Pipeline
Technology for Service Providers.
www.pipelinepub.com Volume 13, Issue 8

Preparing for Tsunami of Data

By: Mark Cummings, Ph.D.

The move to digitize everything is entering an inflection point that leads to an extremely large increase in data volume. Even single applications are beginning to create Exabytes of data. Current planning in the industry around 5G indicates that many participants are thinking that things will be essentially the same, just bigger. This will not be the case. They will be unprepared for the tsunami of data unless there are fundamental changes in the ways CSPs and enterprises architect and operate their networks.

The last time the industry went through a fundamental change of this importance was with the introduction of the iPhone. AT&T was the first — and because of Apple's business model at the time — the only carrier to introduce the iPhone. It planned for things to be the same, just bigger. Years later, as Apple's business model has opened up to others, CSPs have had the advantage of hindsight to guide their preparations.

What is different about the impending data tsunami is all CSPs will be hit at the same time. Without forethought and preparation, there will be serious network and performance ramifications and significant costs associated with recovery.

The human body with its distributed ganglions and lymph nodes provides a model for what CSP preparations should look like, with architecture stores and information processed in a distributed fashion — much of it close to the edges.

Predicting the future is at best, "looking through a glass darkly," so it is helpful to look at indicative current and near-term examples in aerospace, factory automation, medicine, 3D printing, and autonomous vehicles in order to understand the size and scope of the tsunami.

Aerospace: Rolls Royce and Boeing jet engines now have, by weight, more digital equipment than propulsion equipment. A typical transcontinental flight using these engines now generates 6.5 Terabytes of data. Given the number of long haul flights, this means in the order of 10,000 Terabytes of data a day. The amount of data is too big for centralized Big Data.

Automated manufacturing: This industry has progressed to the point that metadata and control plane data are extremely valuable — so much so that manufacturing operators want to ensure it remains in their plants., the security concern being that competitors trying to reverse engineer products would covet this data. And because of security concerns, as well as data quantity, factories are beginning to use machine learning and big data analysis techniques at the factory level to protect against threats and losses.

Medicine: Increasingly digitized, the average hospital patient carries four sensors or actuators, as with insulin sensors, pumps and pacemakers. Additionally, consumer wearable technology is generating large amounts of data, which patients expect to be used in their care. For example, a group of doctors at Stanford Medical Center observed that larger numbers of their patients were coming to appointments with smartphones' and wearables' data that they wanted to leverage in their diagnoses and treatments. At first skeptical, the doctors investigated the accuracy of the data from consumer devices, but they ultimately found that the consumer data fell well within standards for clinical use. As this happens in more places, health care providers will integrate consumer data with that of their own professional instruments.

As that data grows, there will be a commensurate increase in risk. Consider, for example, the WannaCry episode, which demonstrated the susceptibility of medical systems to malicious attacks. It's also evident that the increase in devices also attracts hackers, so vulnerabilities, such as those now known in pacemakers, can have life-or-death implications if not proactively accounted for and prevented.

3D Printing: NASA has sent 3D printers to the International Space Station in order to reduce the stockpiling of certain spare parts; instead, creating parts as needed through design data and 3D printing. The cost of 3D printers varies by the type of material “printed,” with simple monochromatic plastic being the least expensive and metals being the most expensive. However, the cost of materials and 3D printing is coming down, which will increase the number of “products” converted into tangible form, on demand, from information that resides closer and closer to the end user. Because 3D printing files will be quite large and the volume of products sold daily will grow significantly, the level of traffic will be substantial.

Autonomous Vehicles: Maybe the most dramatic example of the coming data tsunami is in the transportation arena. Today, vehicles are being shipped with intelligent cruise control and other features that move us closer to the reality of self-driving vehicles, which are already undergoing on-road testing.

Today, the innovations are pretty much self-contained, but “around the corner” are innovations that require interconnection and constant communication — both of which will trigger the creation of more data.

For example, "platooning" is on the horizon. It is the practice of getting two or more cars close enough to enhance aerodynamics and reduce power consumption. Analogous to “drafting” in NASCAR, assembling of groups of vehicles on the highway will produce many benefits: reduced power consumption, reduced cost, reduced pollution, increased road carrying capacity, among others.

However, achieving those benefits requires constant communication among platooning vehicles, which is why vehicle-to-vehicle communications systems are being developed today. Eventually, cars preparing to accelerate or decelerate will communicate with those in front and behind, in effect, linking the vehicles gas and break pedals to each other.

Each vehicle’s navigation system will have to know where other cars are located, and their destinations, so that platoons are comprised of vehicles grouped according to similar routes and destinations.

The aerodynamic advantage will bring with it a reduction in cooling air. Both electric and internal combustion systems need cooling, and platooning systems in each vehicle will keep track of operating temperatures so that should they get too high, the car will know to break out of the platoon.

Today, traffic and road condition information is gathered into centralized data stores and then delivered back out to the roads with suggestions for “best” routing. With vehicle-to-vehicle communication, information will be delivered daisy chain fashion back down the road resulting in faster, more accurate information.

Roads are becoming “plastic” and dynamic. In the past, the number of lanes in each direction on the Golden Gate Bridge was varied in a way that ensured the commute direction had more lanes than the non-commute direction. This traditionally was done by people, who from the back of a truck, moved plastic sticks indicating where the median should be. Today, highway departments use a moveable safety barrier and a special vehicle that like a zipper, moving back and forth to automatically place the barrier where it should be.

There is still a driver in the truck taking commands for a manual operations center, which uses video cameras on the approaches to gauge traffic conditions. In the near future, we will see automation within the truck and computers utilizing sensor data to decide when to move the barrier. As this happens, the road will start talking to the vehicle, telling it the state of the road’s variable carrying capacity.

With multiple, tightly packed platoons, meeting SLAs around accuracy and timeliness of information becomes a matter of convenience and energy savings, but also of life and death.

We have seen a number of terrorist attacks using vehicles, as that which occurred in Nice, France. Some worry that vehicle information systems could be compromised by criminal sources, who would turn vehicles into weapons. In the United States, the number of vehicles stolen via hacking is now larger than the number stolen via mechanical means, with a recent report detailing the theft of 45 brand new Jeeps from dealerships via hacking.

When hybrid vehicles first came out, they had an automatic switch that had the gas engine start to recharge the hybrid battery as soon as the charge dropped below a fixed threshold. Then, an orchestration overlay system was added to optimize performance and gas mileage by combining information from the gas engine on engine load with information from the battery on state of charge. Then yet another layer of orchestration was added to combine the information in the engine load and battery charge state orchestration with information from the navigation system about what road conditions were like ahead. So, for example, if the battery charge was low, but a steep downhill section was just ahead, using the gas engine to charge the battery was delayed so that the downhill coasting energy could be used instead.

As more layers of orchestration are added to optimize engine temperature, platoon selection, route selection, and road carrying capacity, the data volumes will explode.

Implications:

What these examples point to is a situation where:

1.) Too much data is being produced to store and process in centralized systems. This means that wherever possible, data needs to be collected and processed as close to the edge as possible. Completeness, timeliness, value, and security will play a big role in determining where data goes. The more complete and more up to the nanosecond, the closer to the edge;

2.) One user's control plane and or metadata is another user's data plane and the resulting data can have very high value;

3.) Cyber security will become dramatically more important requiring a network immune system that can detect and respond to malicious attacks extremely quickly and accurately; and

4.) There will be interactions between many disparate systems. Thus, many layers of orchestration and security will be needed to be distributed at various places in networks including the edge. This is similar to our ganglions and lymph nodes example mentioned early on. Network technology planning is taking bandwidth into account, and is seeking to address some of the emerging latency needs. But it has not fully recognized the implications of edge requirements, conflation of the control and data planes, requirements for multiple layers of orchestration and a network immune system. The danger if these issues are not addressed and anticipated is that CSPs, in a fashion similar to AT&T's approach with the introduction of the iPhone, will be seen to have "failed" and will have to start an expensive journey to recover.