Looking Down from the Clouds

John Chidgey
B Eng Elect (1st Hons)
AOCP(A), RPEQ, MISA
5th August, 2018

The advent of generic personal computers running application software for the purposes of control systems HMIs (aka SCADA) is a practice than began during the 1980s, however became the preferred method of human machine interface with a control system by the 1990s to such a degree that customised user interface devices are used only in a handful of situations.

Usually when ruggedness and harsh conditions drive a requirement for physically hardened screens or physical buttons that a personal computer can not provide. This desire led to a dislike for custom HMI devices since an organisation could pay for a SCADA software licence once, then a relatively generic PC could run that application and be inexpensively replaced as needed.

As the decades have passed not only that, but the scale of control systems have also grown to a point where some control system networks equal or exceed the scale of their IT counterparts. Wherever that balance may lie for any given organisation, a divide has developed between two parts of organisations with control systems infrastructure: the control system group and the IT department. The IT department have been in cycles of insourcing and outsourcing for some time with their usage and support patterns having common equipment, common applications, common scripts and so on. IT have been wrestling with Service Level Agreements (SLAs) with IT support companies for years now to varying degrees. Conversely the control systems groups (sometimes referred to as OT: Operational Technology) have to learn every subtle variation of the highly customised equipment they are entrusted with maintaining. In this world in-house knowledge becomes precious and outsourcing considered to be too high a risk and a world where vendor relationships are critical for survival and SLAs are few and far between.

IT Pull Ahead

As decades have passed IT departments have come under increasing pressure from cyber-security threats to keep on top of the obsolescence of their equipment. Businesses demand internet connectivity which drives a huge attack surface for those wishing to do harm from the outside and it must be addressed. In addition the IT sector has been recently offered a simplification of their lifestyle in the last decade: cloud hosting. The old staple of IT departments, budgeting for 3, 4 or 5 year refresh cycles for their server equipment in particular could be changed to a single flat monthly fee as an operating expense with Microsoft Azure or Amazon Web Services (or others) hosting their virtualised machines or through a progressive migration of their services to platforms hosted in the cloud (eg Office 365). Whilst the raw cost economics of this move are still subject to some debate there is no doubt that this leaves IT departments with only standardised

physical end point machines (laptops/desktops) to maintain and replace, which is scriptable and regularly outsourced.

This structure means IT can focus on application updates, have less internal networking requirements and by signing up to an SLA with the cloud hosting provider, the availability of core IT services is no longer directly their problem to ensure uptime. If there are outages in the cloud, they are generally brief and the business impact, whilst non-zero, is ultimately recoverable with minimal risk to the organisation or at least at a risk organisations are prepared to accept.

OT Left Behind

If control systems HMIs had evolved into customised HMI devices that weren't PCs then there would be no discussion to be had. In that reality, there was no choice but to pay the higher prices for the customised HMI device to operate and maintain the facility. However the server and desktop as well as network switch hardware used for OT is often the same or very similar to that used by IT which inevitably leads to the question periodically: can IT and OT converge?

Control systems that operate chemical plants, oil refineries, gas compression facilities and water treatment plants are designed to have a very high availability and a long service lifetime. In some cases controllers can run without incident for 30 years. PCs on the other hand tend to struggle beyond 10 years and that's if a top of the line model is purchased initially, kept in a temperature and humidity controlled environment and fed from filtered, clean mains power for that time.

Equipment that OT groups are charged with maintaining is highly customised and the outcomes of its failure are often dire, hence there is a great level of caution when making changes without understanding all of the detail first. For this reason great care is taken using Management of Change (MoC) reviews with technical design reviews of detail down to the code level in some cases, to ensure that no individuals or operating equipment is put at risk before a change is made. Indeed even rebooting a network switch can lead to a lack of visibility of an operational site and depending on how well designed that site is, it may cause a site to shutdown unexpectedly.

This leads to a drive away from regular updates of software, firmware and hardware that ultimately means that PCs that run the HMIs are rarely, if ever, updated at both a hardware, firmware, operating system and application software level. When they are updated it is carefully undertaken often running HMI PCs in parallel during an upgrade or upgrades are only permitted during complete facility shutdowns when the risk of a loss of visibility is minimised.

Beyond these things HMIs are the windows into the controller and control system without which, no human can see the entire process at a glance. Worse than that, trends in recent decades have shown that there is a growing trust in SCADA HMIs, and more traditional local indicators on instruments like flow meters and pressure transmitters as well as purely mechanical check gauges are being removed from designs to cut costs, simplify and to keep people physically out of dangerous areas. This trend has put an almost absolute reliance on HMI visibility for many plant components, where in the past instrument position was carefully planned to ensure a field operator could physically "walk the boards," this is becoming impossible to perform without the HMI at most facilities.

To address the risks posed by putting so much faith in the HMI there have been two predominant design patterns in the control system space in recent decades: redundancy of SCADA, and co-location of SCADA and controller.

Redundancy

The easiest option to relying too much on a single window into the system is to create a hot standby HMI. A Secondary machine runs in step with the Primary machine ready to take over at a moments notice. Further beyond redundancy risk is often further reduced by using Client machines that communicate with the Server machine, whilst the Server machine does the interaction with the controllers themselves. This means operators use the client machines, often much cheaper desktop PCs, and the servers (the most critical devices) are housed in temperature controlled environments and are increasingly being put on server grade machines chasing higher reliability. To hedge bets even further, multi-client setups are increasingly

common with two redundant servers connecting to two or more client machines, some clients being specifically web-based clients for remote access which is increasingly becoming a requirement.

Co-location

The larger the attack surface the more likely a disruption will occur given sufficient time. In the case of OT networks to reduce the risk of disruption of communication between the HMI server machine and the controllers that the server provides the (sometimes) only window into, the SCADA server is placed on the same local network as the controllers it communicates with. Often this is a closed network or subnet (for Cyber-security reasons) and it is often self contained within a physical facility with fencing around a defined boundary. Strict controls for planned digging at site and MoC is applied to changes to prevent and disruptions. Typically Uninterruptable Power Supplies continue to power the local server machines, their local client machines and controllers to ensure maximum uptime as well.

What About the Cloud

The risk of losing communication between the controller and the SCADA system is significant if you place the server equipment outside the plant boundary. Even if AWS have a relatively physically proximate instance (data center) there's still a significant amount of copper and/or fiber between the facility and the cloud infrastructure. The organisation then must place considerable faith in the telecommunications companies to rapidly fix cable breaks, switch failures and such in a effectively completely uncontrolled environment. Beyond that, much of SCADA relies heavily of accurate time stamping of messages and near-real time command/responses between SCADA and the controller and latency can cause erratic operation in most systems in use today. Unless the SCADA system has been developed specifically to be cloud hosted to handle the additional latency, it is unlikely to perform reliably if it is shifted to the cloud. Of course that's a function of time and industry pressure and whilst more cloud SCADA platforms are becoming available coming in to 2020, migrating between incumbent SCADA software platforms to alternative platforms can be an expensive exercise even if that risk is accepted.

One size never fits all. There are however some scenarios where cloud-based SCADA might work:

1. *Full Local Visibility:* The control system is indicated completely locally as the ultimate backup for a complete loss of SCADA visibility
2. *Full Automation:* A window into the controller has been completely designed out with automation and SIS (Safety Instrumented System) taking care of the plant/process under all operational conditions
3. *Full Diversity and Redundancy:* The cloud server infrastructure is distributed through multiple paths and data centers with full redundancy between all locations and no single point of failure exists

Full Local

For simple plants or highly segmented components of a plant or for simple processes, this is always a possibility. Whilst the trend towards cost-reduction away from local operation panels, having a manual override is always a good idea even if it comes at a price. That said, the larger the facility the less likely this will be the case. Unless it is designed in to begin with, such systems are difficult and expensive to retrofit.

Full Automation

There are some highly repetitive, greatly populous engineering tasks for which a sufficient amount of time and money can be invested into full automation, since the scale and reusability of that technology can be recovered through ongoing sales volumes and design reuse. A good example of this are vehicles from planes down to cars. The unfortunate part of control systems and OT networks is that they are almost always custom designed and built. Whilst not always the case, for the vast majority customised plant automation depth is stopped at a line of cost vs risk.

In many cases HAZOPs and LOPAs are performed to determine the risk and mitigation measures to ensure that a Safety Controller (SIS) can prevent Major Accident Events (MAEs) by shutting down a system or process before the worst can happen. Ultimately this isn't cost effective or even possible for every potential circumstance and the control system will rely on an operator at some point to intervene. The only way operators can be alerted to a condition is via the SCADA either visually or audibly. If the window into the system isn't there,

then the operator won't/can't be alerted of a situation and just as critically, can't directly intervene to prevent an incident from occurring.

It's true that there are some installations where control systems utilise direct paging from the controller to get an operators attention however the paging systems lack of flexibility and cost have ultimately driven these out of favour in recent decades. This leaves full automation highly improbable for facilities that are custom designed, which is unfortunately most of them.

Diversity
To satisfy diversity cloud providers would need to offer a fully diverse hosting platform with multiple data centers in different parts of the same city or different cities, each interconnected and cross-connected via different physical/geographical routes, independent power supplies for each, with application redundancy applied geographically and forced via independent paths. The key is that there can be no common failure mode and as unlikely as it sounds, common failure modes include denial of service via network overload and telecommunication carrier loss of core switching functions, which although unlikely has still occurred. The far more likely common failure mode is large scale brownouts or blackouts, with many points requiring power to function between the data center and site, each of which requires power to carry the data traffic, it doesn't take much for this to fail.

A Hybrid Approach
Rather than give up entirely on the OT cloud idea, given the extreme unlikeliness of the above three to be true for large scale facilities, can a hybrid approach reach a compromise? Provided there are redundant server machines co-located at the facility where the physical controllers are located then all other server machines could be cloud-based, then does that provide the best of both worlds? There will inevitably need to be physical machines, either set up as clients or as thin clients RDPing into virtualised client machines, however they're implemented, these need to also be co-located unless the risk of letting operators use the server machines is considered to be acceptable - which is unlikely.

Beyond SCADA there is also the requirement for local programming via what is often called an EWS: Engineering WorkStation. This also needs to be locally connected to allow for programming and diagnostics to be performed on the local network and often comes with specialised software and a copy of the sites control system code and project files. At this point, physical equipment co-located is a requirement due to plant complexity, lack of local indications and controls and hence we are up for physical host server machines co-located at site.

If this is inevitable then we can make these servers as large as practical, potentially on a Hyper-converged architecture (eg Cisco HyperFlex of HP SimplifVity) with multiple levels of redundancy and entry-level desktop thin clients RDPing to the VMs they host. Any machines used as data repositories and historians can handle a short-term loss of communications with simple local data buffering so they could be moved to the cloud and anything that operators don't require to operate the facility on a minute by minute basis could also be migrated to the cloud. However no matter how you look at it, the Hybrid approach can only go so far and cost savings come more from the Hyper-convergence of infrastructure only if the scale makes cost-effective sense.

Conclusion
As IT are looking down on OT from the Cloud above, OT are left to grapple with the issues that many IT departments have long left behind with physical machines co-located by necessity not by choice. New plants may be designed (in some instances) from the ground up with local controls and indications and sufficient risk treatments to allow for cloud based hosting of HMIs. That said the true end-to-end costs of such a decision need to be considered such that the benefits in cost reduction of cloud-hosting must be significant to justify the additional local instrumentation, design and maintenance costs and personnel risk in the field to enable full OT cloud hosting to be done safely.
However companies that drive for IT/OT convergence for existing facilities for which it was never designed take on significant risk in attempting to push OT equipment into the cloud. Considering all of the implications is critical to success and failure to do so properly can only end in failure.

Peer reviewed by (and with thanks to):

Peyman Radnia
RPEQ, FIEAust,CPEng, TUV FS Eng
October 2018