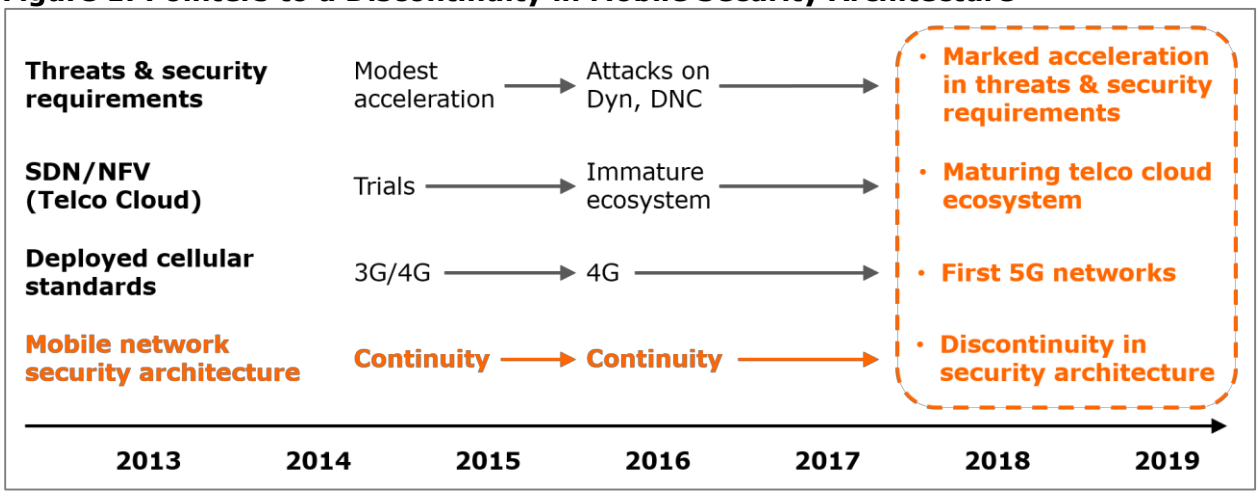
This paper examines trends in the mobile threat landscape, software programmability and cellular standards to point to new directions in network security architecture that mobile operators should be considering.

KEY TRENDS DRIVING MOBILE SECURITY ARCHITECTURE

Despite the ongoing transformation – one might almost say upheaval – in the IT and networking technology landscape, a mobile operator's network security architecture is still driven by essentially the same three fundamental factors as when digital mobile networks were first launched 25 years ago. These are:

1. **The Threat Landscape:** Who or what can compromise the operator's ability to provide a network that assures confidentiality, integrity and availability – and how?
2. **The Overall IT & Networking Technology Landscape:** What are the latest security and networking tools, resources and best practices that are available to the mobile operator? How can these be leveraged for a network security architecture that is fit for purpose for the next several years?
3. **Cellular Standards:** What do the business models and architectural assumptions embedded in upcoming cellular standards imply in terms of specific requirements on the security architecture, as well as specific new vulnerabilities that need addressing?

Figure 1: Pointers to a Discontinuity in Mobile Security Architecture



Source: Heavy Reading

As depicted in **Figure 1**, trends, events and roadmap assumptions emerging across each of these three inter-related ecosystems appear to be converging. Furthermore, they seem to be converging in a way that points to a potential discontinuity how mobile operators deploy and deliver security services – for their enterprise infrastructure, their network infrastructure and their enterprise customers' infrastructure.

2016: A LANDMARK YEAR IN SECURITY THREATS

You don't need to follow cybersecurity closely to recognize that the threat landscape is posing greater and greater risk to more and more businesses and individuals. Unfortunately, 2016 was something of a watershed year in cyber-attacks. Last year saw a marked acceleration in the transition from cyber intrusions being carried out by traditional amateur hacking to what now amounts to the weaponization of the Internet by a host of nefarious actors. These actors range from small-time fraudsters and hacktivists to organized crime and hostile nation-states.

The Dyn Attacks Provide a Warning Around IoT

The October 2016 distributed denial of service (DDoS) attacks on Dyn were notable for the scale of the outages caused to the likes of PayPal, Twitter and CNN. They were also notable because, having been executed using hacked cameras and DVRs that had become part of the Mirai botnet, they fired a huge warning shot across the bows of industry, consumers and governments worldwide. They provided a stark warning that the threats posed by an insecure Internet of Things (IoT) is no longer the stuff of some science-fiction nightmare, but a real-world fact of life – today, right now.

During 2016, America's intelligence agencies also concluded that another nation-state mastered the hacking of the Democratic National Committee's (DNC's) email servers raising the specter that a nation-state may have materially influenced the election of President of the United States.

Mobile Threats Within the Threat Landscape: The Breaching of the iPhone

While most mobile malware exploits are to be found in the Android ecosystem, among the most important mobile security milestones of the last couple of years has been the breaching of the iPhone ecosystem, whose security credentials have been much praised relative to Android's.

- **The "XcodeGhost" malware for iOS**, developed by commercial malware authors was discovered in the App Stores in September 2015. It enables unauthorized third parties to access user information and impacted some users, primarily in China.
- **The "Pegasus" exploit, a zero-day attack on iOS** enabling the iPhones of high-profile individuals to be successfully hacked into and tapped, was discovered in August 2016. UAE human rights defender, Ahmed Mansoor, is among those users known to have been targeted.

Figure 2 depicts just how far the attacker community has come in bringing outages, exposure of customer data, phone tapping and theft of both money and intellectual property to

mobile device users in the consumer and enterprise context – as well as to the mobile operators that serve them.

Mobile malware features prominently in media coverage of mobile security threats, but not all mobile threats involve malware. You don't need to get mobile malware onto a device in order to use the user's number for a phone call or text, forming one leg of a multi-vector attack. High-profile examples include spear phishing or whale fishing attacks that have seen many businesses defrauded. You also do not need to get mobile malware onto a device in order to use it in the enterprise context to access enterprise applications that the device – or the individual who owns it – is not authorized to access.

Figure 2: 2016 – Landmark Year in Mobile Security Threats

Month	Incident or Landmark	Significance
April	TV demo on U.S "60 minutes"	Proves SS7 vulnerabilities allow mobiles to be tapped.
April	Discovery of "Dresscode" malware in Google Play	100,000 downloads of data exfiltration malware
August	"Pegasus" zero-day iOS malware discovered.	High profile iPhone users known to have been spied on.
October	Probably inadvertent T-DOS attack on U.S. Public Service Answering Points (PSAPs) spread by social network links.	911 emergency call centers flooded with bogus calls from affected iPhones.
October	DDoS attacks on DNS provider Dyn, creates huge outages on major websites.	First of several huge attacks leveraging Mirai botnet that infects "things" in the IoT.*
October	African mobile operator suffers severe DDoS attacks	Company claims attack paid for by telecom competitor.
November	EU mobile operator suffers data breach	Mobile user customer records exposed
December	CTIA states U.S smartphone infection rate now 2-5%	U.S smartphone infection rate up from ~2% in 2014-2015

* These weren't actually connected to a mobile network in this instance.

It's also worth noting that **Figure 2** references only those security breaches, incidents and milestones that made it into the public domain in 2016. It references those that are discoverable by threat researchers focused on investigating vulnerabilities in the smartphone operating systems and those that make it into the media. Hence, **Figure 2** doesn't include undisclosed network outages and degradations brought about by attacks exploiting vulnerabilities in the network infrastructure. There are likely to have been some of these, although how many is inherently impossible to estimate.

A STEP CHANGE IN SOFTWARE PROGRAMMABILITY

In relation to both network security functions and network functions more generally, there is near-universal alignment in the communication service provider (CSP) community that software programmability is key to rendering the mobile network flexible, scalable and automated enough to meet customer expectations while sustaining or improving business performance.

Equally, there is near-universal alignment that this roadmap must be based on adopting many of the cloud compute, storage and networking principles implemented by Webscale Internet companies, such as Google and AWS. This roadmap comprises the many software programmable building blocks of software-defined networking (SDN) and network functions virtualization (NFV) that have dominated strategic planning of CSP network architecture and CSP infrastructure products for the last three or four years.

The reality of the last three or four years is, however, that while some large, well-resourced, operators are making good progress in cloudifying their networks, most CSPs aren't moving fast enough. So, while AT&T's John Donovan was able to tell AT&T's own annual security conference last October that he expected "very close" to 30 percent of the company's network functions to be virtualized by the end of 2016, a large majority of other operators around the world are a long way from that milestone.

Heavy Reading runs a twice yearly "Virtualization Index" survey across the same pool of 50 senior CSP executives asking them for an update on the state of NFV implementation in their companies. Some of the highlights of the latest November 2016 survey support the assertions made above as follows:

- Only 18 percent of CSP respondents stated that their companies have completed testing and deployment of all the virtualized functions in high priority areas.
- Fully 62 percent of CSP respondents stated that 10 percent or less of the virtualized functions their company has identified as high priority areas are actually in live production networks.

"The Shift Toward Cloud-Native Requirements"

With a handful of exceptions, the missing ingredient in most NFV deployments to date is the all-important automation and orchestration piece. Yes, virtualized network functions (VNFs) have been deployed on standard servers. Frustratingly, though, the cost savings and speed to market with new services that is promised from greater automation and orchestration is still proving elusive.

As a result, the last year or so has seen a shift in the requirements of leading operators toward so-called "cloud native" VNFs. Operators such as Vodafone, Deutsche Telekom and AT&T have established their leadership in part by slowing or ceasing investment in first-generation VNFs, many of which were ported straight out of appliances into large, stateful, monolithic virtual machines (VMs).

These CSP leaders have had the foresight, drive and resources to turn their back on these first-generation VNFs because, in their view, they consume too much hardware in their pursuit of high availability. Over the last year or so, they have begun insisting on second-generation, cloud-native VNFs that are able to navigate failures themselves according to the well-established best practice developed by the IT sector for cloud applications.

As we show in the next section there are good grounds for seeing 5G as a supportive – or even galvanizing – factor in driving the software programmability roadmaps of the CSP community. This has the potential to materially enhance their cost and revenue outlook, as well as drive new thinking around their mobile network security architecture.

5G: EVOLUTION IN CELLULAR, REVOLUTION IN SECURITY?

This section examines how 5G can serve alongside the heightened threat landscape and the urgent imperative of greater software programmability to sharpen the way operators think about leveraging the cloud to drive better mobile security outcomes. This means looking again not just at how mobile operators leverage the cloud to deliver security services to their customers, but also at how they leverage the cloud to provide security services for their own network and IT assets.

Viewed in the narrow context of different generations of cellular technology, 5G certainly looks a great deal more evolutionary than revolutionary – more continuity than discontinuity. For example:

- The first 5G radios will be capable of being plugged into existing 4G cores (a backward compatibility at the network level that wasn't available to new 3G radios or new 4G radios).
- 5G standards will be designed to formally integrate and support a proliferation of other high- and low-frequency, licensed and unlicensed radio technologies developed beyond the confines of the 3GPP cellular world.

"Cloud Native" Shared Infrastructure in an Era of Internet "Weaponization"

However, viewed from a broader perspective – including the perspective of the evolution of the mobile network security architecture – 5G does have the potential to be revolutionary, or at least transformational:

- **5G is the first cellular generation to launch in the era of the "weaponization" of the Internet.** It's worth recalling that when 2G and 3G were launched most security threats were posed by a small subset of insiders using mostly manual attack vectors, as well as an elite of ultra-sophisticated encryption experts. Even back in 2009 when 4G was launched, the security threat landscape was nothing like today's. 5G is the first cellular generation that will be launched in the era of the Internet being "weaponized."
- **A shared infrastructure leveraging a NFVI**. Consistent with the shift to cloud-native thinking and the need to accelerate automation and orchestration, the assumption now widely shared across the industry is that the 5G network has to be built around cloud-native principles of an automated, orchestrated, shared infrastructure, leveraging an NFV infrastructure (NFVI). This idea may not appear all that remarkable now, but as recently as a year ago, the idea that independently-operated, virtualized servers were a plausible NFV path for some time to come was still fairly common currency.

Of course, very diverse views continue to coexist among many key players in the ecosystem as to precisely what the details of a cloud-native, shared infrastructure architecture for 5G should look like. Nevertheless, the mere fact that there is a strong consensus around this high-level aspect is itself significant. It's significant from the perspective of the mobile network architecture as a whole, as well as from the perspective of the evolution of the mobile security architecture.

Increasingly, 5G is shaping up to be a key inflection point in the history of telecom networks. 5G appears to be the point where wireline and wireless, mobile and fixed networks have the potential to finally converge. 5G could offer up what could appear to the user as almost limitless capacity and performance, albeit according to one or more among several candidate scenarios.

5G appears as the key upcoming staging point in the software programmability roadmaps of CSPs in that they can continue to retro-fit NFV into their existing networks according to suboptimal, semi-cloudified models near term if they want. But for those operators that aren't yet aggressively pursuing a fully-automated, fully-cloudified approach to software programmability, 5G is increasingly looking like the point at which that model runs out of road.

To these operators, 5G increasingly appears as a line in the sand where software programmability is concerned. The industry's leaders are set on a course that's already consistent with that. Its followers now know that there is no 5G future for them without aligning with that trend. And, by implication, that is going to create commercial pressure for those operators to begin to align with that long-term roadmap sooner rather than later, in advance of their first 5G launches.

Additional Security Requirements With 5G

It's worth adding, of course, that 5G will also bring a slew of its own new security requirements. It's beyond the scope of this paper to attempt to address all of them, but a couple of illustrative examples here are useful:

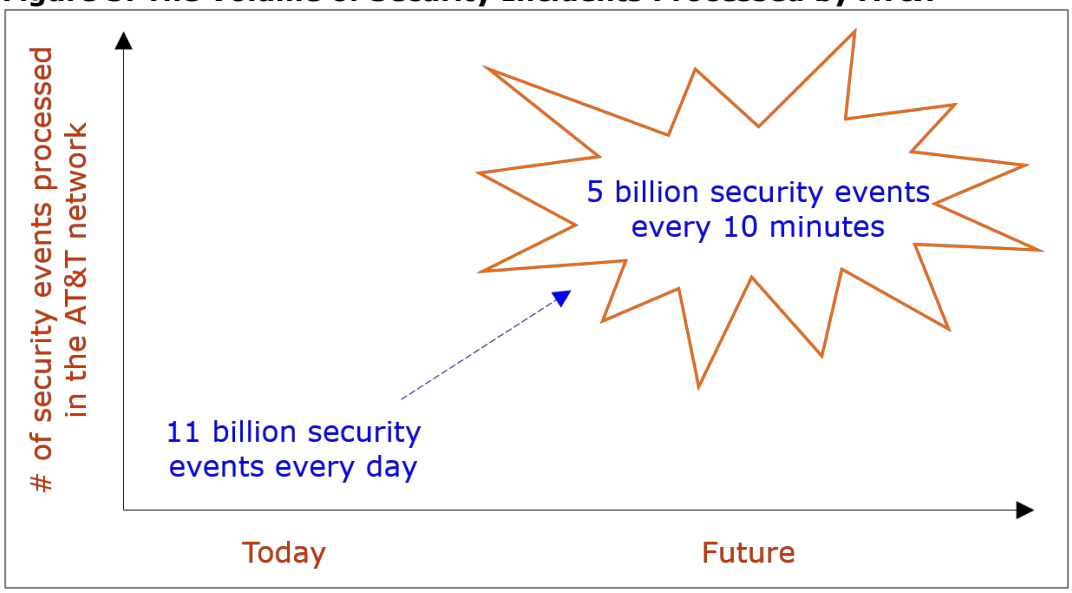
- Unlike previous generations of cellular technology, the 5G business model is predicated on supporting highly sophisticated use cases for enterprise verticals, including the automotive and health sectors where human lives are at stake. Security here will have to align with the stringent compliance and security regulations associated with each of these vertical industries and their variants between different countries (and sometimes between states in the same country).
- One of the key principles of 5G is that of dynamic network slicing, enabling customers to have guaranteed access to tailored network resources sporting customized features. Providing dynamic, tailored security for each and every slice for each individual customer – as well as assuring security for all customers – is among the biggest security challenges for 5G standardization.
- Attacks on the mobile network originating in the radio access network (RAN) have grown in prominence in recent years. How will a mobile operator protect against RAN-side DDoS attacks from botnet-controlled 5G devices when those devices can support speeds measured in units of 100 Mbit/s rather than units of 10 Mbit/s as of today?

A POTENTIAL NEW MOBILE SECURITY ARCHITECTURE

You begin to get an idea of the ramping up in the security challenge that CSPs are facing from a data point shared by John Donovan, AT&T's CTO, at the company's own security conference in October 2016. Donovan told delegates that his company is currently processing 11 billion security events per day. But that data point really is only the most basic

starting point: As shown in **Figure 3**, Donovan went on to tell delegates in the same talk that in its planning assumptions for network security requirements over the next few years, AT&T expects to have to process no less than 5 billion security events every 10 minutes.

Figure 3: The Volume of Security Incidents Processed by AT&T



Source: AT&T (data points) and Heavy Reading (graphic)

There are two important aspects of these data points. First, they demonstrate the sheer scale of the challenge that operators are now facing. Just as importantly, they demonstrate the importance that AT&T attaches to having extensive visibility of all the security incidents arising across all the subscriber and network end points across its networks, including at the level of AT&T's local, regional, national and international networks.

The ability to identify, categorize, correlate and remediate these incidents across an integrated security architecture is critical to the security posture of a leading operator such as AT&T. The critical point here is that AT&T is in a minority of operators in already building out this end-to-end visibility that allows it to accurately measure and forecast the security incidents that it sees down to this level of granularity. Most operators still lack this kind of visibility, which is a form of major, network-wide vulnerability in its own right.

Speed of Implementation, Detection & Mitigation

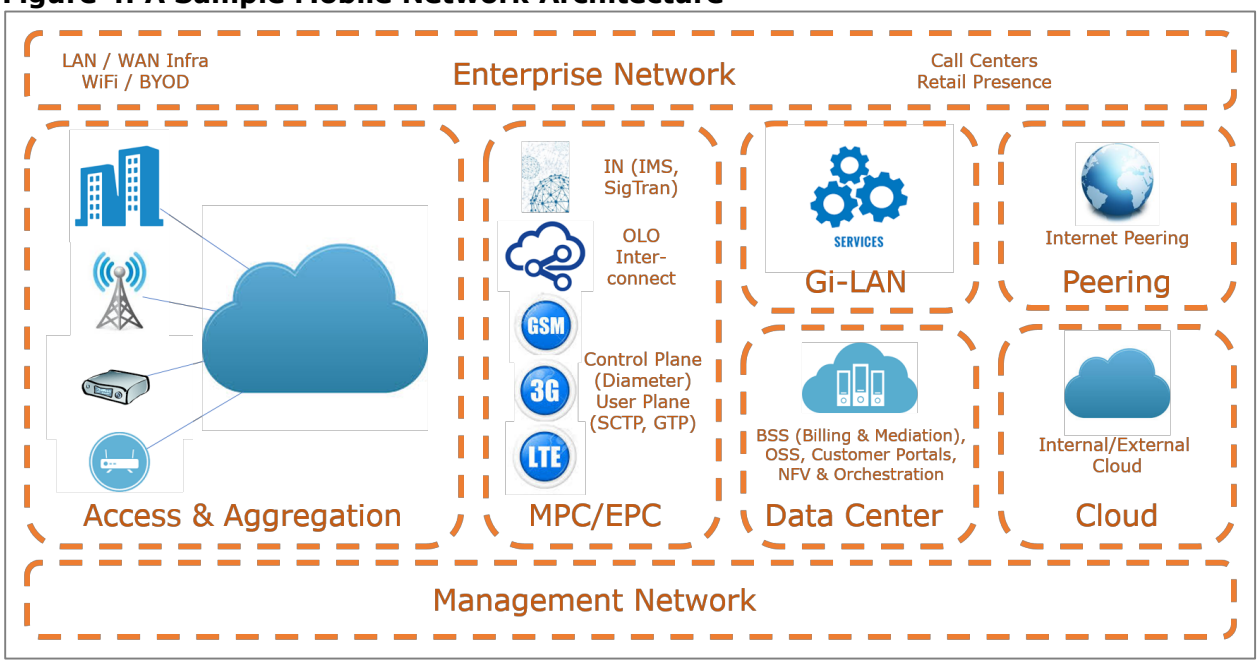
Network-wide visibility is one of the key enablers of speed in the creation and enforcement of security policy, which is becoming increasingly important in security strategy. The more rapidly security policies or outcomes can be implemented or achieved, the smaller the window of opportunity for breaches to inflict extensive damage (or to occur in the first place).

Speed also matters in the commercial environment, in the sense that measurements of speed are metrics that are concrete, persuasive and easily understandable to C-level executives. Making consistent progress in reducing the time it takes to detect this, apply that or patch the other maps well – and looks good – on any CEO's dashboard, whether they be the CEO of a network operator or the CEO of another enterprise. Without network-wide visibility of security threats, the operator is forced onto its back foot – forced into reacting to

incidents in hindsight on the basis of limited information, instead of proactively preventing on the basis of more complete threat intelligence.

Figure 4 depicts today's mobile network architecture as it relates to 2G, 3G and 4G. In the evolution to 5G, the very high-level domains – or boxes, as they are shown here – will remain essentially the same, although there may be transformations within many of the individual boxes (as well as in the relationships and boundaries between adjacent boxes).

Figure 4: A Sample Mobile Network Architecture



Source: Cisco Systems

An Extended Role for the 3GPP SEG Into 5G

Looking at the relationship between the Access & Aggregation domain shown at left and the Mobile Packet Core (MPC)/Evolved Packet Core (EPC) domain in the middle, we expect some sort of reuse of the 4G Security Gateway (SEG) in 5G. The SEG was originally specified by 3GPP as a core network device colocated with the 4G EPC. It was specified to carry out decryption of IPsec tunnels instantiated by eNodeBs to protect against attacks such as unauthorized call interception or eavesdropping. The SEG is also designed to perform validation of PKI authentication certificates, to ensure that eNodeBs seeking to attach to the network are trusted elements and not rogue elements being used in a "man-in-the middle" attack, for example.

Over the years, adoption of the SEG by 4G mobile operators has been mixed. Today a few use it for both decryption and authentication. Most of those that do use it, use it just for decryption. Many operators have rolled out 4G without any encryption or PKI authentication – i.e., no SEG at all.

There's a good chance that IPsec will be carried over in the 5G standards as the recommended encryption technology for the new NG1, NG2 and NG3 interfaces that will be specified. These are set to serve as the nearest equivalent to 4G's S1 interface in the emerging

5G standards. We expect SEG adoption rates and deployment models to evolve in the following ways:

- **More adoption of SEG for both decryption and PKI authentication:** Awareness of vulnerabilities with respect to eavesdropping on mobile phone calls has been heightened of late by the Pegasus and SS7 vulnerabilities cited in **Figure 2**. This has the potential to drive greater adoption of encryption, hence decryption in the SEG as well. The expected proliferation of small cells in 5G also has potential to drive adoption of PKI authentication in the SEG. This will help to protect against rogue cell sites in an increasingly dynamic cell site environment.
- **More additional security applied to decrypted traffic:** In the last couple of years a handful of leading 4G operators have identified the decrypting of traffic at the SEG as an opportunity to apply additional security features to the traffic prior to sending it on to the core. Examples include firewalling and malware detection. Consistent with heightened security risk, we expect additional demand for this type of approach to security policy in the SEG in both the 4G and 5G contexts.
- **A more distributed SEG.** Distribution of the 4G EPC via means of virtual EPC (vEPC) is a key trend driving lower costs, better performance and lower latency. We already see that smaller SEG instances are being distributed in conjunction with these vEPC instances. We expect this trend to accelerate with 5G. In cases where additional security features are also added, this will have the added benefit of addressing threats further out at the edge of the network.

The Contribution of End-Point Awareness to Security Policy

A shortcoming in the security policies of many organizations – including most mobile networks – is that detection systems often flag issues or make decisions in the absence of historical context. In other words, the behavior of a particular end point is judged according to whether or not that specific behavior (e.g., transmitting these specific control packets or opening those specific files) violates the organization's latest security policy. An approach that can take account of the historical behavior of unique end points has the potential to add significant value to a security posture by reducing the number of false negatives and positives.

In today's threat environment, end points can all too easily be taken over by malware and start contributing to attacks on other end points, in addition to being subject to an attack themselves. In these circumstances, it's the change in an end point's behavior relative to its previous behavioral record that can be all-important in determining whether it's behavior is malicious or not. Solutions that can build a baseline with respect to end-point awareness, spot deviations from that baseline, as well as correlate new and past behaviors in support of threat detection, should have increasing value in the emerging threat environment.

The Data Center & the Cloud

In many ways, the bottom right-hand corner of **Figure 4** is the most important – and potentially the most controversial – area regarding the way the mobile security architecture evolves over time. It's in this corner that operators need to consider what they've achieved in terms of real cloudification of their network functions in their data centers as discussed. Operators need to reflect soberly on how their own achievements to date compare with state-of-the-art levels of automation and orchestration in the cloud as practiced by the Webscale Internet companies and many of their enterprise customers.

Operators also need to consider the evolution of the mobile threat environment. Here, they must consider whether the combination of their own security team, support from multiple third-party security vendors, and their own organization's capabilities to get to something close to parity regarding best-practice cloud automation will be powerful enough to meet the security challenges that lie ahead.

For sure, in some cases operators will conclude that they can continue to compete for scarce and expensive network security personnel over the next five to 10 years. They'll conclude that their own sizeable security team can continue to assemble their own multitude of best of breed security partners that have a global presence over the next five to 10 years. And they'll conclude that their own companies have the organizational and technological flexibility and sophistication to remain on-par with – or at least not too far behind – the market leaders in highly automated and orchestrated cloud networking capabilities.

But any operator currently running 2G, 3G and 4G networks – and expecting to deploy 5G in the next few years – that is harboring doubts about whether their company meets the above description should be taking a long, hard look at the bottom right-hand corner of **Figure 4**. They should be asking themselves whether their company is best able to build out the kind of second-to-none security capabilities that are needed to meet the great commercial and technological challenge of our times – that of cybersecurity. And they should be asking whether this uniquely critical area is one they might want to think about potentially buying in outright from security specialists, by way of an infrastructure-as-a-service (IaaS) model.

Considering an IaaS Model for Security

IaaS is typically no more controversial to the IT side of an operator's business than it is in the case of any other enterprise. But run IaaS past the network side of the house, and it certainly is controversial. Mobile virtual network operators (MVNOs) may rely on leasing most of their infrastructure from others, goes the conventional argument, but that's precisely because they've chosen to be MVNOs. Convention dictates that a network operator needs to build, own and control all its own infrastructure.

In fairness, even that isn't entirely true. Mobile operators used to build out all their own cell sites. Now many are built, owned and operated by third parties. Ah, but that's just civil works, some will say. When it comes to live network elements, the operator needs to own and operate all of those themselves. Again, not in the case of transport networks. Some of the world's most successful mobile operators rely on backhaul networks that are built, owned and operated almost entirely by third parties.

There are two unique aspects of the emerging security challenge that are starting to make IaaS a credible model for mobile operators to consider.

- The rate at which the cyber threats posed to critical infrastructure and indeed to human life is clearly serious now, and this threat is only going to intensify.
- The level of expertise required to design, deploy and operate highly competitive threat defenses across a mobile operator's own enterprise, network, cloud and customer domains is very high indeed – and getting ever more sophisticated. Going forward, many operators may not be able to afford the necessary resources to be able to do the job themselves in a highly competitive market for top talent.

KEY CAPABILITIES OF SECURITY IaaS

Done the right way, investing in security according to an IaaS model can give the mobile operator access to a host of security services that combine to inform a security architecture with the kind of visibility, end-point awareness and correlation across networks that is needed to make a mobile security architecture fit for purpose in the migration to 5G. At a more detailed level, the types of capabilities that mobile operators should be looking for can be broken down into the following kinds of prevention, detection and mitigation services.

Threat Prevention

Much like attack is often viewed as the best form of defense, prevention is a key layer in a layered network security solution. While media attention around security tends to focus on the most sophisticated malware-based attacks, the truth is that the large majority of security incidents arise from vulnerabilities that could have been closed off with a better approach to prevention. This is as true in the mobile operator context as any other.

Leading-edge access controls are a critical part of the answer here. In the case of firewalls, it's certainly true that perimeter security alone no longer works. That doesn't mean that it isn't still an important layer of the security solution, though. Firewalling to determine rules of access to the network from external networks is still a key building block in security policy. That remains the case regarding the mobile operator's employees accessing external networks, or indeed its customers accessing the Internet via the Gi or SGi interfaces in 3G and 4G (and its 5G equivalent).

Whether it be within the mobile operator's own enterprise context or that of an enterprise customer, the need to define, monitor and enforce the unique rights of individual employees to access some network or application resources but not others, is very well established in security circles. It remains as important as ever.

In an increasingly dynamic, software-driven environment, however, it's proving increasingly challenging to keep up with the growing volume of rule changes – and avoid an increase in error rates that create vulnerabilities – using traditional techniques. This is because traditional techniques, such as access control lists (ACLs) and VLANs, require a lot of manual intervention by IT admins. Being open to automated access control solutions that enable the operator or its enterprise customers to do away with manual techniques such as ACLs and VLANs opens up an opportunity to improve security by reducing errors, grow revenues faster and reduce opex.

Just as firewalls remain important for the prevention layer, intrusion detection systems (IDS) or intrusion protection systems (IPS) remain important contributors to the detection layer. It's true that they rely primarily on signature-detection techniques that are not well suited to detecting sophisticated or zero-day attacks. However, they are nevertheless effective at capturing the bulk of known attacks that do bear a known, identifiable signature.

Detecting & Remediating Advanced Malware

The use of polymorphism and obfuscation in malware coding has reached the point where it's close to becoming almost standard. Malware authors are increasingly using rootkits to attain persistency on the system level. A lot of techniques are being seen enabling attackers

to bypass OS security protections, stay dormant for a while, and then become active later. Security researchers report increasing use of steganography – concealing a file or message within another file or message with the goal of avoiding detection.

Threat-detection and remediation solutions are therefore needed in addition to signature-detection solutions that can't protect against such advanced malware. These don't just look to block advanced malware from entering the network in the first place; recognizing that some advanced malware *will* make its way into the network, advanced malware solutions continue to monitor different end points – including employee and customer mobile devices, depending on the context – against hundreds of behavioral rules or indicators.

This means that when malware that has already made its way into the network begins to morph its behavior, advanced malware-detection solutions are designed to spot the behavior and investigate it, using sandboxing, for example. Moreover, having identified one instance of a file previously thought to have been benign as malicious, the best advanced malware-detection solutions can also remove all other instances of that same file that are already in the network – and block any other instances from entering in the first place.

Dedicated & Standalone Variants of Anomaly Detection

An increasing amount of investment is also going into anomaly-detection solutions that leverage packet capture, big data and machine learning for breach validation. These are also complementary to signature-based detection, in that they are able to identify and remediate deviations from established norms. Examples include behavior by users or devices that violates access control rules or suspicious packets in the network traffic.

Anomaly-detection solutions come in standalone variants, in addition to those that are integrated into the primary network infrastructure. An advantage of an integrated approach in both the operator and enterprise network context is that embedding anomaly detection functionality into network switches and routers enables all those elements to serve as security sensors throughout the network, including at the remote edge of the network.

This is important because it's at the remote edge that malware and other anomalous packets are most likely to first enter the network. An integrated approach can also amount to an upgrade to the existing installed base of infrastructure in the network as an alternative to investing in a dedicated anomaly detection vendor.

Leveraging DNS Intelligence

The colossal impact of the Mirai botnet attacks on Dyn last year is the single most powerful example of the vulnerability of ICT infrastructure to attacks on the DNS. These days it's no longer a rarity for mobile users to suffer a data outage or a degradation of service arising from an attack on the operator's DNS servers, even if the specific cause of that outage isn't always publicized.

DNS is among the most common protocols used for reflection/amplification attacks. Cisco's analysis of malware validated as "known bad" recently found that the vast majority of that malware – no less than 91.3 percent – uses DNS as a command and control protocol. Among service providers, DNS is among the most common services targeted by application-layer attacks. Indeed, another leading security vendor reports it is the most common attack of all.

Hence, there is substantial value in a DNS security service that can comprehensively monitor, pool and analyze DNS-related activity from billions of DNS requests originating within and between networks throughout the global Internet. The value then lies in leveraging that intelligence to protect users from the subset of that activity that can be identified as malicious.

This can take the form of making it possible to block domains or websites that are known to be malicious at the DNS and IP layers, based on patterns of requests to those domains. From a more preventative standpoint, it can take the form of monitoring domain registrations, DNS server purchasing patterns and IP address acquisitions to watch for patterns of behavior that align with the behavior of attackers – for example, as they assemble the resources they need to carry out an attack.

A resource of this kind requires a lot of investment and a lot of highly-specialized expertise. Most operators are going to be very challenged to develop anything like this kind of competence in house. The same is also true for all but the very largest enterprises.

Threat Intelligence

Threat intelligence underpins any mobile operator's prevention, detection and mitigation efforts. Threat-intelligence feeds come in all shapes and sizes; indeed, the cybersecurity market is overrun with them. For a network operator of any size – including those that want to be competitive in enterprise security – basic threat-intelligence feeds aren't sufficient. An organization of that kind needs to be thinking in terms of a threat-intelligence service of substantial breadth and depth.

Among the characteristics of a threat-intelligence service that's fit for purpose for a mobile operator are that:

- It invests heavily in monitoring the activities of bad actors in order to establish a profile of their behaviors, tools and techniques.
- It captures threat intelligence from the broadest possible variety of sources.
- It enables the company using it to filter out only the threat intelligence that is valuable and actionable relative to its unique security posture.
- It communicates rapidly in order to accelerate time to detection and time to mitigation. That can be in the form of issuing written customer advisory notices promptly. Critically from both an enterprise and CSP perspective, it can also be in the form of threat-intelligence updates that are automatically shared by common network elements throughout either the CSP or enterprise network (or both).

SUMMARY

The software programmability of mobile networks must accelerate in order to respond to the new scale and sophistication of mobile threats that operators can expect in the coming years. 5G increasingly appears as an inflection point for better leveraging the cloud in the mobile network security architecture – in terms of the way the operator drives its own infrastructure for security outcomes and in terms of its openness to considering new security models, such as leveraging IaaS for security.

ABOUT CISCO

Cisco (NASDAQ: CSCO) is the worldwide leader in IT that helps companies seize the opportunities of tomorrow by proving that amazing things can happen when you connect the previously unconnected. For ongoing news, please go to thenetwork.cisco.com.

Cisco and the Cisco logo are trademarks or registered trademarks of Cisco and/or its affiliates in the U.S. and other countries. A listing of Cisco's trademarks can be found at www.cisco.com/go/trademarks. Third-party trademarks mentioned are the property of their respective owners.