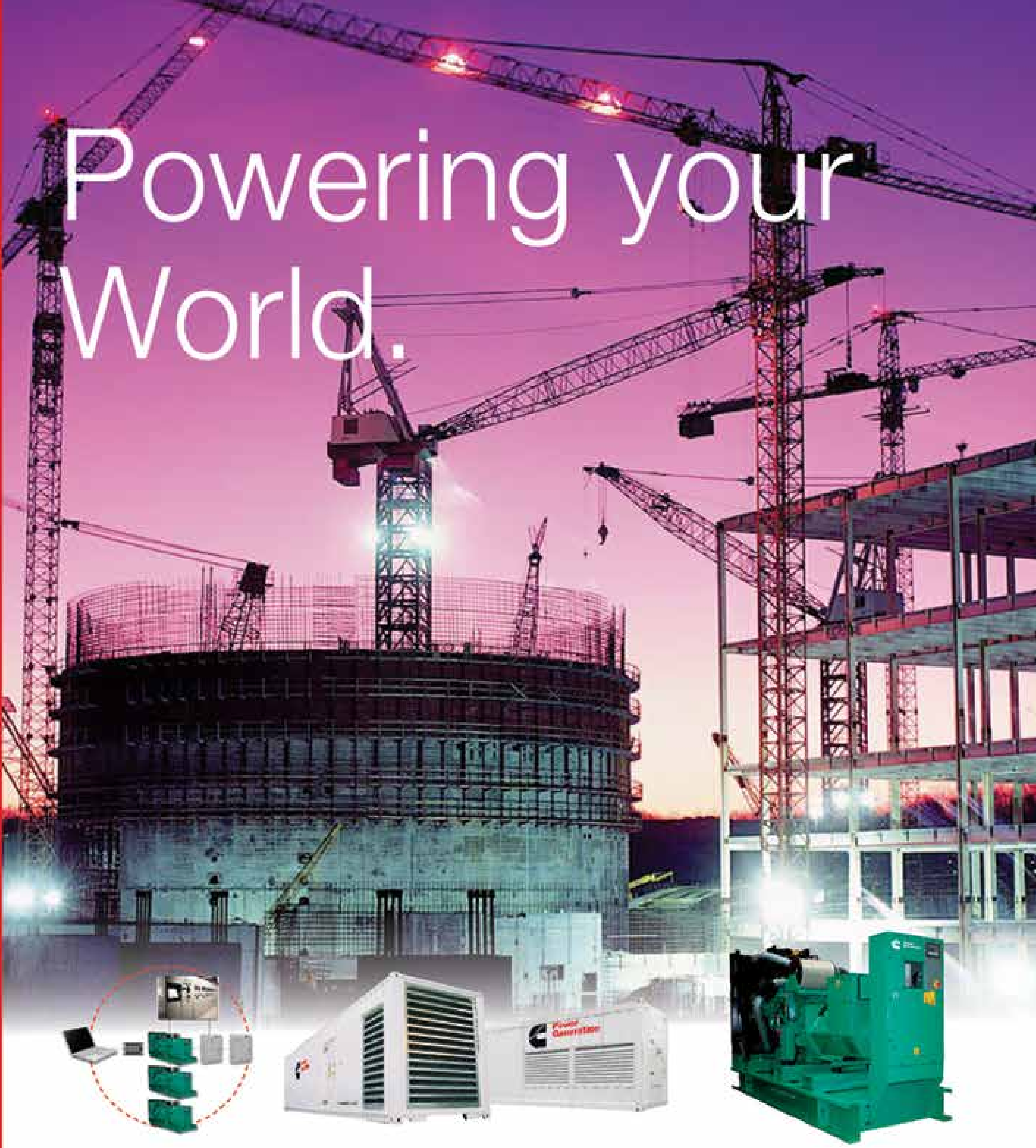
Industrial Communications Handbook

Professor Alan Robert Clark



Powering your World.
Power Generation

Comment

by Professor Ian Jandrell

As with everything in our industry, the way devices communicate continues to evolve. It is also true that over the past decade in particular it has become increasingly important to ensure that data from every part of your process is gathered and analysed.

Why? Because that data becomes the information that helps you optimise the bottom line.

Industrial Communications is more important than it has ever been.

We no longer monitor only levels and temperatures; we monitor energy usage, the location of our vehicle fleet, and the quality of almost every commodity we use – including our own time.

The network has changed significantly in some ways. It has also been impacted by the Internet of Things (IoT) and other emerging trends that often gain traction in less conservative and critical environments first.

As such, the realm of Industrial Communications is very different from what it was a few years ago. This implies the need to understand trends, to appreciate the new technologies that are increasingly coming to the fore, but also to understand the limitation associated with these developments.

The 2016 Industrial Communications Handbook moves away from the Physical Layer and modern high speed copper-based communications systems. Now we need to understand wireless communications – and with that comes the need to appreciate the way wireless works, and how we can secure our networks.

Alan Clark, the compiler, leads you through the fascinating maze of modern industrial communications in a way that is accessible and interesting. Alan has a unique way of making difficult topics simple to grasp.

I would also like to express my sincere thanks to our Project Co-ordinator Wendy Izgorsek, and to our consultant, Tim Craven.

I am certain that you will enjoy this Handbook, and that it will take pride of place next to the previous editions. Taken together, these Handbooks provide a powerful summary of trends in our industry over the past decade – and will continue to provide the insight you need to better manage your plant for many years to come.

Enjoy! Ian

Authors: Alan Robert Clark and Tim Craven
Editor: Wendy Izgorsek

Published by **Crown Publications cc**
PO Box 140, Bedfordview, 2008
Tel: +27 (0) 11 622 4770
Fax: +27 (0) 11 615 6108
e-mail : admin@crown.co.za
Website: www.crown.co.za

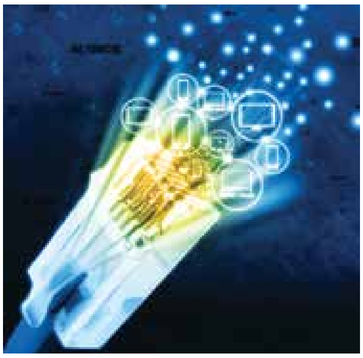
Printed by Tandym Print

Contents

Chapter 1
Introduction

Industrial Communications.

- What is it?
- What makes it Industrial?
- What is being Communicated?
- How?



These are difficult enough questions, and it is important to note that they have not changed much over the years that the Industrial Communications Handbook has been published.

The answers have changed. Dramatically.

1.1 History

In 2005 [1], this Handbook was mainly about the Physical Layer: RS232/422/485, OSI model; and Protocols: HART, MODBUS, AS-I, DeviceNet, InterBus, ProfiBus, FieldBus; and touching on that Awful, Doomed Approach of Ethernet and TCP/IP. Bus Wars at full strength.

In 2008 [2], we saw a 'much of the same' approach, but with a significant reduction in the hardware aspects, and a greatly expanded view on Ethernet, overcoming the major disadvantages of an un-timed Bus by good use of Switched Ethernet, as opposed to mere hubs of the past. In addition, timing was improved by protocols such as EtherCAT. The Awful, Doomed Approach of Wireless (WiFi, Bluetooth, ZigBee, WirelessHART, ISA100, etc) was rearing its ugly head.

In 2013 [3], we started with Ethernet as the only really important thing to worry about, with emphasis on the Protocol that was to be run *ON* Ethernet. Wireless, or Not. Great emphasis is then placed on Mesh networking, and the self-healing ability of the network, of massive importance in a Wireless environment. The Awful, etc, etc, is still an automatic assumption that Wireless is better than Wired under all circumstances; as well as the nebulous 'Smart Grid', which all agree is very important, but no one agrees just what it is (or isn't).

1.2 Going forward

(Don't you *Hate* the term? :-))

Where are we in 2016? The three Whats, and one How, at the beginning are definitely still the Questions.

We have a better idea of what we require from the 'Smart Grid', especially driven by Renewables, which are fickle, and change far quicker than traditional Grid Stability demands. We have visions of the Internet of Things (IoT), where your toaster tells your fridge to order more bread (Ok, pushing it a bit). As this is *Industrial*, we re-define that as Industrial IoT (IIoT). Apart from the obvious insertion of a 'd', what does that mean?

EtherNet itself has come to the party. Not simply relying on being 'fast enough', or 'switched enough', we have Quality of Service protocols such as Time Sensitive Networking [4] and automatic encryption methods in Trusted Wireless [5] that recognise that wireless ethernet is here to stay.

Energy is very much in the picture, the era of 'free' electricity, oil, etc being largely over, how does IIoT save energy smartly? Industrialisation of the Mining process asks how IIoT is applied in the (Electromagnetically) harsh environment of a mine. What you do in a Factory is *VERY* different, but not appreciated by those not Electromagnetically inclined. Robotics can be beautifully controlled by this marvellous system, and like any system, can be marvellously Hacked.

Cybersecurity is not typically at the top of the agenda in traditional Industrial Automation. A suitably hacked network can be used to determine Proprietary Control strategies (Chemical reactions in your soap powder), simple Industrial Espionage (How *MUCH* soap powder you make), and, of course, messing with timing on water valves, heating cycles, emergency shutdowns (Messy, diluted soap slosh).

Like Factories, Buildings are not what they used to be. We need Green Buildings with lighting, energy distribution, information distribution, heating, ventilation and air conditioning, all being addressed by the same IIoT.

All this means something quite a lot different from simply choosing between RS-232 and RS-485!

1.3 'Greenfield'

HOW would one go about specifying, in this day and age, a 'Greenfield' environment, a Gee-Whiz Automation project?

Assume a warehouse-sized facility, a good number of valves, pumps, temperature-level-mass-whatever sensors, ingredient actuators, robotic bottlers, and a Good Olde Fashioned 3-phase supply, with Photovoltaic thrown in for good measure.

What would be measured? What would be controlled? What would make the measurement particulary critical in an Industrial sense? What information would an operator need? Management? Energy Auditor? Energy Backup Strategist? CyberSecurity Auditor?

With the amount of data that can be collected, stored and 'mined', what questions could be asked? What is the saving on my soap powder input costs if I tweak the pH of the surfactant? What is the increased failure rate of the pump? Is it worth changing?

What is the benefit of increasing the surfactant tank size, and only pumping the stuff up the hill at

03h00? What energy offset rate makes it financially viable?

All this becomes available for future analysis if the sensor data is suitably collected, suitably transported using your network of choice, suitably categorised, and suitably stored.

Alternatively, if you are currently stuck with Bus 'x', where to from here? The various Bus's from the Bus War days ain't what they used to be, and have moved on dramatically.

How does one 'break the mould' of 'its always been done this way', when there is a sudden need to double the output of soap powder. Worse: what happens when no one uses soap powder anymore? Think Kodachrome :-)

1.4 Wrap-up

Where are we in the South African framework?

The chap with the screwdriver that used to crawl into awkward bits of the plant to 'set the zero' or 'set full-scale', has put a tie on and sits puzzled in front of a computer muttering about latency and firewall rules.

But surprisingly, one hears tell that although everyone has 'gone digital', a large number of plants have simply got their toes wet: using digital info on top of the olde analogue: digital minimalism perhaps? Ludwig Mies van der Rohe look out!

This certainly gets the job done, levels are checked, pumps controlled, and valves set; but the *metadata* is missing: the ability to model the plant, tweak one input and see what happens to overall cost, or time, etc …

So: a large part of industry still runs 4–20 mA.

The Bus Wars were thought to be settled, that the 'One Bus to Bind Them All' would be found. Ethernet, surely? But the trouble with Ethernet is precisely its universalness. It can carry anything. Where we were expecting a shut down to One, we have instead a proliferation of special-purpose protocols that may run on Ethernet, but do not interact with one another. This 'Lock-in' mentality also dominates the software world, where exclusivity is still seen as a mechanism to lock customers into a particular company's offerings.

Open Source movements, and to a lesser extent, Open Hardware movements have spearheaded attempts at setting Standards, or at least getting interoperability. But the trouble with Open Standards is that they *ARE* open, and hence easily changed, or 'forked'. This remains a challenge.

Wireless has certainly taken off, but suffers greatly from inappropriate use, and ignorance of basic physics. I have seen a high-gain 'omni' antenna bolted on a mine wall, and a dipole on a DIN-rail in a fully metal-plated enclosure. WiFi, it may be; Magic-Fi, not so much.

Many, many years ago, I was involved in controlling a 5 m diameter inclined tube mill, where the crushed ore-height was the measured control variable in the tube. The eventual solution was, wait for it, *slip-rings*. Another maverick project was slope-control/monitoring in a quarry. Wireless would have been an absolute killer-app for those!

This, Fifth Edition, of the Industrial Communications Handbook, attempts to address some of these issues. Chapters 1–3 written by Alan Clark, cover the basics of radio communication.

Chapter 4–6 written by Tim Craven cover the all important aspect of Security, as well as Greenfield challenges.

Chapter 7 looks at some of the changes that make Ethernet a better fit to Industrial Communication.

Chapter 8 finishes off with some concluding comments.

DATALOGIC

Chapter 2
Radio Basics

So many 'networking' installations—Industrial, or otherwise—end up 'going wireless' for all the wrong reasons. A perfectly good setup gets 'upgraded'—and fails dismally.

Most often, it is the simple neglect of the Basics, the elementary Physics, and especially the (not-so) Common Sense that is lacking.
'Going Wireless' can seriously add benefit to an Installation, ***BUT*** it must be *designed*, and not just slapped together with little regard to actual radiation and propagation constraints.

2.1 Time, length, phase

We start with an odd concept that permeates all communication at high frequencies: Time is equivalent to Length which is equivalent to Phase.

Take the single cycle of Eskom's power shown in *Figure 2.1*.

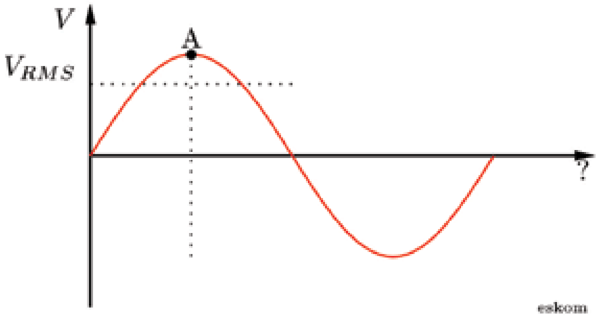


Figure 2.1: Single cycle of 50 Hz, 230 V.

The y-axis is voltage, the Root Mean Square (RMS) value is 230 V, hence the peak (at point A) is 398, or 400 V for short. (Previously Johannesburg was 220, hence 380)…

But what of the x-axis? *IF* it were time, point A would be at 5 ms, since a full cycle at 50 Hz is 20 ms. *IF* it were degrees, then point A would simply be called 90°. *IF* it were length (free-space wavelength), point A would be 1 500 km, since a full wavelength at 50 Hz is 6 000 km.

So point A is simultaneously 5 ms, 90°, 1 500 km, depending on your perspective. In addition, we would call point A a quarter wavelength, or λ/4, for short.

The corollary is that in order for something (at high frequency) to take time to travel to the other end, or to generate phase while doing so, it must be long *(in terms of wavelength)*.

Clark's Rule-of-Thumb is that a 50th of a wavelength begins to require the use of Transmission Line Theory, as opposed to Circuit Theory for shorter things *(in terms of wavelength)*.

Essentially, the speed of light, *c*, is fast, but not *that* fast! A mere 3×10^8 m/s or only 300 000 km per second. But it is finite, and if a length is appreciable *in terms of wavelength*, phase is accumulated, and causes havoc. The higher the frequency, the shorter the wavelength, and the earlier the havoc!

An example is a quarter-wavelength (λ/4) of transmission line representing 90°. Assume it is open-circuited. A wavefront will travel down that transmission line, collecting 90° of phase; it will then reflect at the open-circuit, and come back, collecting another 90° of phase on the way. When the reflected wave reaches the sending end, there is precisely 180° of phase difference—exactly out-of-phase—an open-circuit at the end of the transmission line has magically become a short-circuit at the start of it!

If, even at 50 Hz, one were to connect Johannesburg to Durban directly, and then again via Bloemfontein, the difference in the path lengths would lead to a difference in phase, and grid instability would be the result, if not carefully managed.

At much higher frequencies, like WiFi, the difference in path lengths between a direct path and a reflected one (off another object, like ground) becomes a mess, unless very carefully designed around.

At even higher frequencies, it takes sunlight about eight minutes to reach the earth. What that means is that the beautiful sunrise you are watching has already happened …

2.2 Wavelengths, antennas, etc

Now it turns out that in order to be fed nicely, an antenna needs to be quite long so that it resonates, and radiates nicely. Such a dipole antenna has a sinusoidal current distribution on it when it is a half-wavelength long (λ/2 long). Naturally this depends on the frequency given by Equation 2.1.

$$\lambda(\text{m}) = \frac{300}{f(\text{MHz})} \tag{2.1}$$

At 300 MHz, λ=1 m, and λ/2=1/2 m. Other interesting sizes are shown in *Table 2.1*.

Table 2.1: Frequency and 'interesting' wavelengths.

f (MHz)	λ	λ/2	λ/4
200	3/2 m	3/4 m	3/8 m
600	1/2 m	1/4 m	1/8 m
2 450	122,4 mm	61,2 mm	30,6 mm
5 800	51,7 mm	25,9 mm	12,9 mm

Remembering that λ/4 represents the case where an open-circuit transforms to a short circuit, at 50 Hz (0,000050 MHz), this is 1 500 km, roughly the distance between Cape Town and Ogies, the centre of our generating capacity. So when some nice chap switches Cape Town off the grid, Ogies is in trouble. Grid stabilisation is a challenge on long distance transmission, hence the need for HVDC.

At WiFi frequencies, this calamitous situation occurs at a mere 30 mm in free-space.

Figure 2.2 shows comparative sizes of Sleeve Dipoles at 2,45 GHz, and 5,8 GHz. These are usually terminated with an SMA connector, and this shows the clear dependence of size on the frequency.

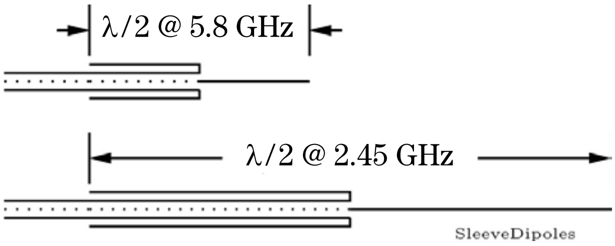


Figure 2.2: Half-wave (sleeve) dipoles.

From *Equation 2.1*, your TV1,2,3 antenna at 200 MHz has elements 3/4 m long, your MNET antenna has 250 mm elements, your WiFi at 2,45 GHz is at 61 mm, and at 5,8 GHz, it's at 26 mm.

Everything in Electromagnetics scales exactly as a function of frequency.

For the vast majority of Industrial Communications, we deal with the unlicenced ISM (Industrial, Scientific and Medical) bands of 2,45 and 5,8 GHz.

So the *choice* of antenna depends very strongly on the frequency of operation.

2.3 Radiation

Quite *what* causes radiation, we don't really know, but we do put forth some theories, almost always associated with accelerating charged particles. What we *do* know is *how* to get it radiating: Time, Length, Phase. (You may have heard that one before …)

Essentially, if we take a transmission line, and split it apart, so the conductors are more than a tenth of a wavelength (λ/10) apart, radiation will happen (intended, or *NOT INTENDED!*). This is illustrated in the simple alpine horn antenna in *Figure 2.3*.

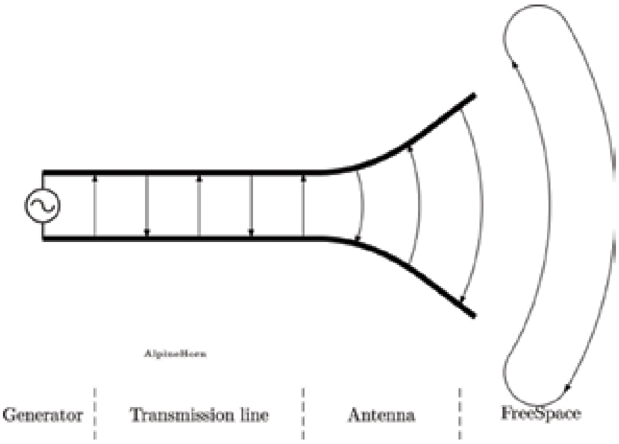


Figure 2.3: Simple Alpine Horn explanation of radiation.

Note that the radiation is launched in a particular polarisation, vertically in the direction of propagation. Additionally, Maxwell tells us that Electric fields get lonely without an accompanying Magnetic field in the plane 90° away from both propagation and the electric field. Thus, sufficiently far away from the antenna, both the electric and magnetic fields are *transverse* to the propagation, as shown in *Figure 2.4*.

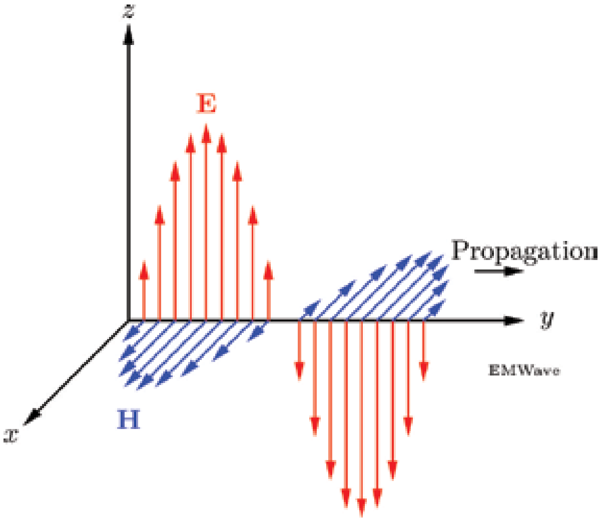


Figure 2.4: A Transverse ElectroMagnetic (TEM) Wave.

Plastic, stainless steel, die-cast zinc: the universal I/O system with IP 67 protection for all applications.

www.beckhoff.co.za/ip67

The wide range of extremely compact and robust Beckhoff I/O modules with IP 67 protected housings is designed for all industrial applications. The broad spectrum of signals ranges from standard digital I/O to complex analog technology to compact Drive Technology.

- **Fieldbus Box (plastic, IP 67):** 12 different fieldbus systems for universal application
- **EtherCAT Box (plastic, IP 67):** high-performance for all applications, directly on the machine
- **EtherCAT Box (stainless steel, IP 69K):** for hygienic applications in the food, chemical or pharmaceutical industries
- **EtherCAT Box (die-cast zinc, IP 67):** for harsh environmental conditions in "heavy-duty" industries

IPC
I/O
Motion
Automation

Beckhoff Automation (Pty) Ltd
Randburg 2194, South Africa
Phone: + 27 (0)11 795 2898
info@beckhoff.co.za

New Automation Technology **BECKHOFF**

2.4 Polarisation

Not only does the antenna size determine the frequency of resonance, but its shape determines the polarisation of the radiated wave.

It is important to note that all man-made radiation is essentially polarised. Astronomical sources (stars, pulsars, quasars, black holes) are generally un-polarised, but such sources are impossible to manufacture. Polarised sunglasses remove the components of the sunlight that are not vertical, thereby removing most 'glare' which is typically horizontal, from e.g. water sources, etc.

A simple dipole produces *linear* polarisation, with far-fields that look like *Figure 2.4*. Since we are ***E***-field-centric, (and the fact that the ***H***-field is 377 times smaller!), we can speak of the field in *Figure 2.4* as being *vertically polarised*, as shown in *Figure 2.5* (e.g. FM radio).

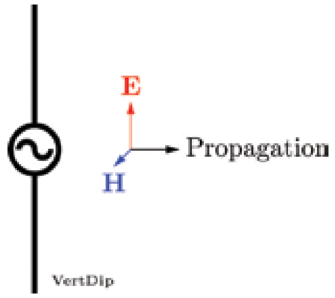


*Figure 2.5: Vertical dipole showing vertical (**E**-field) polarisation.*

If we placed the dipole horizontally (parallel to the ground), the linear polarisation would be horizontal (e.g. TV).

It should then be clear that a *horizontal* dipole will receive absolutely nothing from a *vertically* polarised transmitter.

The corollary is that since it is unlikely that absolutely the same polarisation is used for both transmitter and receiver, there is always some *polarisation loss*, a.k.a Murphy's Law.

So the polarisation, or orientation, of the antennas on both sides of the communications link is important. It becomes more complex in a real environment with many antennas, since radio waves bounce off obstacles, and interact, causing fading and changes in the polarisation of the wavefronts.

We can generate circular polarisation by various means (crossed dipoles, helices, patch antennas with offset feeds), so the polarisation loss (at boresight) is constant: a vertically polarised antenna will have a 3 dB loss, as will a horizontally polarised antenna.

Circular polarisation sounds like a good idea to manage the widely variable polarisation loss, but it must be remembered that (a) a 3 dB loss is half the power, and (b) in any direction other than boresight, it is no longer circularly polarised. In the extreme, at 90° to boresight, the polarisation is again linear.

Many broadcast scenarios utilise 'mixed' polarisation at the base station, in order to give the portable transceiver more options (e.g. Cellular).

An extremely useful case for circular polarisation is down a tunnel, as the extreme nulls do not occur as with linear polarisation, which bounces off the walls, floor and roof of the tunnel. All mine-based Industrial Communication ought to be designed using circular polarisation for this reason. (But rarely is!)

Linear polarisation, particularly 'high-gain omni' antennas are a complete disaster in a mining setup, at least the tunnelled variety.

2.5 Radiation Pattern

The radiation pattern of an antenna attempts to show how an antenna radiates in three-dimensional space. It is purely a function of angle, and nothing is implied as to how far the radiation goes.

It's all a matter of angle. If an antenna radiates better in one direction than another, it is said to have Gain in that direction. Gain is a most unfortunate word since it implies that the antenna is active: i.e. it generates power of its own! In reality, an antenna is a passive device; cannot manufacture power; and the term Gain simply refers to the concentration of power *in one direction* at the expense of power in other directions.

Gain is 'Robbing Peter to pay Paul.'

Gain is measured against an isotopic source that radiates equally well in all directions. Notably, this does not exist, but it is a good reference value which translates to a gain of 1, or 0 dBi.

Gain that is above this reference level is greater than the isotopic, measured in deciBels, and hence positive dBi (in that direction). Gain below this value is thus negative (in that direction).

Integrating the power over an enclosing sphere must therefore always give 0 dBi. What you win on the swings, you *MUST* lose on the roundabouts.

In industrial communications, it is usual to use simple dipoles, as shown in *Figure 2.2* (they look like monopoles, but are not!) since the placement of the equipment is not known.

The Radiation Pattern of a dipole is a doughnut, with the vertical dipole upright in the 'hole' of the doughnut, as shown in *Figure 2.6.* The maximum Gain, in the Azimuth plane, is 2,16 dBi, or 2 dBi for short.

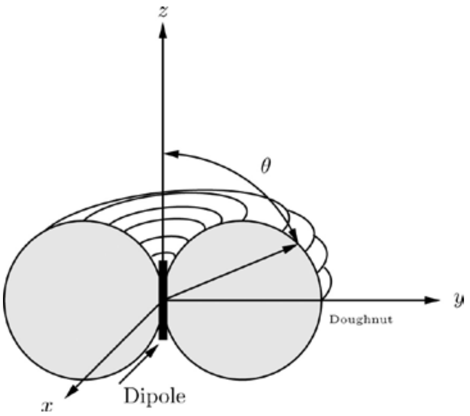


Figure 2.6: Doughnut radiation pattern of vertical dipole.

The problem, of course, is the 'hole' of the doughnut. The dipole does not radiate at all in the axial direction (up and down). So the only reason a dipole has a positive gain of 2 dBi around the middle is because it has massively negative dBi North and South.

Therefore a 'High-Gain antenna' simply must radiate incredibly badly elsewhere. It cannot radiate at high gain in all directions! A High-gain Omni is an oxymoron, unless it is understood that it is Omni in only one plane.

A common antenna with a high gain is of course the Yagi-Uda array, or Yagi, stalwart of terrestrial TV reception. The majority of the radiation occurs in a single beam, but significant amounts of power are still radiated elsewhere: Murphy again. The main beam half-power points at (–3 dB), measured in degrees give an indication of the Gain: the higher the Gain, the narrower the beam width. Remember, you are Robbing Peter to pay Paul. The higher the gain, the more difficult it is to 'point' in the correct direction.

Thus, if your transducers etc are on the periphery of your plant, and the control communications hub is central, directive antennas may be more useful, communicating to an omnidirectional antenna at the centre point.

One more point about Gain, it not only concentrates the energy where you want it, but it also concentrates it *away* from where you don't want it. This is the very simplest form of data security, which, when combined with power control, is often overlooked as 'too trivial', but is vitally important. Do *NOT* automatically set all wireless activity to 'max power'.

So radiation pattern gives an idea of where to point an antenna. It also gives an idea as to where NOT to point the antenna. Remember that metal will reflect any EM wave thrown at it. Hence putting a vertical Omni on the metal walling of a mine tunnel is just plain silly.

Putting an electricity smart meter with a GSM antenna in a metal cabinet likewise. (Yep, its been done …)

Chapter 3

Applying Wireless in Practice

So now that we have covered some of the basics, how do we go about using the stuff intelligently?.

3.1 How far?

The next important question is that of coverage, just how far will it go? Notably, the Radiation Pattern tells you nothing about this. The distance it will travel is largely simply dependent on $1/r^2$. Hence, as a first approximation, the Friis, or Free Space Link *Equation 3.3* gives a reasonable prediction.

For an outdoor situation, the point-to-point link is easier to visualise and plan. Indoor propagation, with multiple reflective and absorptive surfaces becomes an absolute nightmare. Different dielectric surfaces behave differently, depending on frequency, and hence size.

In the extreme analysis, a human is just a large potato walking around a 2,45 GHz microwave oven that you call your plant. All standing wave patterns in the plant are constantly changing as you walk.

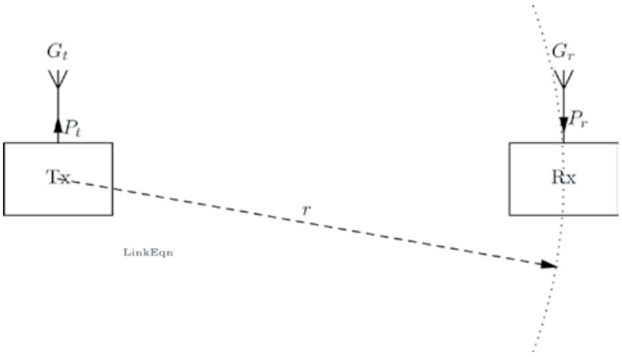


Figure 3.1: Effective Radiated Power (ERP) and the Link Equation (Friis).

Breaking the communication into what is transmitted and what is received is useful: *Figure 3.1* shows that from the receiver's perspective, it is simply sitting in an electromagnetic field of a certain strength.

This field strength at the distance r away from the transmitter is known as the Effective Radiated Power (ERP), given by the first part of the link equation, as shown in *Equation 3.1*. (Pedantically, EiRP, for isotropic …) In log form, it becomes a lot simpler, as we add the dBs as in *Equation 3.2*.

$$\text{ERP} = G_t P_t \tag{3.1}$$

$$\text{ERP} = G_{tdBi} P_{tdBm} \tag{3.2}$$

Thus, from the receiver's perspective, the field strength at r, as given by equations 3.1 and 3.2 could equally well come from a 1 mW (0 dBm) transmitter feeding a 20 dBi Yagi antenna, or a 100 mW (20 dBm) transmitter feeding an omnidirectional antenna!

To increase the ERP seen by the receiver by 3 dB (double the received power), means either increasing the antenna gain by 3 dBi, or increasing the transmitted power by 3 dB (double the transmit power).

The power then received by an antenna in a freespace point-to-point link is given by *Equation 3.3*.

$$P_r = \frac{G_t P_t \; G_r \lambda^2}{(4\pi \mathrm{r})^2} \quad [\mathrm{W}] \tag{3.3}$$

It is much easier to express the Freespace Link Equation in dB form, as shown in *Equation 3.4*.

$$P_r = P_t = G_t + G_r - 32.45 - 20\log_{10} f - 20\log_{10r} \tag{3.4}$$

where P_r and P_t are expressed in dBm, G_t and G_r are in dBi, r is in km, and f is in MHz.

Assume a wireless transducer with a 13 dBm power into a dipole (2 dBi), operating at 2,45 GHz, to another dipole at 100 m. The received power would then be

$$P_{rdBm} = 13 + 2 + 2 - 32.45 - 67.78 - (-20)$$

or –63,23 dBm, or –69,25 dBm at 200 m (0,2 km). At 1 km, this is –83,23 dBm, a full 20 dB lower, way below reception quality on most cheap hardware.

A popular brand of receiver requires –68 dBm to achieve 130 Mbps in IEEE802.11n mode, but can go as low as –85 dBm if only 11 Mbps is required from IEEE802.11b.

Thus, not only will *Equation 3.4* tell you how *FAR* you may go, it also gives an indication of how *FAST* you can go over the distance.

In a similar vein to the ERP discussion, increasing your receiver sensitivity by 3 dB is the same as increasing your receiving antenna gain by the same amount.

3.2 Line of sight

Applying the Friis equation has two main application arenas: outdoor, and indoor.

Outdoor, or at least a relatively large distance, with not much inbetween, the Friis link equation, given above in *Equation 3.4* works best. But it depends on line-of-sight. Now this is not as simple as it may at first seem. At least 60% of the first Fresnel zone must be clear of any obstruction, otherwise the link will be intermittent at best.

Remember that the Fresnel zone is a three dimensional ellipsoid between the transmitter and the receiver. The earth gets in the way, trees, hills, buildings …

Assuming a symmetrical link, with the worst obstruction in the middle of the link (e.g. earth), we can get the First Fresnel zone radius in metres from *Equation 3.5*.

$$F_{1(\mathrm{m})} = 8.656 \times \sqrt{\frac{D_{(\mathrm{km})}}{f_{(\mathrm{GHz})}}} \tag{3.5}$$

Or, since we are only really interested in 2,45 and 5,8 GHz, this reduces to:

$$F_{1_{2.45}} = 5.53\sqrt{D}$$

and

$$F_{1_{5.8}} = 3.59\sqrt{D}$$

The first Fresnel zone under these conditions is shown in *Table 3.1*.

Table 3.1: Fresnel zone radii in metres at the centre of a symmetric ellipsoid.

Link distance [km]	F1 [m] 2,45 GHz	F1 [m] 5,8 GHz
1	5,5	3,2
2	7,8	4,5
3	9,6	5,5
4	11,1	6,3
5	12,4	7,0

Table 3.1 thus also shows why the higher frequency is most often used for longer distance links, as no one wants a 13 m mast at both ends of a 5 km link! (There are other considerations, but this is a positive!)

A point-to-point link that works perfectly in Winter may not work when Summer comes around. Over a kilometre or so link, the refractive index of the air changes, especially with temperature, changing the so-called 'K-factor' which accounts for earth curvature. So the height at the ends of the link may need adjustment. A fully redundant system has several antennas, positioned vertically to account for this. An example is the classic Microwave towers seen dotted around the country for telephony in the days before fibre.

Finally, remember that 2,45 GHz is the frequency used in your microwave oven to cook your potato, designed to make water molecules within it vibrate and cause heating. Since WiFi is at the same frequency, and a tree has leaves containing water, a highly directional link through a tree works in winter in the absence of leaves, but is useless in summer with luxuriant growth. It gets worse in the rain when all the leaves are nice and wet.

Rain, in itself, at these frequencies is a problem, hence the clamoring for TV frequencies that will be freed up after digital migration, as VHF does not suffer from such efficient absorption.

When planning an outdoor link, a link margin is crucial to the success of the link, and strongly influences cost. Hence, how much do you need to spend to get a reliable link? How long is a piece a of string?

A margin of 'only' 10 dB means 10 times the power.

3.3 Indoor and diversity

The majority of industrial communications will occur in an environment that would be classified as indoor. This is defined by lots of clutter, both metallic and nonmetallic. Metallic clutter introduces a fully reflective surface, and strongly reflects the electromagnetic wave, interacting with the strong forward signal, leading to interference: Constructive and Destructive!

Remember from before that a quarter wavelength, 90°, exists between a point of absolute destruction and beautiful addition. Thus, at 2,45 GHz WiFi, that is 30 mm.

On this basis, it can be seen that communication at this frequency in a busy environment is simply impossible.

The only way WiFi actually works is by having *Diversity*.

Diversity ensures that when one antenna is in a destructive interference zone, there is another antenna that can still receive. This of course requires two different radios, and the ability to be able to switch between the signals very rapidly indeed: requiring a computing platform to decide which signal is stronger. Even the cheapest WiFi router has an internal second antenna.

Diversity may be *spacial*, the two antennas sufficiently apart to ensure that they are not destructively interfered with at one time.

They may occupy the same space, but be differently *polarised*. (It is unlikely that both vertical and horizontal polarisation will be in destructive mode at the same time.)

The diversity may be in *frequency*, either within or across the 2,45 and 5,8 GHz bands.

They can also cleverly use *time domain* repetition, as a form of time diversity in a rapidly changing environment.

Naturally, there is all the above, essentially what IEEE802.11n, MIMO (Multiple Input Multiple Output) uses, but noting that it takes computing power either to select or combine the outputs from the antennas, which also consume electrical input power.

Hence intelligent MIMO devices turn off MIMO unless absolutely required in order to conserve power, especially if battery driven.

Another major 'indoor', i.e. not 'outdoor' problem is non-metallic attenuation. The problem here is the wall, cabinet, chair, passage way, tool chest, etc that gets between the transmitter and receiver. Sadly, a human looks like a lump of water at these frequencies, and unfortunately, tends to move about, changing the electromagnetic environment.

This challenge is only met by more power, greater antenna gain, repositioning, or adaptive mesh networking (getting around the obstacle/s).

3.4 Wireless coexistence

Radio systems do not exist in isolation.

We all share the same 'ether'. Even systems that operate at different frequencies can still interact by RF swamping of sensitive receiver stages, etc.

Of particular interest is narrow-band interference killing wide-band systems. A strong Bluetooth signal often kills WiFi. They operate in the same frequency band, but Bluetooth divides it into 95 channels through which it hops in time, whereas WiFi has only 11 channels (three non-overlapping), much wider, but static in time.

Although both spreading and hopping strategies were developed to reduce the possibility of intercept

(military), their combination is often troublesome. Since the 2,45 GHz band, in particular, is extremely noisy and busy, each additional transmitter simply increases the 'noise floor', thus making it more difficult for yet another transmitter to successfully gain access. Hence data throughput rates drop, and annoyingly, that temperature transducer sometimes works, and sometimes doesn't …

It is extremely important to note that the networking systems we have were designed for different purposes, and we ought not to use them for things they were not designed for. Of course, that is a red flag to a bull …Therefore many systems are inappropriately used. Bluetooth was not designed for video streaming. Internet protocol was not designed for low latency communications.

Remember that in the 1980s, Ethernet was all connected to the same network segment, and essentially all communication occurred by talking to everyone on that bus. When the jabber got too much, one put a bridge in, and separated the network into two smaller sub-networks, so that only communication that had to go over the bridge to the other side did so, freeing up each side of the bridge to allow more local communication speed.

Eventually, when costs plummeted, switches arrived, allowing each segment to be quite small, with only a few machines on a common bus. These days, it would be odd to have more than one machine on its own segment, with ALL communication effectively switched over a very fast backbone, such that few collisions occur.

Wireless takes us Forward, slap bang into 1980 all over again. Unswitched hubs is all the atmosphere offers, (OK, there are a FEW non overlapping channels).

In ordinary Telecomms, when bandwidth becomes an issue, lay another cable, it is a tad more difficult to lay another electromagnetic spectrum.

Chapter 4
Security of the Physical Network

Tim Craven

First things first. If you don't want something nicked, lock it up!
Why do we do something different when it comes to networking?

4.1 Communications technology of choice for mission-critical systems

A couple of decades ago, industrial grade communications were handled by serial connections and hardwired IO systems. These communications systems were point-to-point and severely limited by distances. Securing a communications link was a simple case of ensuring that no unauthorised person could access the physical cabling. As TCP/IP networks have become the communications technology of choice for mission-critical systems, security concerns—and the methods to address them—have increased. Add to this mix the rapid and complete adoption worldwide of the Internet, and its use for remote access to these systems, and security becomes one of the most important concerns when designing and implementing a distributed Ethernet network.

This chapter investigates various security concerns that threaten modern communications systems and the methods by which they can be addressed. Communications networks are considered the nerve system of any modern industrial site. An interruption is likely to cause loss of production and threaten human life as more and more end devices rely on communications with surrounding devices to properly monitor and control the site.

4.2 A threat defined

What is a security threat to the network for the purposes of this handbook? In a nutshell, it is an action or event (excluding natural hardware failure owing to use, faulty components or acts of nature) that could cause damage to the network, on either a physical or logical level. A security threat can be intentional or accidental and protection is needed for both.

For instance, a technician accidentally tripping on a cable and breaking it has the same effect as someone breaking in and unplugging the cable maliciously and intentionally. Therefore, when defining security it is necessary to protect against both. Similarly, it does not matter whether a virus on the network emanates from someone hacking in and uploading it or an attachment to an email; the potential damage is the same. While this is a broad definition and not strictly correct, in the scenario of securing a network, it is best to protect against all possible scenarios rather than underestimate and allow avenues of attack to go unguarded.

4.3 Physical security

Physical security is one of the first concerns that must be addressed for any mission-critical system, and a communications network is no different. It starts off at the most basic level, which is access control. Networking equipment such as routers and switches must be kept in controlled areas, accessible only to those who need to commission, troubleshoot or maintain the hardware.

Many networking devices have a serial console port for easy access to the unit's management system, which bypasses any network security in place—such as firewalls, etc. A user with malicious intent and a little knowledge of the hardware would be able to cause serious harm with this type of access. Even without the knowledge and hardware required for console access, an attacker could cause physical damage or interrupt power to the device, which in the best case would remove a layer of redundancy on the network. In the worst case this would cause a catastrophic communications failure to select devices. Establishing concrete site and company policies in relation to these systems is important; for example:

- Who is allowed to access critical communications hardware?
- Should the users be monitored by a local technician or engineer?
- How are device passwords shared?

Other policies may include changing passwords once a month (a controversial practice owing to the effort required to maintain up-to-date password lists in a large organisation), or after any major maintenance of the hardware. Policies need to be established for many of the topics discussed in this chapter. However, the full extent and level of policies adopted depends on the system and the company in question.

Either way, it is important that policies be enforced and not be allowed to be discarded. Complacency is one of the biggest threats to security and it is a good idea to re-evaluate all policies once a year to make sure they are being followed. More often than not security breaches can be traced back to a small mistake, such as someone not locking a door properly, or not disabling a remote connection to a device when he or she has completed data collection. For this reason it is critical that any third-party users be informed of policies they must abide by and that these policies are enforced.

With regard to wireless links, it is important to remember the following points:

- The wireless hardware is as vulnerable as the equivalent wired hardware and so needs to be protected by physical access control security wherever possible.
- Physical access to the radio signal itself now becomes a real threat. A user with a decent high-gain antenna and sniffer software can seriously affect the security of the site.

4.4 External devices

In any discussion on policies and third party users, an important question is: how are external devices handled? A USB flash drive is the easiest and most common way to transfer data physically, yet this type of external storage could be carrying a dangerous virus about to infect your network. A third party laptop may have some kind of sniffer software installed that captures any data travelling through the laptop's network interface, waiting to send this on to unsavoury individuals, whether or not the owner of the laptop is aware of it. There are a wide variety of third party devices that could possibly threaten the network, and we need to be aware of, and protect, against all possibilities.

Policies are particularly significant in such circumstances, and informing outside users (and employees) as to the correct way to handle external storage devices is important; with some viruses, plugging in the USB can be too late. External storage can be handled in different ways, such as having a computer with no connection to the rest of the network (but with an internet connection) running up-to-date antivirus software. Any files needed can be loaded onto this computer, scanned for viruses or malware, and then copied to the relevant machine on the secure network using an authorised clean storage device.

Some advanced firewall manufacturers include similar protection in their hardware, which protects against files incoming from the Internet, such as downloads or email attachments. These files are quarantined and a copy sent to an online cloud server, which checks the file for malware, and opens or runs the file in a protected environment to see what actions are needed. If anything out of the ordinary is discovered, a message is sent back to the firewall which deletes the file from quarantine before it and its attached devices can get to the network.

Even with good policies, one should assume that some sort of malware will find its way onto the network sooner or later. The Stuxnet virus, which shut down entire nuclear enrichment facilities in 2009/2010, was thought to have already infected a large portion of the world's computers at the time it activated; however, it did not activate on those systems (as it was coded to look for a specific target) and was not discovered for a long time.

The number of viruses on the Internet is immeasurable. Viruses range from harmless snippets of code that may do nothing, to system-killers that could cause expansive damage to a site. For this reason all computers attached to the network should be running anti-virus software. Updating the anti-virus solution is critical and must happen regularly, in some cases multiple times a day. The best way to achieve this is to get a solution that has a single server with direct internet access. The server generally resides in a DMZ (Demilitarised Zone), which is essentially a different subnetwork, separated from the rest of the network by a router and firewall. This machine updates its anti-virus definitions from an online server as they become available. The other machines on the network then update their anti-virus definitions from this machine—through a firewall which stops any other type of traffic—and thus are kept up to date yet do not require direct internet access.

4.5 Direct access devices

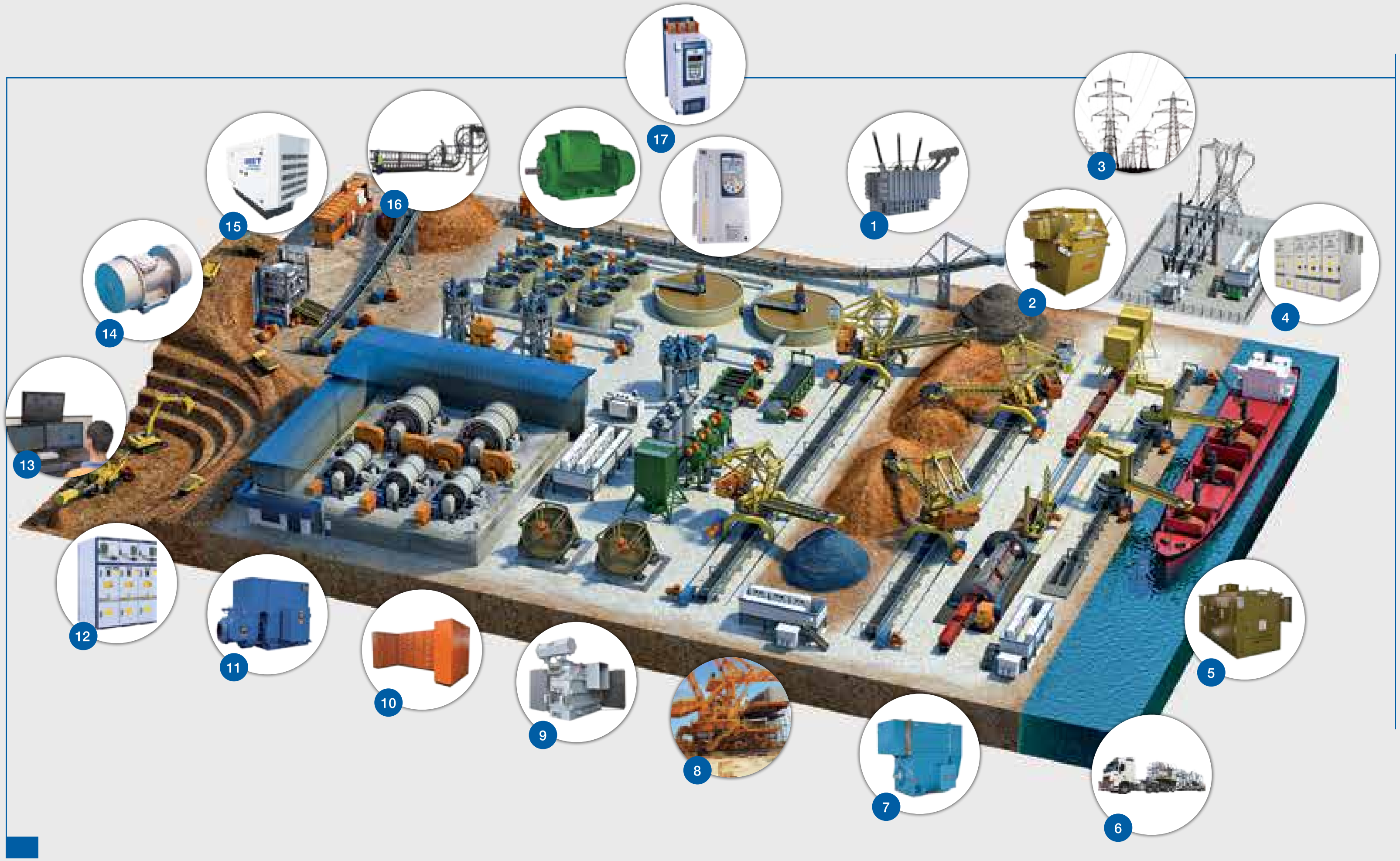
The next step is to protect against other devices that are able to connect directly to the network. Whilst physical access control and company policies are important, there are other, more automated methods that can be used to protect the network from unauthorised devices. Collectively known as AAA (Authentication, Authorisation and Accounting), this technology includes protocols such as RADIUS (Remote Authentication Dial-In User Service) and TACACS+ (Terminal Access Controller Access Control System). AAA collectively refers to three general functionalities:

- Authentication—which is checking people are who they say they are.
- Authorisation—which is checking what those people are allowed to do on the network.
- Accounting—which is keeping a record of who logged in, when they logged in and what they did while logged in.

RELIABLE ENERGY SOLUTION

for the entire mining sector.

1	Power Tranformers
2	Tiger RMU
3	Overhead Lines
4	MV Switchgear
5	Mini Substations
6	Mobile/Containerised Substations
7	MV Motors
8	E-Houses
9	Distribution Transformers
10	Motor Control Centres
11	MV Slipring Motor
12	MV Drive
13	Automation Control Room
14	Invicta Vibrator Motors
15	Diesel Generator
16	Electrical Construction
17	LV Motors & Drives



The Zest WEG Group, a subsidiary of leading Brazilian motor and controls manufacturer WEG, started out as a South African company and maintains its strong commitment to contributing to the development of the African region.

The Zest WEG Group has been servicing the mining sector for more than 35 years and by leveraging best practice engineering and manufacturing capabilities, the group is able to offer a range of standard off-the-shelf products as well as end-to-end energy solutions.

An in-depth understanding of the harsh conditions found within the mining sector and years of experience on the African continent, have ensured that the Zest WEG Group service offering is fit-for-purpose.

From single product installations to individually customised solutions, which are application specific, the latest technology is used to ensure optimum performance and reliability without compromising on energy efficiency.

WEG products are engineered to facilitate a safe and reliable mine and plant with operational stability and the highest possible production levels as an objective. Reduced maintenance and ease of serviceability assist in lowering the total cost of ownership for the mine.

Supporting customers is key and the Zest WEG Group operates a strategically situated network of branches and distributors across the continent. This ensures the highest levels of technical support as well as easy access to product and parts.

SCAN HERE
for more about
our offerings in the
mining sector

ZEST
WEG Group

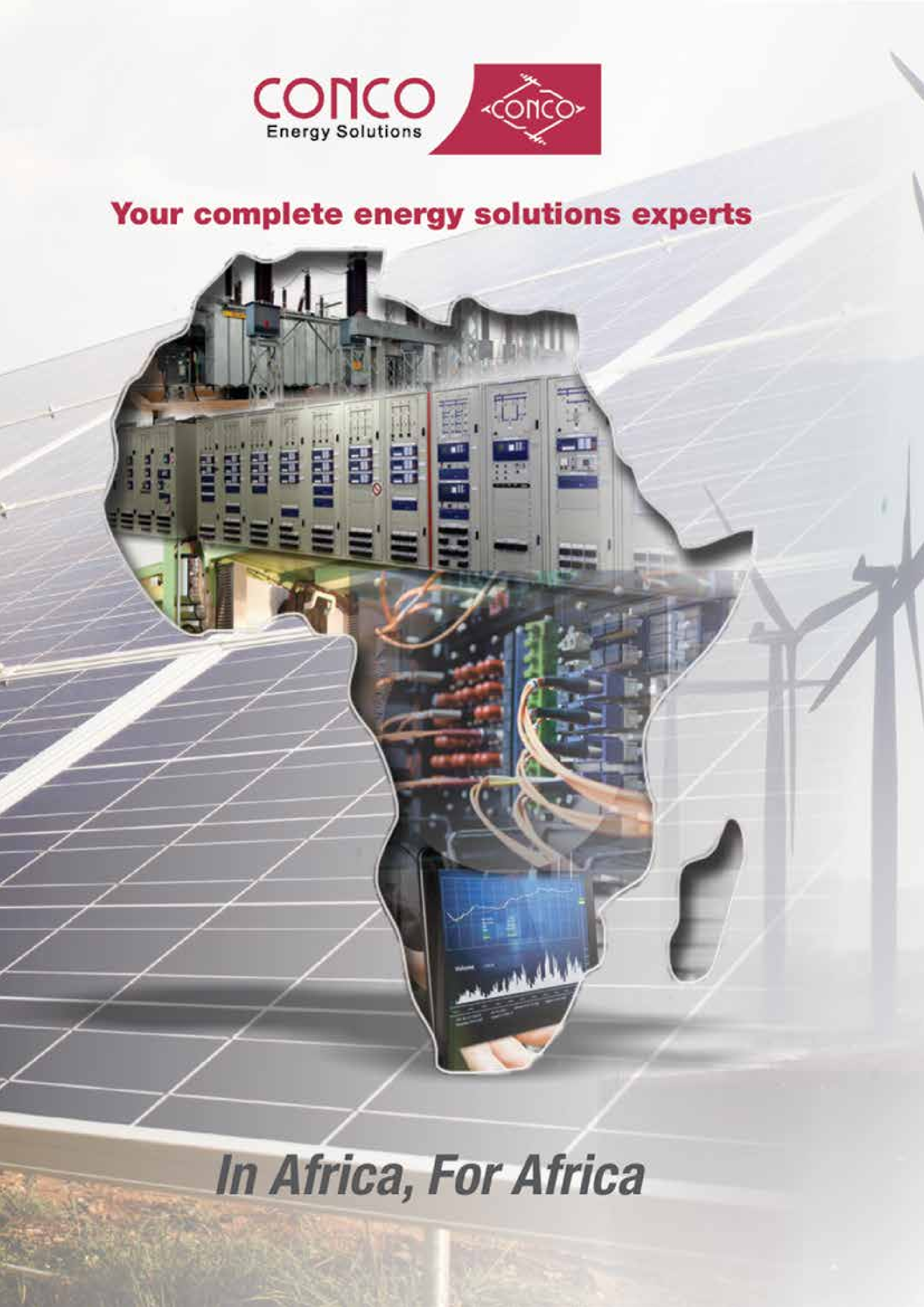
Tel: +27 11 723 6000

www.zestweg.com

For instance RADIUS and the IEEE 802.1x (Port based network access control) standard work together to authorise any laptop plugged into a network switch. Until the user responds with the correct username and password, no networking data is allowed to travel between that laptop and the rest of the network. This authentication process sends authorisation information back to the switch, which states what devices the laptop can communicate with on the network.

This type of functionality can be taken a step further with the use of a Secure Access Management (SAM) system. This system can take over much of the AAA functionality, and provide extra security, logging and access control features. Some SAM systems are able to monitor devices attached to the network, and send an alert if the configuration of the unit changes (as compared with a verified user created configuration) or if the firmware becomes out of date. These systems generally provide an authentication management system, which allows users to keep a single username/password combination to log onto the SAM system, which then controls the user's access to end devices on the network. This means that users are not able to access irrelevant end devices at all, whether intentional or accidental. A misconfiguration of an end device, can quickly and easily be identified by network engineers and rectified with minimal effort, as the SAM is able to keep track of any changes made. These systems not only protect the network against possible security threats, but can increase productivity and facilitate proper time management by removing or automating many of the steps required to maintain a network and the attached devices.

Access control technologies start to bridge the gap between physical and logical security. With physical security the concern is people accessing devices that make up the network and cables interconnecting the devices. With logical security the need is to secure the data itself. Ethernet and distributed networking offer a multitude of benefits to industrial communications systems; however, as they expand they become harder to secure, especially from a physical standpoint. At some point a secure network eventually connects to a less secure—or unsecure—network, such as an uplink to a corporate office for performance monitoring or an internet connection for remote access and control. For this reason a combination of policies, physical, and logical security on a mission-critical control or production network is needed.


CONCO
Energy Solutions
CONCO
Your complete energy solutions experts
In Africa, For Africa

Providers of complete turnkey solutions which include: design, engineering, manufacturing, factory testing, installation and commissioning.

CONCO Energy Solutions (CONCO ES) has expanded its footprint in the industry through the diversification of its product portfolio, which now includes containerised substations, low voltage products, smart grid solutions, batteries and chargers as well as telecommunication solutions.

From the most remote locations on the African continent to highly urbanised settings, CONCO ES is able to provide customisable solutions for each of its customers needs.

The product range includes:

PROTECTION SCHEMES

SUBSTATION AUTOMATION SOLUTIONS

INTELLIGENT HMI APPLICATIONS

TELECOMMUNICATION SOLUTIONS

FACILITY SCADA & POWER PLANT CONTROLLER FACILITIES

ADVANCED TELEPHONY SYSTEMS USING VoIP

CONTROL CENTRE APPLICATIONS

BATTERIES AND CHARGERS

MCC'S, LV AC & DC DISTRIBUTION BOARDS

YARD MARSHALLING KIOSKS, JUNCTION BOXES

AUDITS

SETTINGS AND GRADING STUDIES

COMMISSIONING

SOFTWARE DEVELOPMENT

SOLAR / PV SOLUTIONS

CONTAINERISED SUBSTATIONS

CONCO Energy Solutions

Tel: +27 11 805 1910 | Fax: +27 11 805 2827 |
Email: energy.solutions@concogrp.com | Website: www.concogrp.com

A CONCO Group company

CONCO

Chapter 5

Security at a Wireless/ Logical Level

Tim Craven

You can lock it up, but can’t stop it shouting! Many of our (very local) criminal types organise their networks better from the safety of jail. So can *YOUR* network!

5.1 Wireless meets wires

Another component in industrial networking that is not quite *physical* yet not quite logical is *wireless Etherne*t. The general recommendation for wireless on an industrial scale is to try to avoid it. Wireless is a great technology for use in a corporate or home environment, making it convenient for users to connect quickly and easily. However, in an industrial environment it becomes more of a hindrance than a help, for a variety of reasons.

Many of the reasons are technical, such as interference, latency, etc. However, it is the *security* aspects that are of particular interest in this chapter. Previous sections of this handbook cover protection of the physical network from unwanted users connecting to the network from an external location. Using wireless connections effectively negates much of this security. Wireless APs (Access Points) are accessible from anywhere, provided their signal is strong enough. This means that if a *wireless signal* leaks out of the site's property, someone with the right equipment and know-how, from outside the access control perimeter, can possibly gain access to the network. As this access is effectively local (i.e. it is the same as connecting via a cable to the network, rather than coming in via the internet), it bypasses some of the other logical security features, *like firewalls*.

Wireless access requires credentials, and there are other ways of making it more secure, such as hiding the SSID (Service Set Identifier) from being publically broadcast. However, even without full access, someone could potentially capture the data travelling through the air and break the encryption. All in all, the benefits and convenience of wireless do not outweigh the security and other *technological flaws* when considered for an industrial mission-critical communications system, and should be avoided unless absolutely necessary.

In some cases, using wired communications is not feasible, making the wireless route the only option. In these cases it is important that a *specialist* company be contracted to plan out and commission the wireless in the most secure way possible. This could include details such as using *directional* antennas rather than *omnidirectional* (directional pushes a much narrower beam of wireless signals, rather than broadcasting them everywhere). Hiding the SSID makes it harder for unwanted attackers to discover the wireless system, and implementing proper security such as RADIUS, rather than the older WEP (Wired Equivalency Protocol), makes it harder for anyone to crack the security and gain access to the network through the wireless link, should they be able to intercept the signal. The WEP protocol, when released, was cutting edge and enough to secure most wireless networks. Today, a WEP secured AP can be cracked in under a minute with software freely available online and a standard entry-level laptop. WPA and WPA2 with shared-key access are better, but not much as they still rely on point-to-point keys. A RADIUS server, bypassing hardware access entirely, is based soley on actual User authentication.

5.2 Outdated firmware

This leads into the next important topic, which is *correct maintenance* of the firmware of networking hardware to keep up to date. Industrial networking is a competitive market, and hardware manufacturers are constantly working on bug fixes and improvements to their devices. New protocols and ways of implementing various functions are constantly emerging and evolving, and potential security concerns in the devices addressed.

Firmware updates are the method by which a manufacturer rolls out these improvements to the customer, and are highly essential to keeping a network running optimally. In many ways this applies more strongly to security than other areas. As quickly as security experts find ways to block device exploits and improve their security, so malicious persons work to break through this security. New firmware releases for a device should be monitored in terms of the changes they introduce, and updates should be performed when deemed necessary. At the very least, firmware should be updated once a year as well as whenever a firmware release addresses any known security flaws.

5.3 (pa$$.w0rds)

Another extremely relevant point is the *changing of passwords* on devices. Again, strong company policies are needed. Although the trend is slowly starting to change as users become more aware of the need to properly secure the communications network, the majority of engineers and technicians are still guilty of one of the cardinal sins of industrial security: *leaving device passwords set to default*. The main reason for this is convenience; with many different contractors working on a single system made up of many different manufacturers' devices, it can become painful and inconvenient to keep track of every device's username and password.

Leaving devices on default authentication settings means that if someone can communicate to the device, they could probably gain management access to the device. Local security in devices, especially industrial grade mission-critical communication devices, is very strong, and should not go unutilised. Deciding that the inconvenience of tracking and maintaining a list of device authentications outweighs the extra security will come back to haunt you if someone manages to gain access to devices and ends up shutting down or damaging the communications system.

On the other hand, changing passwords too frequently can lead to poor password choices, such as reusing the same password many times but changing a number on the end of it to reflect the current month and year, or using simpler passwords that are easier to remember. This is another instance where a SAM solution is useful, as it can automatically handle the password changes of end devices and keep a database of all passwords. A SAM system does not become bored or fed-up with constantly changing passwords, and is not lazy about changing the passwords according to a schedule and a list of password complexity requirements.

5.4 Secure versus unsecure networks

In the early days of industrial communications, when Ethernet was still a *fledgling technology* and serial was the choice for mission-critical communications, resetting, checking or reconfiguring a device required being in physical proximity to the device. A small error could cost a few man hours, especially in travel time and production would drop or come to a complete stop. The small error could end up costing thousands of Rands or even more. The switch to Ethernet started to eliminate travel time since much of the work could be done from a central control room over the distributed network. The introduction of the Internet has taken this a step further, making it possible for users to connect from any location with an Internet connection.

With remote access, these small problems can be identified and addressed in minutes, and a user can obtain real time assistance from a head office located on another continent. In other cases, various different geographic sites can be linked by a private corporate network, allowing a central HQ to collect information and control everything from one location. This type of network is classified as WAN (Wide Area Network), and generally is not under the direct control of the same people in charge of a site's mission-critical secure LAN (Local Area Network).

The local network under direct control is considered to be a secure network, while any WAN this connects to is considered an unsecure network. It is clear why we consider the Internet an unsecure network, but it is important to realise that any corporate or similar network that connects to the secure network should be considered unsecured. Corporate networks have different requirements from mission-critical networks, such as high bandwidth and less strict firewalls to allow office workers to perform their jobs. Policies regarding virus checking may be more lax, and malware probably exists on the corporate network in some shape or form.

5.5 Firewalls

What exactly is a firewall? Originally (long before personal computing or networks) a firewall was a specially designed wall in a boat or building that was designed to prevent the spread of fires between different rooms and compartments. Skip forward a few decades and the word has been adopted to mean a *logical* or *physical* 'wall' that stops the spread of harmful data between different subnetworks. This can exist directly on a PC in the form of a software package such as Windows Firewall. In a mission-critical environment, however, it is generally a special hardware device running its own operating system and advanced protection software.

In its most basic form, a firewall is made to monitor traffic flows between different networks and allow or reject traffic based on a set of rules. Firewalls come in two major varieties: *stateful* or *stateless*.

Stateful firewalls not only monitor each packet travelling through the firewall, but also keep track of individual connections between devices in different networks.

Stateless firewalls simply inspect each packet as an individual entity.

A stateful firewall is better equipped to detect spoofing attacks, where a device intercepts a traffic stream and then sends its own, modified stream to the end device, pretending it was the original sender of the traffic. A stateful firewall picks this up because it monitors the actual connection and not just the packets individually.

A properly configured firewall is set with a general policy to deny any traffic passing through it, with specific rules created for each traffic stream that needs to be allowed. These rules should be as tightly specific as possible. For example, if two individual devices in a subnet need internet access to an online server, make the rule from only those two devices' IP addresses to the server's address on the Internet, rather than the entire subnet to the Internet. If they only talk over a single TCP port, set that port as another restriction. This makes it harder for anyone to find access through a loophole in the rules.

Modern firewall packages come with various other services, each of which needs to be licenced separately. This allows users to pick and choose the specific security options they want to integrate and use on their systems, and can include services such as built in anti-virus checkers, which check all incoming files, or traffic managers which can be configured with a set of rules determining which websites can be accessed from inside a secure network. Other services include monitoring of outgoing connections to see which devices are performing what services and actions. Whilst many of these features are meant more for corporate environments, they are becoming useful for mission-critical networks as well.

5.6 Virtual Private Networks

Another important feature of firewalls is their ability to function as a VPN server. A VPN, or Virtual Private Network, is exactly what the name implies. It is a virtual connection between another router or an end device, and the firewall or VPN server. This creates a virtual network connection that is private: it requires authentication (username and password) to connect to, and encrypts all traffic travelling across it. A VPN can be created over an unsecure network such as the Internet and sensitive traffic can be sent across it without fear of this traffic being intercepted. These days VPNs are popular for civilian use to obscure Internet activities and to bypass various geographic

restrictions (certain websites and features that are not available in some countries); however, from a security perspective the *authentication* and *encryption* features are of interest.

VPNs allow engineers to work from anywhere in the world where they have an Internet connection, and due to the way VPNs operate, it is as if the engineer's laptop is plugged directly into the network (when configured correctly). VPNs are becoming more prevalent on mission-critical systems, and are quite essential when dealing with international companies and hardware, as they allow the companies' support teams to commission, monitor and troubleshoot devices remotely, without the need for expensive dedicated links between areas. Actions that once cost thousands of Rands and required days of travel time as well as accommodation can now be undertaken in a matter of hours using a properly secured VPN solution. It is highly critical that VPN connections be properly configured and maintained. For mission-critical VPNs, IPSec (IP Security) is currently the best protocol from a security standpoint. While more complicated to set up than something like PPTP (Point-to-Point Tunnelling Protocol), IPSec is much more secure. Users should look at using security certificates rather than a username and password. Certificates are computer files that identify a device, and allow secure, encrypted communications only between correct certificate holders. VPNs are extremely convenient and should be utilised where they can save time and production hours. It must be remembered that they are effectively *opening tunnels into the network*, and if not configured correctly, pose a serious security threat.

Certificates are not only for securing VPN connections. They can be used to secure other types of communications, such as email. Email can be set up to digitally sign emails and encrypt the content of the email. A digital signature is proof that *'you are who you say you are'*, and that the email originated from your machine. Encryption means that the content of the email can be decrypted and read only by someone with the correct certificate on their side. *Note*, it is carefully stated that this set up only proves that the email has come from your machine, not necessarily from you. This, once again, highlights the need for correct company policies, such as not leaving PCs unlocked and email programs open and unattended. Logical security is important and very useful, however it protects only to a certain level. The human factor must always be considered and addressed.

5.7 An ounce of prevention

Two other system types that have gained popularity in recent years are an IPS (Intrusion Prevention System) and an IDS (Intrusion Detection System). These are similar systems and are sometimes confused. Add to this the fact that different vendors implement these tools in different ways and the line between them gets increasingly blurred. The difference is in the name: *Prevention* versus *Detection*. An IPS is very similar to a firewall in that it sits between two or more networks and monitors traffic passing between them. However, where a firewall inspects each packet and connection based on a series of access control rules, an IPS uses a set of rules to look for specific types of attacks and prevent those. For example, there is a type of attack known as a DDOS, or Distributed Denial Of Service attack, where a malware is first distributed to a number of online PCs. This malware allows a central controlling PC to initiate an attack where all of the 'slave' PCs send a flood of traffic to a certain address, effectively bottlenecking the target connection with junk information. This causes useful data to be slowed or stopped completely. A firewall, even if configured to drop each of these junk packets, still needs to spend time and processing power inspecting each of the packets to confirm it can be discarded. This means that the firewall itself is affected and slows down the inspection and transmission of useful traffic. An IPS could be configured to identify this type of attack and rather shut down each connection where possible, dumping all packets without inspecting each.

An IDS, on the other hand, is a more passive system. It sits on the side of a network rather than at an uplink, and monitors the network for various types of security red cards. For instance, if a set of devices uses 10% of its network capabilities for a year, and suddenly starts using 50%, this could be flagged as a possible issue. If a device is only using UDP traffic when operating normally and suddenly starts flooding the network with TCP multicast requests, this too could be flagged. All this monitoring is presented in a format that is easy to read and analyse and passed on to a network security engineer. This allows possible threats to be identified and addressed before they create a serious problem. Because IDSes look for symptoms rather than just causes, they can help identify problems that have not previously been encountered.

5.8 Monitoring

Now for the final point that is always critical and not only from a security standpoint: monitoring of a network and attached devices. Networks are becoming highly complex entities and they need to be properly maintained. The first step to properly maintaining a network and its attached security features is by having a full view of the network. Large security breaches are often preceded by smaller breaches as attackers test different components of the system. If the smaller breaches are identified early, they can be addressed—and the larger breach deferred or prevented completely. The IDS mentioned in 5.7 is one type of monitoring system; however, a host more are available and should be considered. On a simpler level users could implement a syslog collector—a central server that collects the system and event logs from devices on the network and consolidates them. Some of these systems can help flag concerning events, allowing an engineer to quickly identify possible problems.

There exists a protocol in Ethernet devices called SNMP (Simple Network Management Protocol), which is an open standard and should be supported by all Ethernet hardware, especially industrial grade hardware. The SNMP standard works off dictionaries of OIDs (Object Identifiers) known as MIBs (Management Information Bases). These OIDs are simply numerical codes which translate to a certain query, i.e., the OID 1.3.6.1.2.1.2.2.1.8 is for the query Interface Operational Status, or `ifOperStatus()`. Further codes appended to this identify which interface is being queried. This OID is then sent to a switch, for instance, that responds with an OID stating whether the interface is up or down. A central NMS (Network Management System) receives all the responses from different queries to devices around the network. These are consolidated and presented to a network engineer, normally in a quick to understand visual format. The engineer is able to assess the status of the entire network, and identify problematic areas and devices instantly. While these systems are more important from an operational standpoint, they are another example of a monitoring system that should always be implemented.

Since the OIDs and MIBs are part of an Open Standard, they are unfortunately sometimes carelessly put together by manufacturers, meaning that a LARGE database of such identifiers is necessary. Manufacturers do not always publish these (for various reasons) and, as a result, swapping out a network component for a different one with exactly the same functionality, but from a different manufacturer, may degrade the SNMP reports.

On a wider front this is true for most mission-critical networks, which start off well planned and documented, but later start to suffer from small changes here and there that are not documented (people forget to document the changes, or think they are so insignificant that they need not to be documented). After a period of time, this lack of updating of documents and maintenance of the network means that what remains is an unsecure mess of a network that has *vulnerabilities* and *flaws* throughout.

Remember that one single security breach is all it takes …

It is clear that security on a modern communications network is extremely important and cannot be underestimated. Industrial Ethernet brings a host of benefits and improvements; however, if not secured properly it is more hindrance than help. In the best case scenario, unauthorised individuals will be in the network and able to view confidential data; in the worst case, individuals could cause damage to company buildings and themselves. Securing a network properly leads to increased peace of mind whilst utilising the benefits that Ethernet networks provide.

Connecting Devices around the world

Manage field equipment online!

With Netbiter Remote Management you can view and control field equipment remotely — Manage your sites and analyze data online.

Savings are immediate as a Netbiter gateway is about the same cost as a single service visit.

Save time, save money, save resources

netbiter®

Get connected!

Anybus gateways allow you to quickly and easily connect any machine to any industrial network. You configure the communication in the Anybus Configuration Manager software. No programming is necessary.

Connect. Configure. Done!

Anybus®

PC Interfaces, Gateways & Bridges

With up to four CAN and CAN FD channels, HMS offers the best CAN interface for your application. Customized cards are available with APIs for CAN, CANopen and SAE J1939 and support our analysis and configuration tools.

Furthermore, IXXAT repeaters/bridges increase system reliability, reduce wiring costs and enable connection of CAN systems via e.g. Bluetooth or Ethernet.

IXXAT®

HMS Connecting Devices™

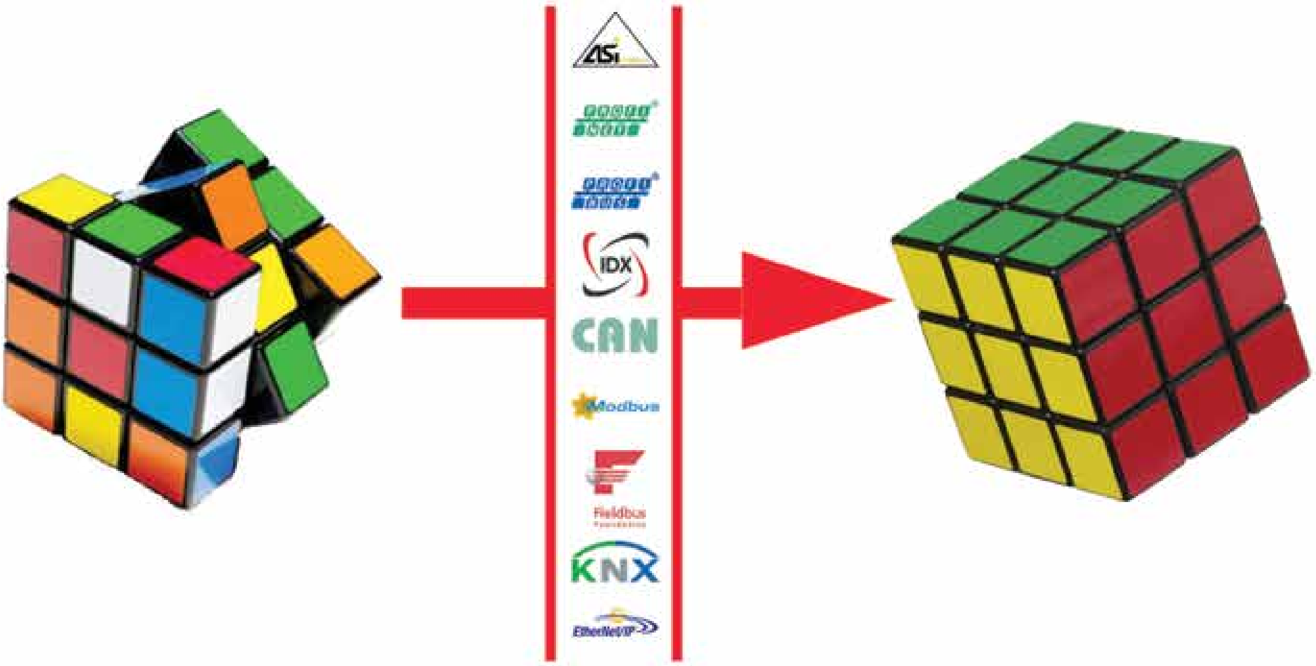
www.hms-networks.com

South Africa distributors
Industrial data Xchange
Sales@idx.co.za
+27 11 548 9960

IDX

Industrial Data Xchange

Data connectivity that works



Let IDX solve your data communication needs...

PROTOCOL CONVERTERS AND COUPLERS

PROFIBUS/PROFINET CALLOUTS AND AUDITS

CUSTOM LEGACY SYSTEM SOLUTIONS AND GATEWAYS

INTERNATIONALLY CERTIFIED TRAINING

REMOTE ACCESS AND ASSET MONITORING SOLUTIONS - IIOT

ADVANCED TROUBLESHOOTING AND TEST TOOLS

www.idx.co.za

Chapter 6
Greenfield

Tim Craven

In 1993, South Africa was introduced to GSM – cellphones! The wild, ridiculous five-year penetration targets of all the fancy business plans were breached *in three months*. We did not have an olde analogue system of phones to migrate from: this was a Greenfield, and we Boldly went Digital.

6.1 A fresh start

When undertaking a major stockroom cleanup, one of the easiest methods is to clear everything from the room so that there is a clean, empty room to work with. Stock can be packed into the room one piece at a time in a neat, organised and controlled fashion. This is much simpler than trying to work around the boxes while cleaning and organising the stockroom, as the existing clutter gets in the way, confuses everyone involved and yet still has to be catered for in the final stock count. The same is true for any type of project, from the simple restructuring of a single room or process, to creating an entirely new plant for a specific application. Working on a project that has no constraints imposed by prior work (commonly called a 'Greenfield' project) is always easier than the alternative, where a system already exists that has to be expanded or altered to fit the customer's requirements.

This applies equally to a mission-critical network implementation and in the following paragraphs we take a look at the differences between implementing a Greenfield network and changing or expanding an existing network.

Ethernet networks work on the multi-layered OSI reference model, and we will use an analogous approach when discussing the implementation of a network, by starting at the physical layers (cabling, hardware selection, etc) and moving up through the data link layer (logical topology, redundancy, etc) towards the network, transport and application layers (IP structuring, routing, etc).

6.2 Hardware

When expanding an existing network, the existing hardware is the first point to be considered when selecting new hardware. The biggest restrictions come from any proprietary features on the existing hardware, and whether the new (expansion) hardware needs to comply with these proprietary features. If this is the case, it would limit hardware selection as only manufacturers that comply could be used. This is known as becoming vendor locked, where one is effectively locked into using a single manufacturer's products and unable to consider other options.

Solving this issue can be unnecessarily costly, as most existing networking hardware will need to be replaced and the expansion or upgrade will also need to be catered for. This could mean that thousands of Rands of networking equipment may be mothballed when it could have provided years of operation. Alternatively, one could look at downgrading existing functionality, so that instead of the proprietary features, one uses openly available standards. This could create unplanned-for issues, as proprietary features are often more effective than open standards (mostly owing to the fact that manufacturers do not need to cater for integration with other manufacturers and can focus on the feature itself). For instance, many proprietary redundancy protocols are much better (provide quicker recovery times) than their open standard counterparts. Adding new devices means they either need to support the existing (proprietary) protocol (which means one is vendor locked) or one could move to an open standard such as RSTP, but with a drop in performance. In some cases the proprietary protocol used will have a level of backwards compatibility with open standards (although in these cases a loss of performance on the non-proprietary devices can be expected). This decision would be preceded by an analysis of the system and its requirements, with the final decision depending on the outcome of the analysis. This would itself add extra time and cost to the project.

Looking at the same point from a Greenfield project simplifies matters greatly (as it will in most cases with a Greenfield project). Instead of worrying about existing hardware and integration with the same, the hardware selected could be based simply on what will suit project requirements best (whilst being mindful of budget). At all times it is necessary to keep in mind future expansion or additions, and cater for them where possible. It is at this stage that active steps should be taken to avoid becoming vendor locked. Wherever possible, use open standards rather than proprietary, as long as the open standards provide the performance required by the project specifications.

6.3 Cabling

Once the hardware has been selected, the physical connections between this hardware need to be considered. Cabling is one of the few components that can be easier to cater for on an existing network than on a Greenfield project (under certain conditions). The reason is that in an existing system, a certain level of cabling will already exist. In most industrial cases, when pulling cabling for communications (particularly when dealing with fibre optic cables) multiple extra cables will often be pulled. In the case of fibre optics for backbones, rather than simply pulling through a single fibre optic pair, a multicore cable will be pulled. These cables consist of multiple inner fibre optic cores protected by an insulating, hardened cover. This means that existing cabling runs will have unused fibre optic pairs which can be used for new links, meaning that time and effort are saved by not needing to run a separate cable.

However, in cases where existing cabling cannot be used for backbone links and new cabling must be installed, once again a Greenfield project has a huge benefit over an existing site; i.e. the space to dig up trenches and run the required cables. Existing sites start to become cluttered with above ground elements that interfere with cable runs, and these elements need to be dealt with. This can mean either skirting the obstacle (which adds length to the required trenches and cables, increasing cost and labour requirements) or moving the obstacle to trench underneath it (which can be costly if the obstacle cannot easily be moved by hand). If planned correctly, a Greenfield project can have all relevant cabling trenched and buried, before extra clutter is moved in. Either way, when dealing with cabling it is important to try and cater for future expansion, e.g. by using multicore cable that will provide additional fibre pairs when required, or by burying cables in covered trenches that can be (relatively) easily accessed in the future.

6.4 Logical topology and redundancy

Ethernet networks are highly customisable, and there are usually many ways a network can be laid out. The physical topology is restricted by physical factors such as building locations, cable runs, etc, but the logical topology of a network is more defined by how we want data to travel around the network. For instance, we may have a physical mesh of devices and interconnecting cables; however, from a logical standpoint, many of these links may be disabled for redundancy. Some redundancy protocols will even split data across multiple redundant links (such as MSTP), with different VLANs using different logical paths to traverse the network. So our logical topology, while ultimately reliant on the physical topology, is best defined by the location of various end devices and the different flows of traffic relating to different processes/functions.

When expanding an existing network, the choice of logical topology will depend heavily on how the current network is laid out logically. For instance, if there is an existing redundant backbone ring, new devices will need to be added as part of the ring (if they are also to be redundant), or added on the edge of the ring, in which case they are not redundant on the existing backbone. One could look then at having a different section of the network run a different redundancy protocol; however, this gets complicated to implement and maintain, and requires research beforehand to ensure that the different redundancy mechanisms are compatible with one another (or at least will not interfere with one another). Running different redundancy mechanisms on a single network is not recommended, as it is too easy to have a situation where they clash and data is lost erroneously.

Once again, it becomes important to avoid becoming vendor locked (especially in the case of redundancy, which can be heavily affected by other protocols on the network) and open standard redundancy protocols are recommended (as long as they provide the required performance).

A Greenfield project once again provides a greater level of choice, especially when the cable layouts can be planned according to the redundancy to be implemented, rather than existing cables defining the redundancy options. Being able to define cable layout and redundancy implementation before any physical work takes place allows more freedom and allows the data flow requirements and the ultimate purpose of the network to define the physical and redundancy components, rather than the other way around. Wherever possible, it is preferable to let the requirements define the network to provide the best possible solution for the application in question.

6.5 VLANs

As a natural progression from the design of the logical network and associated redundancy we move on to consider traffic segregation. TCP/IP networks use different types of traffic, broadly split into unicast, multicast and broadcast traffic. Without going too much into the technicalities, which are not the focus of this chapter, unicast traffic will be between two individual devices, whilst multicast and broadcast traffic will often be sent to many or all devices on the subnet, whether that traffic is relevant or not. If the end device needs the data then it is not an issue. However, even if the end device does not need the data and simply discards it, it still needs to inspect each data packet, which uses processing power and time. While the time and processing power may seem negligible, if this happens hundreds or thousands of times a second, the device can become flooded with non-relevant data, whilst relevant data sits in the incoming or outgoing queue.

The solution is to segregate the traffic to ensure that the end devices never receive the irrelevant packets and thus do not waste resources inspecting them. The way to do this is by using VLANs, specifically Layer 1 or port based VLANs (VLANs refer specifically to Layer 1 VLANs, unless otherwise specified). There is a common misconception that IP subnetting will also segregate the traffic but this is not completely true. With IP subnetting, a device in a different subnet from that of the sender will not be able to communicate with the sender; however, it will receive multicasts and broadcasts sent by the sender. The packets will be inspected and discarded still using up end device resources and therefore not providing proper traffic segregation. With Layer 1 VLANs, on the other hand, the switches themselves will simply not transmit the data to non-relevant end devices, meaning no resources are utilised by the end devices inspecting and discarding these packets.

Designing VLANs is another step where there is not much difference between Greenfield and existing networks. VLANs are configured on each switch on the network (plus routers if inter-VLAN routing is required) and each physical port requires a small VLAN configuration as well. Expanding the network means that each new switch/port being utilised will require a small amount of configuration. Existing switches will already be configured (although some tweaking may be required). On a Greenfield project all switches will need to be configured, so more time is required. This means that a more important point is that the initial VLAN design must be properly planned, and should cater for long term future expansion as much as possible.

It may take longer to configure VLANs on a Greenfield network (compared to an existing network), but it can be argued that the configuration time has already been done for the existing network, and so the total configuration time required will be pretty much equal. Changing existing VLANs can be more complicated, and will incur downtime on a live network. For this reason, time should be spent making sure the original VLAN design is optimal for traffic on the network, and thus reconfiguration will be kept to a minimum.

6.6 IP address structures

Whilst VLANs are not directly related to IP addresses in any way (the switches on which VLANs are implemented are generally Layer 2 devices, and so are not IP address aware) the two share a connection, especially when considering routing on the network. The recommended practice is to assign a different IP range to each VLAN on the network. While this is not required from a functional point of view, it does offer two major advantages:

- It is possible to route data between the VLANs where required, and
- It is easier to identify which VLAN a device belongs to if it sits in a different IP range.

IP addressing structures can be of the most painful points to address

when upgrading an existing network, and one of the most satisfying when working on a Greenfield project. From a Greenfield perspective, one needs a list of devices (or expected devices) on the network, how these devices will be segregated (VLAN assignments), and any restrictions on what ranges can be utilised (sometime ranges may be assigned on a site-by-site basis, for instance). Once all this information has been gathered it becomes a simple matter of subnetting IP ranges to break them down into smaller groups, which should then cater for the number of devices in each VLAN. With the free tools available online it becomes a straightforward task. IPs are then assigned to the actual devices and all the assignments are documented. It is necessary to add a bunch of devices to an existing network, things can become more complicated, depending on how well the original design catered for expansion. In a best case scenario no extra VLANs (and thus IP ranges) are needed and all existing ranges have open IP addresses for each new device to be added. In many cases however, the existing ranges are too small for the expansion required. This is where things can become painful. Sometimes it will suffice to make small changes to existing subnet masks to include other devices, as long as this does not overlap with existing ranges. In other cases, redesign of the IP structuring for the network, and reassignment of all existing IP addresses to cater for the new design, will be required. Furthermore this will involve large amounts of reconfiguration of existing devices on the network to accommodate the new structure. Finally, it may be necessary to reconfigure some software to work with the new ranges—such as SCADA and other monitoring or control software—as well as firewall and routing configurations.

6.7 Routing

Routing is directly and intrinsically related to the IP structuring on the network. On a Greenfield project designing routing is similar to designing the IP structuring for the network. The same information is required as for designing IP structures (list of devices and segregation framework), as well as a list of what devices need to intercommunicate with each other. From here, it becomes simple matter of planning out the routing table and configuring the routing rules on all the routers. In some cases it may also be possible to clean up the routing tables by super-netting multiple IP ranges together, depending on the network design.

In an existing network that is being expanded, routing is similar to IP structure design, in that it can prove to be extremely easy or monumentally hard. In some cases it may be a simple case of adding a new route or two to the routing table to cater for the new devices on the network, whilst in others a complete redesign of the network is required depending on the complexity of the routing system. Furthermore, a redesign will require reconfiguration of all end devices that require routing, as their gateway IP address will need to be changed to match the new network design. This again shows the importance of correctly planning not only for the original network in the starting phases, but also for future upgrades and expansion.

6.8 Firewalls

A router is set up to move data between different subnets. However, it only gives control over what data will be routed where, not control over what data is allowed to be routed where. For this, a firewall is required on the network. Most industrial routers these days include a firewall, but it is important to make the distinction between the routing component of the device and the firewall.

Setting up a firewall on a mission-critical site is closely tied in with the router setup, and since both are performed together, the firewall setup follows a similar pattern to the routing setup when viewed from a Greenfield/non-Greenfield perspective. On an expanding network, if we have originally catered correctly for the expansion, firewall setup will involve either setting up a few new rules on an existing firewall, or implementing a completely new firewall if none exists. Both options are pretty straightforward as long as all the requirements are clearly laid out. If the entire network needs to be redesigned, the firewall will need to be reconfigured entirely and this could be a time-drain and can cause downtime on the mission-critical network.

On a Greenfield project, the firewall will also need to be configured from scratch. As long as the previous steps and all firewall requirements have been laid out in a clear fashion, the actual configuration should not take too long a period of time. Because firewalls (and routing) are so closely tied in with the IP structure, the amount of work that needs to be invested in the router setup is largely dependent on the complexity and changes on the IP structure of the network. However, as with most other points, the ‘clean slate’ provided by a Greenfield project will generally provide time and production savings, and also give the option of designing the network to fit the application, rather than designing the application to fit the network.

6.9 Less stress

Greenfield projects are almost always easier to work on than existing networks, and if the benefits they provide are understood and utilised, commissioning a stable and secure network can be created from the get go. If all steps are properly documented along the way, maintaining this network should also prove to be straightforward. On the other hand, constantly making unplanned changes and expansions on an existing network can lead to a frustrating mess that constantly causes problems, especially if the changes are not properly documented. One of the most important considerations of a Greenfield project, as has been stressed many times, is to plan properly for future expansion and upgrades. This will lead to time and production savings, as well as greatly decreased frustration and stress for those who are in charge of maintaining the network and attached devices.

Chapter 7
EtherNet on Steroids

As mentioned, Ethernet, or EtherNet for emphasis, has certainly come to the party. Originally driven by Quality-of-Service (QoS) needs of voice and video, this is a painful illustration of using something for which it is **NOT** designed, but JUST BECAUSE IT IS THERE.

Ethernet, especially in conjunction with TCP/IP was designed, come hell or high water, to get the packet across.

For Voice or Video, you simply couldn't care less about an errant packet that got stuck at Gillooly's Interchange for five extra minutes. If it ever does arrive, it simply gets thrown away.

In classical industrial networking, time-critical information is crucial to the design of control algorithms. So we take a brief look at how to re-design ethernet for timing, as well as a replacement for standard WiFi based Ethernet. Finally we look at some interesting statistics on market share, provided by HMS via IDX, locally [7].

7.1 Time sensitive networking

The major shortcoming of Ethernet often boils down to a lack of real-time capability [4]. The IEEE task group Time Sensitive Networking (TSN) intends to change this. The intention is for real-time to become an integral part of the Ethernet standard, rather than a non-standard-compliant add-on. But is this really a feasible approach?

Ethernet is specified and continually developed in the working groups of the IEEE 802 [ ] project. Some years ago, the task group was established, seeking to make Ethernet usable for time-critical applications. However, IEEE 802 does not offer a complete solution, but instead provides standards for the data transfer layer that require integration into an application concept.

The original plan envisioned the projects of the TSN task group be completed by the end of 2016. However, in addition to the six originally proposed extensions to the Ethernet standard, further projects are under discussion.

For example, the group is developing a procedure that involves forwarding of time-critical messages only to the immediate neighbour during each cycle (IEEE 802.1Qch). This is advantageous if the cascading depth is low. The approach can help integrate wireless devices or other components with latency that is difficult to determine, and it is more robust than time control.

An additional aspect discussed by the experts is how to limit the effects of nodes that act incorrectly. To this end, the incoming side (ingress) of the nodes must monitor the partners (IEEE 802.1Qci). Ethernet itself is also subject to changes: particularly noteworthy is the new two-wire transmission technology (100 Mbps: IEEE P802.3bw, 1 Gbps: IEEE P802.3bp), for which unshielded cables can be used. The main drivers for new projects here are car manufacturers. If the forecasts of half a billion Ethernet ports installed in vehicles by 2021/2022 come true, this will have a lasting effect on other markets – not just direct suppliers.

7.1.1 Does TSN really help automation companies?

The procedures defined in TSN are not suitable for efficient distribution or for gathering small data quantities. Compared with an EtherCAT solution, for a typical data volume below 10 bytes per device, TSN would result in a tenfold increase in protocol overhead, even in a best-case scenario. The TSN approach, with its significantly poorer efficiency, is therefore not really suitable for conventional I/O or drive applications.

However, it can have advantages in heterogeneous environments with data quantities of more than 100 bytes per transfer. Such an environment can be found, for example, in the networking of controllers, in robot cells, or in the integration of camera systems into automation systems.

Since standards are unable to take into account individual cases and special requirements, some functions may not be particularly suitable for specific automation applications. For example, although IEEE 802.1Qca includes a provision for the distribution of topological information, this protocol contains so much functionality that there is significant transfer and memory overhead. The lack of scalability limits the usability for simple nodes, as the important information regarding the topology could be distributed with significantly lower overheads.

The degrees of freedom for synchronisation were limited in IEEE 802.1AS. Yet there is no restriction on the forwarding delay of the individual nodes, which may have a very negative impact on the clock control loop. The delayed adjustment of the clock may cause increased inaccuracy in individual nodes.

Synchronised sending in a reserved channel can eliminate the impact of other protocols on time sensitive streams, the real-time traffic would have to be scheduled, however, to avoid additional delays in the cyclic data exchange. This is a complex optimisation task, and thus it is not feasible to find the optimal schedule even with a limited amount of data streams in a reasonable time.

7.1.2 Integration in applications

The focus of IEEE 802 is data communication. An application layer is required to integrate communication into a control environment. At cell level, however, proprietary application protocols of the control system manufacturers dominate. There are several standards at the I/O level with similar structures which may be used. The limited addressing volume may present challenges, but the model itself can be applied. CANopen-based protocols with certain extensions would be a candidate as an intermediate level. This would facilitate the transition to the I/O protocol world and would therefore be efficient, both in the cyclic and acyclic ranges.

Moreover, there are new efforts underway concerning the OPC UA standard, focusing on enhanced real-time capabilities, and also allowing for support of TSN. OPC UA offers a powerful infrastructure which can be extended with real-time capabilities. Thus it enables a unique approach to define a common real-time protocol suite at higher levels of automation.

7.1.3 TSN – a success story?

Industrial communication has been a key driver for progress in automation technology. However, it has also produced a number of 'ghost towns', such as the Manufacturing Automation Protocol (MAP), or the attempt to network with .NET components. All failed approaches were characterised by unnecessarily high complexity of protocols with relatively low efficiency and a lack of focus on the needs of automation vendors. TSN also has a tendency towards more complex procedures. Nevertheless, there are quite a few companies that have a strong interest in standardised real-time Ethernet at cell level. However, at the field level working solutions are already available. The willingness to establish an additional fieldbus will likely be limited. Still, TSN could well play an important role higher up in the automation pyramid.

It makes sense therefore to grapple with TSN and associated activities, even if many key questions remain. Automation companies and automation providers should build on the achievements at the I/O level to-date. If TSN is to become a successful model for automation in a heterogeneous cell infrastructure, there is a need to agree on an application protocol and to select appropriate real-time mechanisms from the TSN pool.

7.1.4 Technology in detail: The real-time standards

To date, the TSN group has initiated six standardisation projects:

Improved synchronisation behaviour (IEEE 802.1ASbt)
The previous version of IEEE 802.1AS had already specified a synchronisation protocol for the timing of distributed clocks, based on the IEEE 1588 standard. It had promoted the integration into a standard Ethernet environment. However, compatibility with other 1588 Ethernet profiles was lost. The new version will incorporate the accepted features of one-step transparent clocks. The main area for improvement right now is the response to error situations, such as failure of a communications line or a master. The new version should also be able to deal with different time domains in a device.

Frame preemption (IEEE 802.1Qbu)
A major problem for deterministic transfer of time-critical messages is legacy traffic on the same network segment, where an individual frame can be more than 1 500 bytes long. This can result in delays of up to 125 µs per node cycle. The problem can be addressed by means of a frame interruption mechanism (specified within the IEEE working groups in Ethernet project P802.3br). Ultimately, this mechanism will require not only new network components, but also new Ethernet ICs in the end systems.

Enhancements for scheduled traffic (IEEE 802.1Qbv)
The time control of send operations plays a key role in TSN. Just like in 'real life', there may be traffic jams on information highways and, even with high-priority real-time data and preemption, there may still be some variation in transmission times. Since the time-sensitive streams are transmitted cyclically, largely undisturbed communication can be realised by blocking less time-critical data just before cyclic communication. The procedure is comparable to traffic light control.

Path control and reservation (IEEE 802.1Qca)
In order to get from A to B as quickly as possible, you need a map and a route planner. Just like in everyday life, a network requires one to capture the way in which

components are arranged and determine how to select the communication routes in the most efficient manner. The protocol can be based on the 'Intermediate System to Intermediate System' (IS-IS) concept, which is also used by routers. It involves gathering and distribution of topology information. After several iterations, all nodes have all the topology information from the entire network. If there are several possible routes that lead to the destination, the procedure can be used to find the shortest one. It can also be used to identify redundant routes. This project was initiated outside TSN using built-in TSN-features.

Seamless redundancy (IEEE 802.1CB)

Although the IEC already provides specified protocols for seamless redundancy such as High-Availability, Seamless Redundancy (HSR), or the Parallel Redundancy Protocol (PRP), it requires that the complete data exchange between stations be designed for redundancy. This can cause problems, because the order of the messages is not respected in the event of a fault. In addition, troubleshooting is quite complex. For IEEE 802.1, it was therefore decided to explicitly apply seamless redundancy only to individual critical data streams. This makes it possible to reduce the protocol overhead, and critical points are easier to identify.

Stream bandwidth reservation (IEEE 802.1Qcc)

A major problem with Ethernet is found with overload situations, such as when data are received through two channels and forwarded over a single output. A large memory is also sub-optimal, since the delay increases with the number of bytes stored. This delay (best effort) cannot be controlled by increasing the response time in automation technology. If real-time data streams have high priority, there is the risk that the rest of the communication be delayed forever. For this reason, the required stream bandwidth is determined and reserved. The reservation protocol allows a real-time load of up to 80% of the bandwidth. It is an extension of the existing reservation protocol. It has become clear, though, that it will not be feasible to meet all the extended requirements of TSN by merely extending the existing reservation protocol. This means that it will still be necessary to find additional mechanisms for implementing real-time channels in the future.

The two new projects .1Qch and .1Qci have been initiated; they are, however, not yet clearly defined in terms of detailed objectives.

7.2 Trusted wireless 2.0

The automation industry mostly uses wireless technologies that can be used worldwide. Due to national frequency regulations, only a few frequency bands meet this requirement. The so-called ISM (Industrial Scientific Medical) bands can be used without a license, but only the 2,4 GHz ISM frequency band is available worldwide. This is why most wireless technologies in automation technology use this band.

Moreover, with a bandwidth of 83 MHz, this frequency band enables the parallel operation of several wireless systems—even if different wireless technologies are used. Other bands, such as the 433 MHz band (bandwidth 1 MHz) or the 868 MHz band (EU only, bandwidth 2 MHz) offer a smaller bandwidth. Thus, there are fewer ways of several wireless systems coexisting in these lower-frequency bands. However, coexistence is one of the vital properties of wireless technologies as the number of wireless systems constantly increases while available resources (wireless spectrum) are limited.

However, the attenuation of electromagnetic waves is less in the lower frequency bands. Therefore, the coverage range decreases as the frequency increases.

Trusted Wireless 2.0 [5] is a wireless technology that has been developed especially for the industrial use. It is particularly suitable for the transmission of sensor/actuator information without a cable infrastructure or for the transmission of small or medium data amounts—even over large distances from a few hundred metres to several kilometres. A few features of Trusted Wireless 2.0 are:

- Rugged communication thanks to FHSS.
- Automatic and manual coexistence mechanisms.
- Secure communication thanks to 128 bit AES encryption and authentication.
- Long range thanks to high receiver sensitivity and variable data transmission rates.

7.2.1 Frequency hopping spread spectrum

Every user wants to use a reliable and rugged communication connection for his/her application. The terms 'reliable' and 'rugged', however, are perceived rather subjectively. Characteristics such a reliability, latency, determinism, data throughput, etc play an important role for the user. Generally, this is called 'reliable'.

It is important to know and be able to classify the real application requirements. Available wireless technologies have different key aspects and performances and have to be selected according to the application requirements. It is also vital to know which factors impede the 'reliability' of a wireless path and how the different wireless technologies deal with these problems.

There are two major factors that can influence a wireless connection: Firstly, the disturbance of the wireless signal by other electromagnetic waves, triggered by other wireless systems or unwanted emissions of other electric devices (EMC disturbances); secondly, 'fading', which is caused by the free space attenuation and especially by reflections.

7.2.2 EMI

In the 2.4 GHz band, wireless systems benefit from the fact that EMC disturbances caused by general industrial applications do not reach this high frequency range. Frequency converters, ballasts and other EMC-producing devices, which usually are a problem, do not disturb the GHz band. Their energy transmissions play a role for frequencies in the kHz and MHz area.

With Direct Sequence Spread Spectrum (DSSS), the useful signal to be transmitted passes through a spreading code generator, which transforms the narrow-band interference signal with high amplitude into a wide-band signal with lower amplitude. Together with the useful signal, the incoming narrow band interference signal with high amplitude passes the same spreading code generator in the receiver. This way, the wide-band useful signal with low amplitude is converted again into a narrow band signal with a high amplitude and simultaneously, the interference signal is transformed into a wideband noise. One benefit of this procedure is possible transmission with a very high data rate. The disadvantage is the fixed transmission frequency as well as the fact that this procedure is only useful up to a certain interference signal level. If this level is exceeded, the receiver cannot make a distinction between the useful signal and the interference signal.

With the FHSS, many different individual frequencies are hopped through in a pseudo-random pattern. This way, an interference signal only blocks one or a few neighboured individual frequencies—no matter how high the level. The transmission can be implemented without interferences using the remaining frequencies. If disturbances become worse, only the data throughput is reduced in the FHSS system. In the DSSS system, however, transmission might be blocked completely.

In the 2,4 GHz band, Trusted Wireless 2.0 uses a Frequency Hopping Spread Spectrum (FHSS) with up to 440 possible individual frequencies, with the devices using a selection of up to 127 channels. The number of frequencies used within the pseudo-random hopping pattern depends on further settings and mechanisms such as the exclusion of certain frequency ranges (black-listing) for the coexistence management, or the use of several frequency groups (RF bands) to optimise the parallel operation.

7.2.3 Fading

Both DSSS and FHSS combat Fading by the simple expedient that the path length between the direct signal and the reflected signal are different at different frequencies.

FHSS shows this fading margin more strongly as it is unlikely that a catastrophic loss of communication will be repeated at the next random frequency. DSSS has a bundle of frequencies, hence a softer, less catastrophic fading. Thus, although DSSS technically sounds better at handling fading, in practice, a decent FHSS kills DSSS.

7.3 Industrial network market share

Industrial Ethernet is growing faster than before [7] and now accounts for 38% of the market. EtherNet/IP is in first place within industrial Ethernet globally, followed by PROFINET. However, classic fieldbuses are still dominating the fragmented world of industrial networks with 58% and PROFIBUS is still the most widely used industrial network. Furthermore, the Internet of Things is driving wireless technologies, which are now on the chart for the first time as shown in *Figure 7.1*.

Some of the trends highlighted by HMS (in *Figure 7.1*) are:

Fieldbus is still growing: Looking at new installed nodes within factory automation globally, fieldbuses are still the most widely used type of networks with 58% of the market. Fieldbuses are still growing by approximately 7% per year as users ask for simplicity, tradition and reliability.

Industrial Ethernet is growing faster than before: And taking more market share. At a growth rate of 20%, Ethernet now makes up for 38% of the global market compared to 34% in 2015. EtherNet/IP is the number one Ethernet network with 9% followed by PROFINET (8%). Runners-up are EtherCAT, Modbus-TCP and Ethernet POWERLINK.

Wireless is coming: For the first time, wireless technologies are on the chart with 4% of the worldwide industrial network market. WLAN is most popular, followed by Bluetooth.

The Internet of Things is a big driver for wireless technologies, opening up new automation architectures and is increasingly being considered for machine connectivity and control, including Bring Your Own Device (BYOD) solutions via tablets or smartphones.

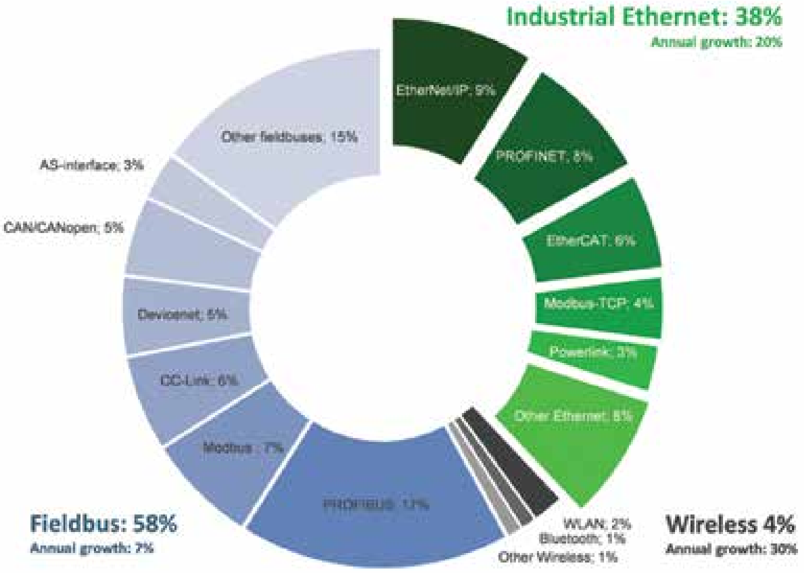


Figure 7.1: 2016 Industrial Networking market share according to HMS.

7.3.1 Regional breakdown

In Europe and the Middle East, PROFIBUS is the leading network while PROFINET has the fastest growth rate. Runners-up are EtherCAT, Modbus-TCP and POWERLINK. The US market is dominated by the CIP networks where EtherNet/IP is overtaking DeviceNet in terms of market shares. In Asia, no network stands out as truly market-leading—PROFIBUS, PROFINET, Ethernet/IP, Modbus and CC-Link are widely used. EtherCAT continues to establish itself as a significant network, and there are early signs of CC-Link IE Field being adopted.

Chapter 8
Conclusion

"Professor Google is your greatest friend.
Professor Google is your greatest enemy.
Balance these two."

8.1 Concluding remarks

This edition of the Industrial Communications Handbook has deviated from those in the past by not specifically focusing on the nitty-gritty of specific hardware, or protocols.

The actual nitty-gritty information can very often be found: [1, 2, 3] are excellent reference works, with [1] still being the best overview of RS-232 I have ever seen. Remember that the original RS-232 specification was for a 25-pin connector. This implies 2^{25} different combinations!! They didn't sell 'patch-boxes' for nothing! Remember too, that although the pared-down 9-pin connector arrived later, the combinations remained the same!

At my University, a new-fangled all-singing, all-dancing printing 'solution' has been rolled out. In theory, I can sit in my office, hit the print button, walk to any printing device anywhere on campus, swipe my card, and hey-presto, out pops my print job.

(Ignoring for the moment that the print job could be my exam, and that the card could have been suitably purloined …)

But a closer inspection of this marvellous, brand-new distributed system reveals (Dramatic Drum Roll):

RS-232

Yes, it's Ethernet as well, but the control bit contains the first 9-pin DIN I have seen in years. So someone, somewhere, very recently consulted[1], probably via Professor Google.

But what Professor Google will not tell you is really, really why you ought in this day and age to have considered using a USB port rather than an RS-232 port. Professor Google is not good at giving years worth of 'No, don't do it that way: this way is better' (for the following 12 reasons …) Professor Google is particularly not good at 'In myyyyy dayyyyy', whereas I excel there.

So what this incarnation of the Handbook attempts to do is give an overview of some of the Why's. The handbook, together with its predecessors, also contains a treasure trove of contacts: the first-port-of-call that make up the fine folk that advertise in these pages. Stop that hysterical, cynical laughter: you know its true!

If the handbook prevents you from bolting an omnidirectional antenna on the side of a metal cabinet; if you have a good idea of how high to mount your antenna; if you understand that you are employing wireless not just because its new; if you have a good argument for why a Secure Access Management system, or a basic VPN is required; if you have a considered argument for or against Proprietary/Open; if you understand that the S in SNMP is actually a C, then the handbook has done its job.

8.2 References

[1] Mick Crabtree (Ed. Alan Clark). Mick Crabtree's Industrial Communications Handbook. Crown Publications, second edition, 2005.

[2] Mick Crabtree (Ed. Alan Clark). Mick Crabtree's Industrial Communications Handbook. Crown Publications, third edition, 2008.

[3] Mick Crabtree (Ed. Karen Grant). Industrial Communications Handbook. Crown Publications, fourth edition, 2013.

[4] Karl Weber. Time sensitive networking: Deep impact or mission impossible? PC Control, 2015. Beckhoff Automation.

[5] Frank Hakemeyer. White paper: Trusted wireless 2.0—basics and practical applications. Phoenix Contact, 2013. Phoenix Contact.

[6] Tim Craven. Data acquisition from remote sites requires data security. Web: www.h3isquared.com.

[7] Anders Hansson. Industrial network market shares 2016 according to hms. HMS, 2015. www.idx.co.za.

[8] IEEE 802.3. Ethernet standard.

[9] IEEE 802.11. Wireless Local Area Network standard.

[10] IEEE 802.11a,b,g,n. Versions of WLAN standard.

[11] IEEE 802.1X. Port-based Network Access Control.

Ex
Available

Ex
Available

Abbreviations/ Acronyms

AAA	Authentication, Authorisation and Accounting
AES	Advanced Encryption Standard
AP	Access Point
BYOD	Bring Your Own Device
dB	decibel
DDOS	Distributed Denial Of Service
DMZ	DeMilitarised Zone
DSSS	Direct Sequence Spread Spectrum
EiRP	Effective isotropic Radiated Power
EM	Electromagnetic
EMC	Compatibility
EMI	Electromagnetic Interference
ERP	Effective Radiated Power
FHSS	Frequency Hopping Spread Spectrum
FM	Frequency Modulation
GSM	Global System for Mobile communication
HMS	Her Majesty's Ship
HQ	Headquarter
HSE	High Availability Seamless Redundancy
HVDC	High Voltage Direct Current
I/O	Input/Output
IDS	Intrusion Detection System
IEEE	Institute of Electrical and Electronic Engineers
IIoT	Industrial Internet of Things
IoT	Internet of Things
IP	Internet Protocol
IPS	Intrusion Prevention System
IPSec	Internet Protocol Security
IS-IS	Intermediate System to Intermediate System
ISM	Industrial, Scientific and Medical
LAN	Local Area Network
MAP	Manufacturing Automation Protocol
MIBs	Management Information Bases
MIMO	Multiple Input Multiple Output
MSTP	Multiple Spanning Tree Protocol
NMS	Network Management System
OID	Object Identifier
OPC UA	Open Platform Communications Unified Architecture
OSI	Open Systems Interconnection
PPTP	Point to Point Tunnelling Protocol
PRP	Parallel Redundancy Protocol
QoS	Quality of Service
RADIUS	Remote Authentication Dial-In User Service
RMS	Root Mean Square
RSTP	Rapid Spanning Tree Protocol
SAM	Secure Access Management
SCADA	Supervisory Control And Data Acquisition
SMA	SubMiniature version A
SNM	Simple Network Management
SNMP	Simple Network Management Protocol
SSID	Service Set Identifier
STP	Spanning Tree Protocol
TACACS	Terminal Access Controller Access Control System
TCP/IP	Transmission Control Protocol/ Internet Protocol
TEM	Transverse ElectroMagnetic
TSN	Time Sensitive Networking
TV	Television
UDP	User Datagram Protocol
USB	Universal Serial Bus
VHF	Very High Frequency
VLAN	Virtual Local Area Network
VPN	Virtual Private Network
WAN	Wide Area Network
WEP	Wired Equivalency Protocol
WLAN	Wireless Local Area Network

About the authors

Alan Robert Clark, Pr.Eng.

Professor Alan Robert Clark was born in Derby, England on the 18 June 1964. Emigrating to South Africa in June 1969, he became a South African citizen in 1984. He attended Brakpan Technical High School, and matriculated with full exemption in 1981. He attained the degree of B.Sc.(Eng) in Electrical engineering in 1987 from the University of the Witwatersrand, Johannesburg. In this period, he was employed by Unilever, implementing several industrial control projects. Joining the Department of Electrical Engineering as a Professional Assistant in 1988, half his time was spent developing antennas in the HF/VHF/UHF band for an industrial company. He was appointed as a lecturer in January 1990, and attained the degree of Ph.D. in Electrical Engineering in 1993, and was promoted to the post of Senior Lecturer in January 2000, and to Associate Professor in October 2002. Current lecturing includes a first course in Electromagnetics in the third year of study, and a High Frequency Techniques course in the fourth year.

Alan has supervised twenty-five M.Sc. postgraduates, one Ph.D postgraduate, and has one current Ph.D candidate. He has authored eight journal articles, four refereed conference publications, and eight other conference publications. He is a registered Professional Engineer (9 April 1992), a member of the South African Institute of Electrical Engineers, and a member of the Institute of Electrical and Electronic Engineers (USA). His research interests are in the field of computational electromagnetics, with particular emphasis on improving a Method-of-Moments code:SuperNEC; general antenna design; simulation and measurement techniques; modelling of electromagnetic propagation in conductive media, and recently, Electromagnetic Compatibility.

Additional interests include the InterNet, Unix (Linux), Microchip PIC microcontrollers, general electronic design, piano, birding and gardening. Married to Lesley on 20 March 1988, he has a son, Robert James, born 23 October 1999, and a daughter, Kathleen Brenda, born 18 June 2001.
Email: AlanRobertClark@gmail.com

Tim Craven

Tim Craven joined H3iSquared in 2008 and has been tinkering with Ethernet switches and routers ever since. He provides training, troubleshooting and consulting services in the field of industrial and utility Ethernet in and around South Africa.
Enquiries: Email tim@h3isquared.com

INDEX TO ADVERTISERS