Data Communities on AT&T Network 3.0 Indigo

This paper examines trends and capabilities driving the emergence of data communities on AT&T Network 3.0 Indigo, the key architectural components of the platform and example use-case communities.

Data Communities on AT&T Network 3.0 Indigo

Table of Contents

Executive Summary

The pace of digital data generation is accelerating exponentially, driven in large part by the proliferation of IoT devices. This acceleration has improved the ability to derive valuable insights from data. People across different entities are more frequently looking to form communities of common interest to share ideas, information and resources to aggregate data and insights. As data usage grows, security, identity management and privacy play a key role. The surge in available data, cloud computing and new open platforms has also led to the emergence of practical machine learning (ML) and artificial intelligence (AI) applications.

AT&T's network evolution from hardware to software-based functions has enabled new services like AT&T FlexWare, introducing greater agility in the network. The combination of software-defined networks, big data analytics and advancements in automation create the foundation for significant transformation throughout AT&T's platforms, products, processes and people. These capabilities can be extended beyond AT&T to other entities. This is what AT&T envisions as data communities on our AT&T Network 3.0 Indigo (hereinafter "Network 3.0 Indigo" or "Indigo").

Indigo itself is the next generation of modern networking that AT&T is constructing. It includes access technologies like LTE Advanced and 5G, but it's much more. It includes software-defined networking and orchestration, analytics and the replacement of bulky applications with microservices tapping into shared pools of data. The data communities platform, one part of Indigo, will be a cohesive set of functionality to enable closed communities of sharing and collaboration, with underlying SDN-enabled capabilities and infrastructure services.

The community platform will enable dynamic, on-demand combinations of data to be sourced from multiple entities and merged into shared communities to derive insights in a highly secure environment. It will connect the best human intellect and ML capital in order to scale capacity for learning and enhance collaboration among community members to help solve problems. AT&T has already enabled early, internal beta

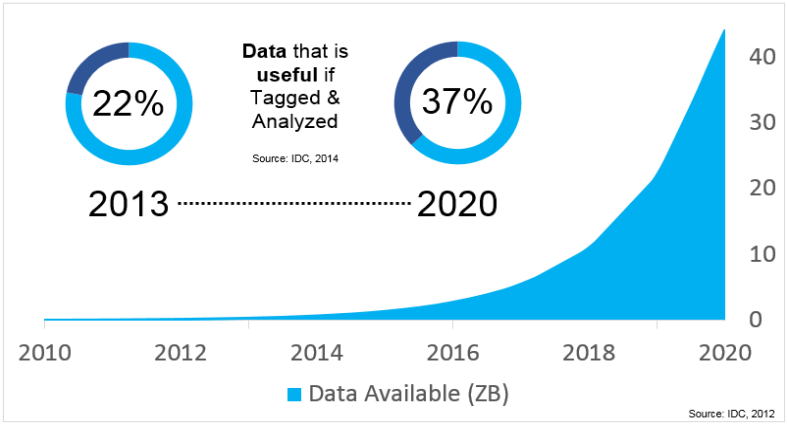
versions of Indigo communities to address initial use cases in technician dispatching, customer care, threat analytics and video delivery optimization.

This paper outlines the trends driving Indigo data communities and explains the key architectural components of the platform.[1] It also provides use case examples to further illustrate the vision for the platform and the transformative power it holds.

Trends Driving the Evolution

The pace at which data is generated and stored is accelerating exponentially, driven in large part by digital adoption and the proliferation of IoT devices. By 2020, there will be an estimated 20 to 50 billion connected devices.[2] Nearly 1.7 megabytes of new information will be created every second for each human being in the world, and digital data is forecasted to exceed 44 zettabytes - approximately ten times that of today (see Figure 1). In the past, the world's data doubled every century; now, it doubles every two years.[3]

Figure 1: IDC data growth forecast



As data usage grows, security, identity management and privacy play a key role. The World Economic Forum has listed cybercrime as the top global risk,[4] and the average cost of a breach has risen to $7 million per incident or $221 per record lost or stolen.[5]

---

[1] This paper lays out the platform and architecture of data communities on AT&T Network 3.0 Indigo as envisioned today. It will continue to evolve.
[2] http://www.gartner.com/newsroom/id/3165317
[3] "Why a Connected Data Strategy is Critical to the Future of Your Data", Hortonworks White Paper Value of Community Insights
[4] http://www.nytimes.com/2016/04/07/us/politics/homeland-security-dept-struggles-to-hire-staff-to-combat-cyberattacks.html?_r=0
[5] http://fortune.com/2016/06/15/data-breach-cost-study-ibm/

The Department of Justice's Internet Crime Complaint Center recorded nearly 270,000 cybersecurity-related complaints in its 2014 report – an increase of over 1,500% since 2000.[6] Consumer privacy and security concerns have begun to influence internet activity as well. Nearly half of consumers are more worried about privacy than a year ago, and 74% have limited online activity due to privacy concerns.[7]

Nevertheless, despite these security and privacy concerns, communities are more frequently being used to derive value from data. A community is a self-organized network of people with a common interest who collaborate by sharing ideas, information and resources. Nearly every industry is trending towards "open innovation" in which companies collaborate with people, academic institutions and other entities.[8] The boundaries between a company and its ecosystem are becoming increasingly invisible.[9]

The surge in available data, the growth of open-source software, widespread virtualization and cloud adoption has led to the emergence of practical ML and AI applications.[10] These technologies, when fed massive amounts of data, are demonstrating significant advancements in areas such as object detection, speech recognition and natural language processing. Machine and deep-learning techniques are designed to be iterative in nature, constantly learning and optimizing outcomes.[11] Instead of static rule or policy-based automation to guide the system on the next best step, AI is able to make a more intelligent recommendation based on data, context and patterns to determine outcomes proactively.

Emergence of Network 3.0 Indigo

---

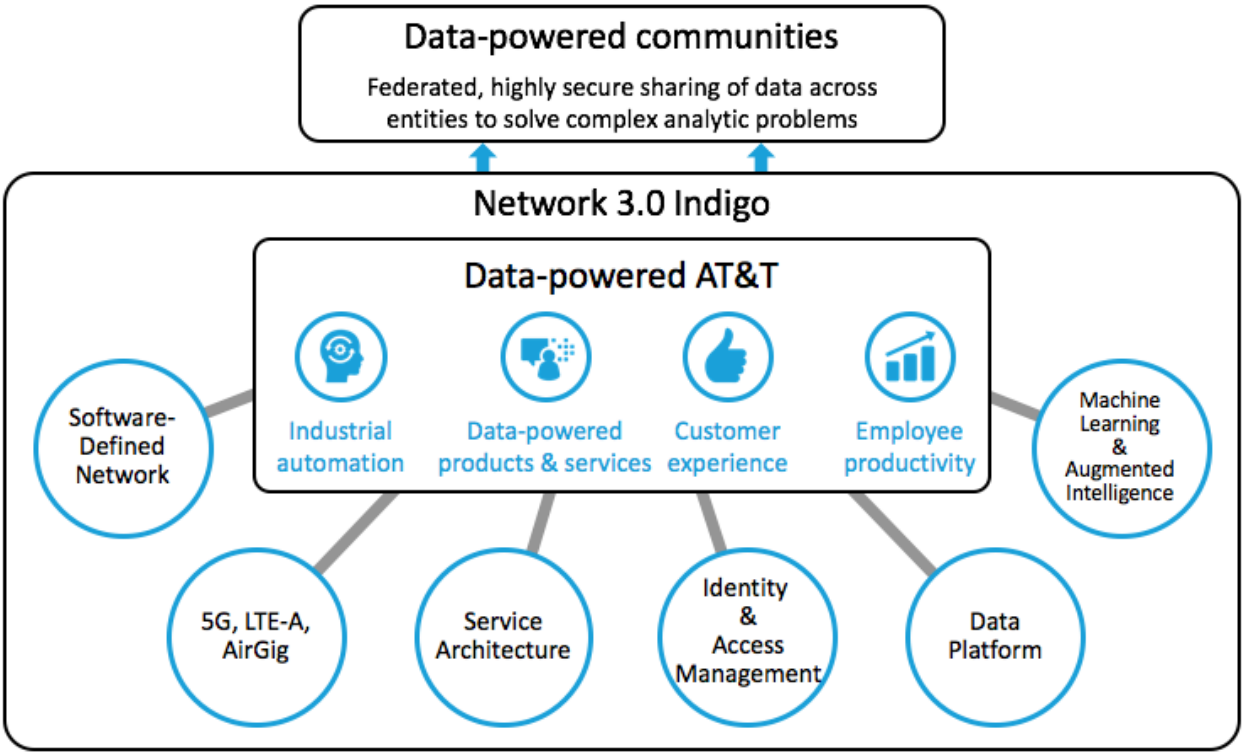
[6] http://www.foxbusiness.com/features/2016/04/27/cyber-attacks-on-small-businesses-on-rise.html
[7] http://www.kpcb.com/blog/2016-internet-trends-report
[8] http://thomsonreuters.com/en/press-releases/2015/05/tr-analysis-shows-concerning-trend-for-global-innovation.html
[9] "Open Innovation, Open Data, and New Business Models", Hans-Dieter Zimmermann & Andreja Pucihar
[10] http://www.economist.com/blogs/economist-explains/2016/07/economist-explains-11
[11] http://www.forbes.com/sites/louiscolumbus/2016/06/04/machine-learning-is-redefining-the-enterprise-in-2016/#7aded0d35fc0

As AT&T's network continues to evolve from hardware-based functions to software-based functions, new services like AT&T FlexWare have enabled greater agility in the network. In parallel, advances in data analytics and automation have enabled AT&T to better solve complex internal business challenges. AT&T's progress in this area was made possible through internal data sharing and collaboration platforms with application programming interfaces (APIs), microservices and security. These platforms allow knowledge workers to collaborate across business units by enhancing access to data and providing AI and ML tools. This evolution is driving a significant transformation throughout AT&T's platforms, products, processes and people.

In addition, these platforms can be configured to allow the participation of external entities within the communities to optimize collaboration and drive solutions to complex problems. This is what AT&T envisions with Indigo data communities. Data sourced from multiple entities could be merged into shared data communities where each data contributor retains data ownership, granting controlled access to community members to derive analytics and insights in a highly secure environment. The complexity of a community's stated objective could attract the best human intellect and ML capital to scale the overall capacity for learning, and to augment human intelligence to make better decisions. Mixed reality could enhance the solutions by providing an overlay of synthetic content on the real world that is anchored to and interacts with the real world, creating new experiences to solve complex business problems.

Network 3.0 Indigo data communities are the natural progression of AT&T's internal journey toward data-sharing, collaboration communities and a recognition of the importance and transformative value that this may bring to others across industries.

Figure 2: Emergence of data communities on Network 3.0 Indigo



Defining Network 3.0 Indigo Data Communities

Indigo data communities will be a cohesive set of functionality composed from three distinct technology components: new platform capabilities, SDN-enabled capabilities, and infrastructure services. The key functionality contained within these components, as envisioned today, is further defined in the following sections.

New Platform Capabilities

Leveraging the success of our NetBond product and other associated services, new platform capabilities will enable the dynamic creation of federated communities to share data in a highly secure environment in order to solve complex analytic problems. Each community will have a community owner who will be able to invite other participants to join the community, such as those from other parts of their enterprise, commercial institutions, academia, industry consortiums or independent researchers. Community

members will access shared data sources, set their own sharing policies and bring their own data to share.

Data owners will be able to retain ownership of their data and have control over who has access to it. Access to the community will be managed through AT&T's next generation identity and access management platform. Through federated single sign-on, users will be able to leverage existing identity credentials once the identity provider is on-boarded to the platform using standards-based approaches including SAML 2.0 and OpenID Connect. If an existing identity is not appropriate or available, the user can register and authenticate directly using AT&T multifactor authentication capabilities. The Indigo data community platform will check that user credentials meet the required identity proofing and authentication level for the specific data community that is accessed. A policy engine will enable the owner of the community to set policy for sharing and accessing information among members of the community, with the ability to control data access and privacy at the data set level or even at the record or field level.

Communities will be preconfigured with a set of ML and AI tools that can be used to build, validate, deploy and run analytic microservices generated within the community. Members can also bring their own analytic tools, conditional on completing a certification process so that the tools are compliant with the community's security standards. The platform will also provide a set of connectors to approved data sources with an ability for communities to extend this capability to custom data sources. All data brought into the community will be stored in a highly secure privacy enabled environment so that community members only access the data they are allowed to see. The data owners within the community will have control over allowing any data to stay within or leave the community. To facilitate communication among community members, AT&T will provide federated, highly secure unified communications capabilities.

SDN-Enabled Capabilities

Communities on the new platform will be able to run on most virtualized networks so long as certain requirements are met.[12] A community running on AT&T's SDN will inherently have these capabilities; however, any virtualized network that also cares for these requirements may be used.

For example, AT&T's SDN capabilities are secured by AT&T's identity and access management platform, which provides risk-based access controls and multifactor authentication options. The same identity and access management platform will be used to create highly secure environments for Indigo data communities. Networks with weak identity and access management controls would render many of the security features ineffective since they will be working with an untrusted identity.[13] Securing access to the network itself is critical to keeping uninvited devices and actors out of the community. AT&T's SDN comes with highly secure network access channels including NetBond, ANIRA and AVPN.

The network layer serving Indigo data communities will provide access to network, compute and storage facilities as well as API-driven orchestration services that enable the platform to provision the necessary functions on behalf of the data sharing community. The network layer will also need to have a level of programmability and expose that capability to the Indigo data community.

The programmability inherent in AT&T's Enhanced Control, Orchestration, Management and Policy (ECOMP)[14] enabled SDN network is built from the concept of microservices. The core principles behind microservices are reusability, encapsulation, use of lightweight protocols and infrastructure independence. The capabilities produced by network microservices are exposed via infrastructure independent APIs that allow service consumers to be completely abstracted from the underlying network layer.

[12] Details regarding requirements will be coordinated with the future release of Network 3.0 Indigo
[13] This is discussed in further detail in Trusted Identity and Access section of Multi-Network Extensibility below.
[14] ECOMP white paper: http://about.att.com/content/dam/snrdocs/ecomp.pdf

These features will enable analytic services to be provisioned directly on top of AT&T's SDN without needing to be tightly coupled with the network service layer, and will allow the Indigo data community analytic microservices and network microservices to form a common development framework.

All the features described in this section come standard with AT&T's SDN using the open-source ECOMP framework and industry-leading network capabilities provided by AT&T's core network. Other network providers running the ECOMP open-source virtualization platform will also have access to the type of capabilities that will be part of the requirements to run Indigo data communities.

Infrastructure Services

Data communities on Network 3.0 Indigo will need access to compute, network and storage facilities in order to provide analytic and data services. These services can be enabled by any Virtual Private Cloud infrastructure provider including on-premise or third-party public cloud providers. The infrastructure provider will need to have API accessible cloud orchestration services to provide the auto-scaling features of the platform. For a high level of security, the infrastructure provider can also enable AT&T NetBond to provide access to the platform and to automatically deliver all the capabilities described in the SDN section above.

Multi-Network Extensibility

A Matter of Trust

At their core, data communities on Network 3.0 Indigo will provide a highly trusted infrastructure, giving users confidence that data is only accessed and processed in accordance with the policies defined by the community and authorized by the data owners. To extend Indigo data communities across multiple networks requires careful trust management to maintain this confidence. This requires network security and identity confirmation. Transparency to the community about the validated level of security will be important when extending Indigo communities access across additional networks.

There are different levels of trust that may be extended. For instance, AT&T's SDN could extend trust to another network to authenticate a user. Trust could also be extended to another network to provide highly secure VPN service to a community member accessing the Indigo data community infrastructure. Finally, AT&T's SDN might extend trust to a third party in order to store or process data on its infrastructure.

Trusted Identity and Access

Today's typical identity methods focus on multi-factor authentication. By adding additional factors based on network-derived information and patterns such as physical location for wireline networks and SIM-based access management for mobile networks and devices, AT&T's network could provide enhanced identity management and verification. For AT&T customers on AT&T's network, AT&T will utilize its world-class security assets to perform and verify the trust level of the authentication required for access. However, what happens if a user wants to log in from another network? What if the log in request comes from a shared device? The value of Network 3.0 Indigo communities will be maximized when the same levels of authentication confidence can be extended across multiple network operators. This will require consistent standards to establish trust across operators. Just as user authentication will have different levels of confidence, trust across networks will have different levels of confidence.

The various data communities will have a range of data classifications and security needs. AT&T's identity and access management capabilities will help members enroll with the appropriate identity and authentication proofing as prescribed by the community policies. Moving beyond static policies that could require specific multifactor authentication, the platform will also apply threat intelligence and behavioral analytics to optimize risk-based controls that can increase authentication requirements based on perceived risk. To create an effortless and highly secure experience, the member can use a smartphone application as part of the authentication process. For AT&T provisioned mobile devices this will allow AT&T to incorporate network-based authentication functions and detect anomalies such as location-based risk. If the rest of

the wireless network operators would adopt these capabilities for their mobile devices, these features would be ubiquitous for all community members.

Through federation standards, the platform will be able to accept third-party authenticated identities that meet certain proofing and authentication requirements. If the third party is unable to provide the appropriate authentication capabilities, AT&T can construct a hybrid access pattern where the user initially authenticates with the third party, but then requires additional credentials by the AT&T access management layer.

Trusted Network Computer and Storage

Another dimension to establishing a highly trusted infrastructure is how the user connects to the microservices and data in the community. The network itself must be highly secure and trusted. Again, there are multiple trust levels an Indigo community can establish based on the specific access mechanism within a single operator network. This is based on the network service, as well as the degree of monitoring and threat analytics conducted on the network. If access crosses multiple networks, the overall security and trust level of the access depends on the security of each network as well as the security of the interconnection between the networks.

Finally, networks can be extended to use third-party cloud infrastructure to store and perform analytics on data made available to a community. The overall trust level of the community and associated storage and compute infrastructure will be based on the authentication of this infrastructure and the security of the network connection to this infrastructure.

Trust Across Network Operators

The trust level associated with connecting a user to community data across networks run by multiple operators depends on the trust level of each of the networks and the links between the networks. At the transport level, the highest levels of trust on a single network can be obtained by physical VPN connections and SIM devices operating on highly secured private home networks. Levels of trust decrease with the use of internet connections across multiple networks. The highest levels of trust among networks can

be achieved when the network operators rely on common trusted processes and policies. This is possible when network operators use common platforms for operations and provisioning policies. AT&T's open-source ECOMP platform provides an ideal mechanism for achieving this by allowing the same level of trust in policies implemented in each of the networks and joint compatible provisioning, operations, and policies of the connections between the networks. If the interconnected operators are not running the same ECOMP platforms, a range of lower levels of trust and security can still be achieved by establishing detailed technical, operations, provisioning and testing requirements, along with certification and auditing processes across operators.

Network 3.0 Indigo data communities will allow for a flexible federation of resources and infrastructures to support diverse requirements of different communities, keeping security front and center through policy, transparency and validation.

Community Examples

The innovation that will be made possible through this platform is different than innovation that has been achieved to date. By breaking down data silos, Indigo communities will create a new way for people and enterprises to share data, access AI tools and collaborate. Innovations will be easily transferred inward and outward, with the data owner controlling what is shared. The platform will foster a trusted and diverse set of Indigo community-certified contributors, where platform extensions will deliver exponential growth of innovations and value, similar to what the app-store model did for mobile applications. This open innovation may lead to new business models, redefining the entire logic in which data's value is created and captured.[15]

Data communities on Indigo are best understood through concrete use cases. The following sections discuss several use cases. It is expected that Network 3.0 Indigo will enable use cases where AT&T is the community owner as well as use cases where the community owner is unrelated to AT&T.

---

[15] "Open Innovation, Open Data, and New Business Models", Hans-Dieter Zimmermann & Andreja Pucihar

Indigo Data Community Use Case #1: Technician Dispatching

AT&T employs one of the largest technician fleets across North America. It needs to be efficiently dispatched on a daily basis for both new installation work and repair jobs. The work of technician dispatching is a classic and complex Operations Research optimization problem. Variables include technician skills, job types, equipment availability, customer availability, workforce constraints, training needs, vacation times, routing, weather and a host of other considerations to make an optimal schedule.

Other industries also have the need to dispatch technicians with a high degree of variability in the complexity, time and skills required to solve the business need. Some examples include large-scale plumbing companies, HVAC companies, construction crews, and utilities. A community could be formed where companies come together to share their data and expertise in order to develop analytic algorithms yielding better technician assignments. Real-world data would allow community members to develop new approaches to solving their optimization problems. All participating members would benefit from the enhanced analytics developed through sharing data and analysis with strong policy controls on how raw data is used.

Subsets of data could be used to address more localized improvements. For example, if all companies that dispatched trucks into a city shared live routes and drive times between neighborhoods, every member could get better routes for their drivers and drive times that could be used to improve the job assignments. The output would be a derived data set benefiting all members participating in the community. Having the improved routing data built into the optimization engine that schedules jobs would offer an additional benefit over only using the routing provided by common navigation applications available today.

Additionally, members' customers could also participate by sharing their calendar, location data, commute times and workplace information. A member's customer would have the ability to control their data and revoke the rights to use it at any time and the community would automatically honor that request. Members' customers would benefit by getting a better match between their availability and the technician's availability.

Members would benefit by achieving a more efficient dispatch process and improving customer satisfaction.

Indigo Data Community Use Case #2: Security Threat Analytics

AT&T protects its network from myriad security threats. The nature of these threats change every day. AT&T has access to an enormous amount of data and some of the best security analysts in the world.

Establishing a collaboration community with other large network operators, government agencies, security companies, researchers and academia would provide a way for AT&T and others to improve network security. The members would share their data, algorithms and threat indicators within the community. An output could be an enhanced threat indicator feed created from the best data and algorithms within the community. In addition, detailed knowledge of the attackers' methods and procedures to attack a network could also be shared among members. All community participants would benefit from better algorithms that produce more timely and comprehensive threat analytics and mitigation.

Typically, getting companies to share threat indicators, attack methods and procedures is difficult due to the risk of this information falling into the wrong hands. However, within a highly secure community environment, it is more likely that this kind of collaborative effort would be successful.

Combining Indigo community capabilities with SDN-controlled networks provides the added benefit of being able to integrate analytics derived on the platform with the runtime execution of these models on SDN-defined networks, changing the networks in response to information derived from the platform. Using a platform like ECOMP would yield a closed-loop analytics environment from initial exploration all the way through deployment and execution.

Indigo Data Community Use Case #3: Customer Care

Customer care is a key driver for customer satisfaction in any industry. This can be a daunting task as the challenges in providing effective customer care lie in a multi-step process. There is a need for a frictionless channel for a customer to communicate an issue, a method to determine the actual problem, a determination on whether to use an automated system or human representative, and a selection of the best representative based on skill sets and demand. Each of these steps requires utilization of data for the best action. Diagnosing the issue and proposing a solution requires domain knowledge and fast decision making. The data sets to address each of these factors may lie in different parts of the company organizationally, such that managerial and technical silos can hamper the overall system.

Data communities in Indigo will create a structure to help solve multi-organizational problems, like those that arise in customer care. In an enterprise, there can be varied methods for storing and accessing data. In some cases, different organizations might have decidedly different rules for access and permissions. As an example in telecommunications, network usage data or call detail data may be necessary for a representative to troubleshoot a network issue, but access may not be allowed for direct sales and marketing to upselling purposes. Adherence to strict policies about personal private data usage is crucial to legal and regulatory compliance. Indigo communities will enable different organizations to share data and control who has permissions, and what level of aggregation is required. The policy management layer will provide a structure to help compliance with internal and external requirements.

Sharing data across silos and with community members creates exciting opportunities for predictive analytics. Using customer history and other outside factors, like weather or local behavior of similar customers, will allow for the creation of predictive models for customer needs. Social network data could be integrated either in aggregate, or personalized, if permitted based on customer opt-in or other applicable requirements. Advanced ML algorithms will recommend a personalized next-best action based on historical interaction, changes on customer accounts or comparisons to like-minded customers. These algorithms will behave as real-time learning engines to drive adaptive changes based on current conditions.

For example, a community member could take a corrective action on behalf of a customer before they have a chance to notice and report a problem. Third parties could bring relevant data and algorithms to advance the learnings and create a true 360-degree view of consumers and segments across different industries. These third parties could become community members collaborating on customer care solutions. Indigo communities will manage the certified identities of these members. Methods for creation of customer journeys across multiple platforms – store visits, online chat, call centers and IVR systems – may be leveraged across different industries that have multiple customer touch points. The methods developed on the community platform could be leveraged by each community member.

Indigo Data Community Use Case #4: Optimized Video Delivery

A compelling use case for customer care communities comes from the telecommunications industry. Many customers use applications on mobile devices to consume video content. If they have a complaint about their service experience, they might complain to the content app creator or to the service provider. In this case, useful data may reside between the two entities – the app owner knows what content was consumed, and the service provider knows about the state of and impact on the network. These two stakeholders would benefit from the sharing of customer care data – to make the app quicker and more efficient in consuming resources, to lower the impact on the network and ultimately to improve customer experience.

A large volume of traffic that flows across the network is video traffic. Therefore, network and content providers have a vested interest in efficiently moving video content to optimize their capital investment while delivering the best possible experience to the customer. Currently, it is hard to predict what video content will be requested by customers, how long it will be consumed, when it will be requested, and what type of device and network will be used for the video consumption. This complex issue needs to take into account the video source location, the network core, the last mile network and the device capabilities. Ideally, each piece of the delivery process needs to be optimized and frequently these steps have competing needs. There is additional

complexity with the presence of the content delivery network (CDN) and caching considerations for frequently used content.

A community could be established among network providers, content providers, device manufacturers and CDNs. Each of these collaborators could bring data into their community to optimize video delivery to the end consumer. Each link in the value chain would offer information to yield a complete picture of the video delivery process. Content providers could bring information on the most popular videos by device type, customer type, and location, along with information about soon to be released shows and their projected viewing audiences. Content delivery networks could share which videos are being cached and at what locations along with their cache hit rates. Network providers could share performance information related to the network, including low usage times. Device manufacturers could provide metrics on storage capacities and usage as well as what percentage of video pushed to the device was actually viewed. Adding data from the community members would allow for advanced analytics that could optimize each step in the video delivery value chain and provide benefits to each member as well as the end consumer.

Additionally, consumers could be invited into the community to share their viewing habits across all media providers linked to their devices. The community and consumers could set limits so that the consumer data could only be used for the purposes of optimizing the network delivery of video and not for marketing purposes. The consumer's device could also be included in the caching network so content that is highly likely to be viewed within some reasonable time would be cached locally on the device. The consumer would get the best possible video delivery experience for their device by sharing their information while maintaining control over the use of their information.

Conclusion

Trends in data generation, security and identity management, ML, AI and communities of open collaboration are converging to enable the next evolution of technology innovation: data communities on Network 3.0 Indigo. AT&T's network evolution from

hardware-based functions to software-based functions has enabled new services like AT&T FlexWare introducing greater agility in the network. The combination of software-defined networks, big data analytics and advancements in automation create the foundation for significant transformation throughout AT&T's platforms, products, processes and people.

These capabilities can be extended beyond AT&T to other entities. This is what AT&T envisions with data communities on Network 3.0 Indigo. Leveraging software-defined networking, the platform will enable a highly secure data sharing environment, where security is promoted via identity management and authentication capabilities. As envisioned, the platform will enable useful combinations of data to be sourced from multiple entities and merged into shared communities to derive insights without compromising privacy. It will attract the best human intellect and ML capital to scale capacity for learning and enhance collaboration among community members to help solve problems. AT&T is defining new platform capabilities to enable the dynamic creation of federated communities to share data in a highly secure environment to solve complex issues. Use cases in areas such as dispatching, customer care, threat analytics and optimized video delivery demonstrate innovation opportunities when communities come together to derive value from massive amounts of previously siloed data.

