THE BANDWIDTH BOTTLENECK

Researchers are scrambling to repair and expand data pipes worldwide — and to keep the information revolution from grinding to a halt.

BY JEFF HECHT

On 19 June, several hundred thousand US fans of the television drama *Game of Thrones* went online to watch an eagerly awaited episode — and triggered a partial failure in the channel's streaming service. Some 15,000 customers were left to rage at blank screens for more than an hour.

The channel, HBO, apologized and promised to avoid a repeat. But the incident was just one particularly public example of an increasingly urgent problem: with global Internet traffic growing by an estimated 22% per year, the demand for bandwidth is fast outstripping providers' best efforts to supply it.

Although huge progress has been made since the 1990s, when early web users had to use dial-up modems and endure 'the world wide wait', the Internet is still a global patchwork built on top of a century-old telephone system. The copper lines that originally formed the system's core have been replaced by fibre-optic cables carrying trillions of bits per second between massive data centres. But service levels are much lower on local links, and at the user end it can seem like the electronic equivalent of driving on dirt roads.

The resulting digital traffic jams threaten to throttle the information-technology revolution. Consumers

can already feel those constraints when mobile-phone calls become garbled at busy times, data connections slow to a crawl in crowded convention centres and video streams stall during peak viewing hours. Internet companies are painfully aware that today's network is far from ready for the much-promised future of mobile high-definition video, autonomous vehicles, remote surgery, telepresence and interactive 3D virtual-reality gaming.

That is why they are spending billions of dollars to clear the traffic jams and rebuild the Internet on the fly — an effort that is widely considered to be as crucial for the digital revolution as the expansion of computer power. Google has partnered with 5 Asian telecommunication companies to lay an 11,600-kilometre, US$300-million fibre-optic cable between Oregon, Japan and Taiwan that started service in June. Microsoft and Facebook are laying another cable across the Atlantic, to start service next year. "Those companies are making that fundamental investment to support their businesses," says Erik Kreifeldt, a submarine-cable expert at telecommunications market-research firm TeleGeography in Washington DC. These firms can't afford bottlenecks.

Laying new high-speed cable is just one improvement. Researchers and engineers are also trying several other fixes, from speeding up mobile networks to turbo-charging the servers that relay data around the world.

THE FIFTH GENERATION

For the time being, at least, one part of the expansion problem is comparatively easy to solve. Many areas in Europe and North America are already full of 'dark fibre': networks of optical fibres that were laid down by over-optimistic investors during the Internet bubble that finally burst in 2000, and never used. Today, providers can often meet rising demand simply by starting to use some of this dark fibre.

But such hard-wired connections don't help with the host of mobile phones, fitness trackers, virtual-reality headsets and other gadgets now coming online. Data traffic from mobile devices is increasing by an estimated 53% per year — most of which will end up going through mobile-phone towers, or 'base stations', whose coverage is already spotty, and whose bandwidth has to be shared by thousands of users.

The quality is spotty, as well. First-generation mobile-phone networks, introduced in the 1980s, used analogue signals and are long gone. But second-generation (2G) networks, which added digital services such as texting in the early 1990s, still account for 75% of mobile subscriptions in Africa and the Middle East, and are only now being phased out elsewhere. As of last year, the majority of mobile-phone users in Western Europe were on 3G networks, which were launched in the late 1990s to allow for more sophisticated digital services such as Internet access.

The most advanced commercial networks are now on 4G, which was introduced in the late 2000s to provide smartphones with broadband speeds of up to 100 megabits per second, and is now spreading fast. But to meet demand expected by the 2020s, say industry experts, wireless providers will have to start deploying fifth-generation (5G) technology that is at least 100 times faster, with top speeds measured in tens of billions of bits per second.

The 5G signals will also need to be shared much more widely than is currently feasible, says Rahim Tafazolli, head of the Institute for Communication Systems at the University of Surrey in Guildford, UK. "The target is how can we support a million devices per square kilometre," he says — enough to accommodate a burgeoning 'Internet of Things' that will range from networked household appliances to energy-control and medical-monitoring systems, and autonomous vehicles (see 'Bottleneck engineering').

The transition to 5G, like those to 3G and 4G before it, is being coordinated by an industry consortium that has retained the name Third

DIGITAL TRAFFIC JAMS THREATEN TO THROTTLE THE INFORMATION-TECHNOLOGY REVOLUTION.

Generation Partnership Project (3GPP). Tafazolli is working with this consortium to test a technique known as multiple-input, multiple-output (MIMO) — basically, a way to make each radio frequency carry many streams of data at once without letting them mix into gibberish. The idea is to put multiple antennas on both transmitter and receiver, creating many ways for signals to leave one and arrive at the other. Sophisticated signal processing can distinguish between the various paths, and extract independent data streams from each.

MIMO is already used in Wi-Fi and 4G networks. But the small size of smartphones currently limits them to no more than four antennas each, and the same number on base stations. So a key goal of 5G research is to squeeze more antennas onto both.

Big wireless companies have demonstrated MIMO with very high antenna counts in the lab and at trade shows. At the Mobile World Congress in Barcelona, Spain, in February, equipment-maker Ericsson ran live indoor demonstrations of a multiuser massive MIMO system, using a 512-element antenna to transmit 25 gigabits per second between a pair of terminals, one stationary and the other moving on rails. The system is one-quarter of the way to the 100-gigabit 5G target, and it transmits at 15 gigahertz, part of the high-frequency band planned for 5G. Japanese wireless operator NTT DoCoMo is working with Ericsson to test the equipment outdoors, and Korea Telecom is planning to demonstrate 5G services when South Korea hosts the next Winter Olympics, in 2018.

Another approach is to make the devices much more adaptive. Instead of operating on a single, hard-wired set of frequencies, a mobile device could use what is sometimes called cognitive radio: a device that uses software to switch its wireless links to whatever radio channel happens to be open at that moment. That would not only keep data automatically moving through the fastest channels, says Tafazolli, but also improve network resilience by finding ways to route around failure points. And, he says, it's much easier to upgrade performance by replacing software than by replacing hardware.

Meanwhile, a crucial policy challenge for the 5G transition is finding a radio spectrum that offers adequate bandwidth and coverage. International agreements have already allocated almost every accessible frequency to a specific use, such as television broadcasting, maritime navigation or even radio astronomy. So final changes will have to wait for the 2019 World Radiocommunication Conference. But the US Federal Communications Commission (FCC) is trying to get a head start by auctioning off frequencies below 1 gigahertz to telecommunications companies. Once reserved for broadcast television because they are better than higher frequencies at penetrating walls and other obstructions — but no longer needed after television's shift to digital — these low frequencies are particularly attractive for serving sparsely populated areas, says Tafazolli: only a few base stations would be required to provide broadband service to households and driving data to autonomous cars on motorways.

Other bands in the 1–6-gigahertz range could be opened up for 5G use as 2G and 3G technologies are phased out. But the best hope for dense urban areas is to exploit frequencies above 6 gigahertz, which are currently little-used because they have a very short range. That would require 5G base stations up to every 200 metres in dense urban areas, one-fifth the spacing typical of urban 4G networks. But the FCC considers the idea promising enough that on 14 July, it formally approved opening these frequencies for high-speed, fast-response services.

⟳ **NATURE.COM**
To listen to a podcast about the bandwidth challenge, visit:
go.nature.com/2axbk00

Ofcom, the UK regulatory body, is considering similar steps.

Companies are particularly interested in these higher frequencies as a way to extend 5G technology for other uses. In the United States, wireless carrier Verizon and a consortium

of equipment-makers including Ericsson, Cisco, Intel, Nokia and Samsung have tested 28-gigahertz transmission at sites in New Jersey, Massachusetts and Texas. The system uses 5G technology to deliver data at 1 gigabit per second, and Verizon is adapting it for use in fixed wireless connections to homes, which it plans to test next year. The company has been pushing fixed wireless as an alternative to wired connections, because connection costs are much lower.

BIGGER PIPES

"When I take out my cell phone, everyone thinks of it as a wireless communications device," says Neal Bergano, chief technology officer of TE SubCom, a submarine-cable manufacturer based in Eatontown, New Jersey. Yet that is only part of the story, he says: "Users are mobile, but the network isn't mobile." When someone uses their phone, its radio signal is converted at the nearest base station to an optical signal that then has to travel to its destination through fixed fibre optics.

These flexible glass data channels have been the backbone of the global telecommunications network for more than a quarter of a century. Nothing can match their bandwidth: today, a single hair-thin fibre can transmit 10 terabits (trillion bits) per second across the Atlantic. That is the equivalent of 25 double-layer Blu-ray Discs per second, and is 30,000 times the capacity of the first transatlantic fibre cable, laid in 1988. Much of that increase came when engineers learned how to send 100 separate signals through a single fibre, each at its own wavelength. But as traffic continues to increase over heavily used routes, such as New York to London, that approach is coming up against some hard limits: distortion and noise that inevitably build up as light passes along thousands of kilometres of glass have made it effectively impossible to send more than 100 gigabits per second on a single wavelength.

To overcome that limit, manufacturers have developed a new type of fibre. Whereas standard fibres send the light through a 9-micrometre-wide core of ultrapure glass running down the middle, the newer design spreads the light over a larger core area at lower intensity, reducing noise. The trade-off is that the new fibres are more sensitive to bending and stretching, which can introduce errors. But they work very well in submarine cables, because the deep sea provides a benign, stable environment that puts little strain on the fibre.

Last year, networking-systems firm Infinera in Sunnyvale, California, sent single-wavelength signals at 150 gigabits per second through a large-area fibre for 7,400 kilometres — more than 3 times the distance possible with a standard fibre, and easily enough to cross the Atlantic. They also transmitted 200-gigabit-per-second signals a shorter distance.

The highest-capacity commercial submarine cable now in service is the 60-terabit-per-second FASTER system that opened in June between Oregon and Japan. It sends 100-gigabit-per-second signals on 100 wavelengths in each of 6 pairs of large-core fibres. But in late May, Microsoft and Facebook jointly announced plans to beat it with MAREA: a large-area fibre cable spanning the 6,600 kilometres between Virginia and Spain. When completed in October 2017, the cable will link the two companies' data centres on opposite sides of the Atlantic at 160 terabits per second.

Another approach to reducing performance-limiting noise was demonstrated last year by a group at the University of California, San Diego. Fibre-optic systems normally use separate lasers for each wavelength, but tiny, random variations can generate noise. Instead, the group used a technique known as a frequency comb to generate a series of uniformly spaced wavelengths from a single laser (E. Temprana et al. Science 348, 1445–1448; 2015). "It worked like a charm" to reduce noise, says group member Nikola Alic, an electrical engineer. With further development, he says, the approach could double the data rate of fibre-optic systems.
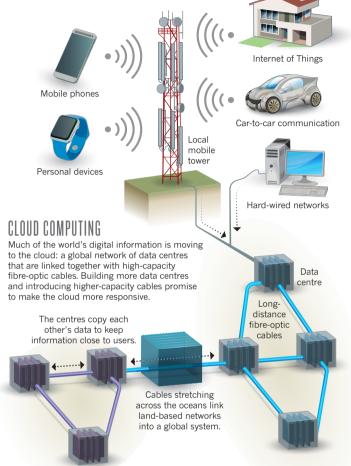
TIME OF FLIGHT

Impressive bandwidth is useful, but promptness also matters. Human speech is so sensitive to interruption that a delay of one-quarter of a second can disturb a phone or video conversation. Video requires a

BOTTLENECK ENGINEERING

The Internet was built on a century-old telephone system, leaving many choke points that have to be eliminated to keep the bits flowing.

MOBILE EVERYTHING

Demand for wireless connections is exploding, with ever more devices coming online. Engineers hope to meet that demand with fifth-generation (5G) networks that will increase data rates from millions to billions of bits per second.

CLOUD COMPUTING

Much of the world's digital information is moving to the cloud: a global network of data centres that are linked together with high-capacity fibre-optic cables. Building more data centres and introducing higher-capacity cables promise to make the cloud more responsive.

fixed frame rate, so streaming video stalls when its input queue runs dry. To overcome such problems, FCC rules allow special codes that give priority passage for packets of data carrying voice calls or video frames, so that they flow quickly and uniformly through the Internet.
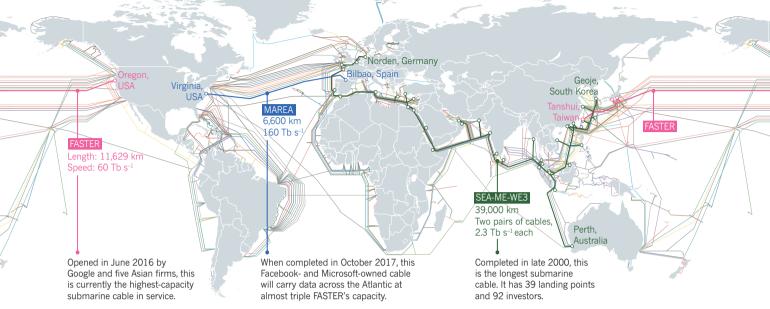
New and emerging services including telerobotics, remote surgery, cloud computing and interactive gaming are also sensitive to network responsiveness. The time it takes for a signal to make a round trip between two terminals, often called latency, depends largely on distance — a reality that shapes the geography of the Internet. Even though data travel through fibre-optic cable at 200,000 kilometres per second, two-thirds the velocity of light in the open air, a person tapping a key in London would still need 86 milliseconds to get a response from a data centre in San Francisco, 8,600 kilometres away — a delay that would make cloud computing crawl.

Emerging mobile applications require both broad bandwidth and low latency. Autonomous cars, for example, need real-time data on their environment to warn them about hazards, from potholes to accidents ahead. Conventional cars are becoming wireless nerve centres, needing low latency for 'hands-free' voice-control systems.

A potentially huge challenge is the emergence of 3D virtual-reality systems. Interactive 3D gaming requires data to travel at 1 gigabit per

THE SUBMARINE WEB
Much of the world's Internet traffic passes under the oceans, through fibre-optic cables that can run along the sea bed for thousands of kilometres. Companies are constantly laying more and better cables.

Norden, Germany
Bilbao, Spain
Geoje, South Korea
Tanshui, Taiwan
Oregon, USA
Virginia, USA

MAREA
6,600 km
160 Tb s⁻¹

FASTER
Length: 11,629 km
Speed: 60 Tb s⁻¹

FASTER

SEA-ME-WE3
39,000 km
Two pairs of cables,
2.3 Tb s⁻¹ each

Perth, Australia

Opened in June 2016 by Google and five Asian firms, this is currently the highest-capacity submarine cable in service.

When completed in October 2017, this Facebook- and Microsoft-owned cable will carry data across the Atlantic at almost triple FASTER's capacity.

Completed in late 2000, this is the longest submarine cable. It has 39 landing points and 92 investors.

second — 20 times the speed of a typical video feed from a Blu-Ray Disc. But most crucially, the image must be rewritten at least 90 times per second to keep up with users turning their heads to watch the action, says computer scientist David Whittinghill of Purdue University in West Lafayette, Indiana. If the data stream slips behind, the user gets motion sickness. To keep that from happening, Whittinghill has installed a special 10-gigabit-per-second fibre line to his virtual-reality lab.

To speed up responses, big Internet companies such as Google, Microsoft, Facebook and Amazon store replicas of their data in multiple server farms around the world, and route queries to the closest. Video cached at a local data centre is what allows viewers to fast-forward as if the file was stored on a home device, says Geoff Bennett, director of solutions and technology for Infinera. But the proliferation of these data centres is also one of the biggest drivers of bandwidth demand, he says: vendors' efforts to synchronize private data centres around the world now consume more bandwidth than public Internet traffic. The Microsoft–Facebook cable is being built expressly for this purpose (see 'The submarine web').

So far, most data centres are where the customers and cables are: in North America, Europe and east Asia. "Many parts of the world still rely on remote access to content that is not stored locally," says Kreifeldt. South America has few data centres, he says, so much of the content comes from well-wired Miami, Florida: traffic between Chile and Brazil might be routed through Miami to save money, but at a cost in latency. The same problem plagues the Middle East, where 85% of international traffic must travel to centres in Europe. That is changing, says Kreifeldt, but progress is slow. Amazon Web Services launched its first cloud data centre in India this year, in Mumbai; it has had a similar centre in São Paulo, Brazil, since 2011.

INTERNAL COMMUNICATIONS
Bandwidth is also crucial on the very smallest scale: on and between the chips in the banks of servers in a data centre. Expanding the flow here can help information to move more quickly within the data centres and get out to users faster. Chip clock speeds — how fast the chip runs — flat-lined at a few gigahertz several years ago, because of heating problems. The most practical way to speed up processors significantly is to divide the operations that they perform between multiple 'cores': separate microprocessors operating in parallel on the

"USERS ARE MOBILE, BUT THE NETWORK ISN'T MOBILE."

same chip. That requires high-speed connections within the chip — and one way to make them is with light, which can move data faster than electrons can.

The biggest obstacle has been integrating microscale optics with silicon electronics. After years of research on 'silicon photonics', engineers have yet to find a way to efficiently generate light from silicon, a key step in optical information processing. The best semiconductor light sources, such as indium phosphide, can be bonded to silicon chips, but are very difficult to grow directly on silicon, because their atoms are spaced differently. Optical and electronic components have been integrated on indium phosphide, but so far only on a small scale.

In an effort to scale up photonic integration to a commercial level, the United States last year launched the American Institute for Manufacturing Integrated Photonics in Rochester, New York, which is supported by $110 million from federal agencies and $502 million from industry and other sources. Its target is to develop an efficient technology to make integrated photonics for high-speed applications, including optical communications and computing.

Separately, a Canadian-funded team earlier this year demonstrated a photonic integrated circuit with 21 active components that could be programmed to perform 3 different logic functions (W. Liu *et al. Nature Photon.* **10,** 190–195; 2016). That's an important step for photonic microprocessors, comparable in complexity to the first programmable electronic chips that opened the door to microcomputers. "Compared to current electronics, it's simple, but compared to photonic integrated circuits it is quite complicated," says study co-author Jianping Yao, an electrical engineer at the University of Ottawa in Canada.

Further development could lead to varied applications. For example, Yao says that after the chip is optimized for manufacture, it could convert a 5G smartphone signal received at a base station into an analogue optical signal, which could be transmitted by fibre optics to a central facility, and then digitized.

The quest for faster chips, like other parts of the Internet problem, is a daunting challenge. But researchers such as Bergano see a lot of potential for improvements. After 35 years of working on fibre optics, he says, "I remain a complete optimist when I think about the future." ■

Jeff Hecht *is a freelance writer in Auburndale, Massachusetts.*