URSI GASS 2020, Rome, Italy, 29 August – 5 September 2020

The Time Fabric API: Using Metrology to Create Trust in Data

Richard G Hoptroff [(1)], Simon D Kenny [(1)]
[(1)] Hoptroff London Ltd, London, UK, *www.hoptroff.com*

Abstract

We present a system based on traceability to UTC that allows data to be immutably watermarked with the key identifiers of when and where the data was created. This creates a new class of data using metrology: data that can be believed to be true because it is verified by a physical standard.

1 The Need for Trust

Our world is increasingly governed by events that begin and end on computers. The only record of what happened is the data record the computers create. We are asked to trust this data record because the computer is trustworthy, and that the computer is trustworthy because the correct data record it produced proves it. The logic of the trustworthiness of machine records is circular, unless it is verified by some external reference

Regulations such as MiFID II and Consolidated Audit Trail recognise the need for data that can be trusted to reconstruct events. To establish causality in cascading financial events, this requires that computer's clocks are synchronised to UTC, in some cases to within 100µs, so that the sequence and interval of events is accurately recorded. At least where causality applies, it quickly becomes clear when data are wrong because it ceases to make sense as a description of the trading process.

Legal frameworks such as the EU's General Data Protection Regulation (GDPR) and the Privacy and Electronic Communications Regulations (PECR), and the California Consumer Privacy Act (CCPA), which limit what we can do with personal data, are more challenging.

These regulations are new and companies are still adjusting to what is required but they will need trustable data if they are to demonstrate meaningful compliance. For example, a recent report [1] demonstrated how advertisers typically do not fully explain what they will do with the personal data when they obtain user consent to its use; every time they auction an online ad spot, as they do hundreds of billions of times a day, they need to pass on the private user information to hundreds of potential bidders, usually with no back-to-back obligations on how the data is used, and when it must be deleted; detailed records of how data was used are very hard to reconstruct because they are fragmented across many parties and many machines.

2 Using Metrology to Trust Data

When an event happens on a computer that will have an effect in the real world, it should be recorded in such a way that can be audited to confirm when and where the digital event happened. This needs to go right down to the Edge, which might be a consumer's web browser or an IoT device whose data is used as a reference in analysis of a process or a dispute.

Without such a system, the records created have no logical support. If the reported outcomes of automated systems are to be trusted, it must be possible to audit the digital business world in as much detail as accountants audit the physical business world today.

We have developed a system to tackle this problem by combining 3 elements: *traceable time*, *traceable place*, and *data immutability*. At its heart is metrology: traceability to BIPM's Coordinated Universal Time (UTC) [2].

3 Key Concepts

3.1 Traceable Time

Traceable time is a clock that can create timestamps known to be correct by way of an unbroken chain of comparisons back to the national standards institutes who contribute to UTC. While this has been long used by industries such as telecoms and power generation for synchronization, it was only with the introduction of the MiFID II regulations that it was applied to the verification of digital event chains.

To create an unbroken chain of comparisons to UTC resiliently, globally, accurately and at low cost requires a time feed network that distributes time from multiple chains of comparisons to different UTC sources. As shown in figure 1, Hoptroff London built such a network to serve the financial services industry, where three clock sources are distributed, and developed the Loop Test method of monitoring time delivery over distance [3].

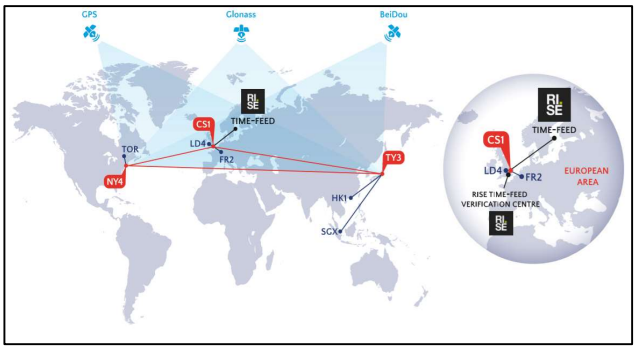

Figure 1. *Time disseminated using PTP [4] from GNSS satellite sources via timing hubs in New York, London and Tokyo, via telecoms extranets to data centres worldwide, monitored by Sweden's RISE standards institute over a terrestrial connection for resilience.*

On arrival at the data centre, the Time Fabric API uses software timestamping to filter the jittery long-distance PTP and allows applications to synchronize, or at least measure the offset of, their system clocks, so that the server's time is traceable back to UTC. We developed three Key Performance Indicators to measure the quality of this chain of comparisons:

Source Traceability compares the three UTC sources once they have completed their journey to the server to provide a measure of the loss of traceability over the time feed network. This loss is typically 1µs – 20µs depending on the distance and the type of network used [5].

Accuracy measures the clock offset between the median of the three UTC sources and the server's system clock used to create timestamps. This can be used to steer the system clock, with accuracies of better than 10µs being typically achieved. In some non-regulatory circumstances; steering may not be necessary. It may be sufficient to record the offset of the clock [6]. This is often needed at the Edge where we have little knowledge or control over the device in question.

Granularity measures the accuracy with which an application can obtain a timestamp from the system clock. Usually this is negligible, of the order of 100ns, but with heavily loaded virtualized systems, this measure can potentially become dominant [7].

3.2 Traceable Place

To prove where digital events happened in the physical world, we need the inextricable link between place and time. If a server tells some of its neighbours about an event, and the round-trip time of the message is measured, we can be confident that the event happened where the server claims it did.

Consider the London tube map (figure 2). It is a logical representation, much as we think of the internet as a logical network of IP addresses and routes we are traffic-controlled through within in a cloud, with no idea of the world above us.

Figure 2. *The logical map of the London underground, as originally conceived by graphic artist Paul Garbutt. Area of figure 4 shown dotted.* *(Transport for London)*

The physical reality is of course quite different (figure 3), as is the internet. Not just in geography, but the interchanges and branch lines; congestion in rush hour; unanticipated breakdowns; density proportional to demand.

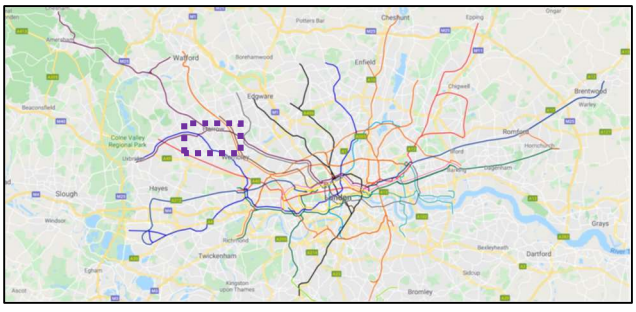

Figure 3. *Physical map of London underground. Area of figure 4 shown dotted.* *(Google LLC)*

Suppose you are a blind traveller; with little knowledge and no control of the network you are travelling on. The best way to be confident of where you are is if you time your travel from a known place a short distance away. You can be even more confident if some of your colleagues travelled short distances from other places, allowing confirmation by triangulation (figure 4).

Figure 4. *As with a transport network, confirming the place where events occurred within the internet fabric can be reduced to traceable measurements of journey times.* *(Google LLC)*

Just as transport networks have congestion, the internet fabric can be capricious in terms of latency. Without doubt, the shorter the journey, the better the location accuracy. But ,machine learning models, that can map network latencies across the global IP fabric, can be used to reduce doubt and provide likelihood estimates of actual location. This is an area of active research, see section 4.5 and [8].

3.3 Data Immutability

To demonstrate immutability in time, place and identity requires ledgers that can be trusted by all concerned and have a mechanism to guarantee they cannot be manipulated. Unlike time and place estimation, which only need to be accurate enough to be fit for purpose, immutability must be indisputable and impregnable. There is no margin of error.

In order to be trusted, data must be available to all concerned and be distributed evenly enough that the likelihood of manipulation of the record is vanishingly small.

Hash chain ledgers, first reported by Bayer, Haber and Stornetta [9], achieve this well by hash coding the event content with the hash code of the previous event (figure 5). Time and place can be recorded in the ledger at regular intervals irrespective of digital events, to prove a ledger's identity; hash ledgers only need to be self-consistent, unlike blockchain, since there is no proof of work burden.

Traceable Timestamp	Prior Hash	Hash
2019-11-03 12:34:36.945 715 TAI	3170958	1648979
2019-11-03 12:34:37.276 349 TAI	1648979	7807668
2019-11-03 12:34:38.167 649 TAI	7807668	9461354

Figure 5. *Hash chain ledgers render data immutable. By including the previous hash code in the next hash code calculation, any attempt at manipulation, resequencing or deletion is immediately apparent. We enhance the hash ledger by traceably timestamping it at regular intervals.*

4 Practical Implementation

4.1 Building a Trust Network

Since 2017, we have distributed time to data centres worldwide to accuracies of ~10µs, via timing hubs in New York, London and Tokyo, from multiple GNSS and terrestrial sources to serve MiFID II regulatory needs.

To extend this to immutability in time, place and identity, we have developed a Linux and Windows API (figure 6) for software developers that provides traceability and immutability services to software applications.

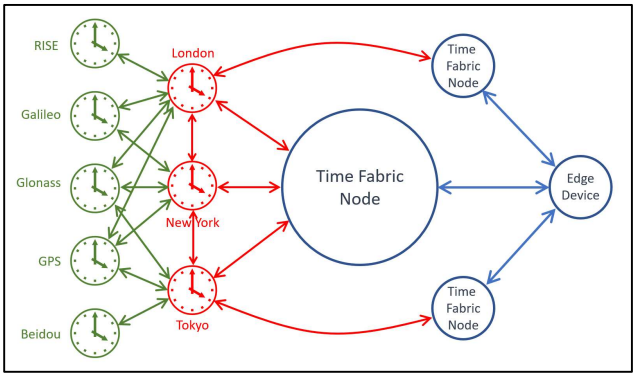

Figure 6. *The Time Fabric API network consists of traceable time distribution and network of Node devices to confirm the locality of any edge device.*

Time is distributed to secure servers using the Time Fabric API, ('Nodes') which then connect to other Nodes running the same API to provide accurate triangulation to edge devices where events must be recorded.

Neighbour Node discovery requires several considerations. Physical proximity is needed to maximise location accuracy; diversity, *ie* choosing Nodes that your other neighbours are not connected to, must be encouraged in order to build a self-organising mesh; nodes with too many connections should be avoided in the interests of load balancing; nodes may expire, so self-healing is needed.

While Nodes can be trusted to be stationary, edge devices cannot; their connections to Nodes need to be dynamic, and may be wireless. Based on initial tests, we anticipate wireless delivery accuracies of ~5µs over the wireless hop, given smart enough software support at the receiving end.

The same Node network is used to share traceability hashes in order to create an impregnable web of trust (figure 7).

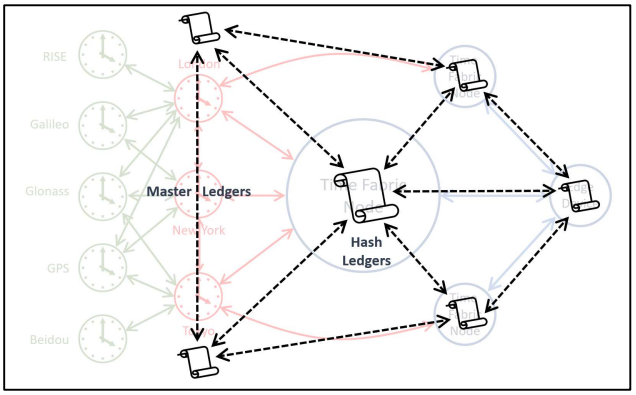

Figure 7. *The Time Fabric API cross-pollinates ledger hashes to build a network of trust.*

4.2 Trusting Edge Device Identity

We have little trust or control of Edge devices, so their codebase must be miniscule: "Tell me what time your clock says, and if you think you know where you are, tell me where". That's it. Optionally, the clock on the Edge device can be synchronised, but this is not a requirement [6]. Equally, the edge device can share its own hash ledger to prove its identity if desired (figure 7) [8].

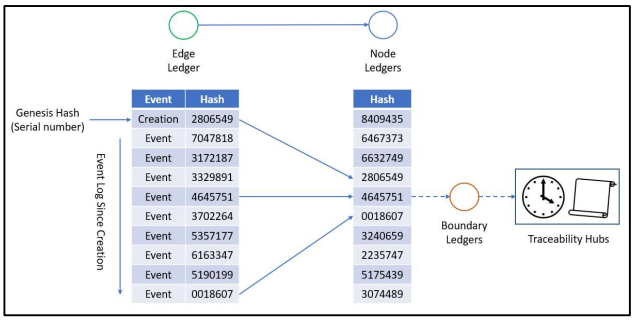

Figure 7. *An edge device can maintain a ledger from its inception; by sharing this with Nodes, its identity can be assured.*

Communication with Edge devices may be intermittent, *eg* a sensor such as a body-mounted video camera or a food transportation temperature sensor, occasionally being plugged in via USB for charging and data uploads. If intermittent, Nodes record the Edge device's clock offset so that the edge devices' timestamps can be adjusted for when it uploads reports of events.

4.3 The Ledger Audit Process

Figure 8 shows how the hash codes are woven into a hierarchy of ledgers and across Nodes. Data can be easily audited to confirm the time, place and identity of any event within the Node network.

The design is inherently resilient; even if half of the system ceased to function, the remainder would retain its integrity.

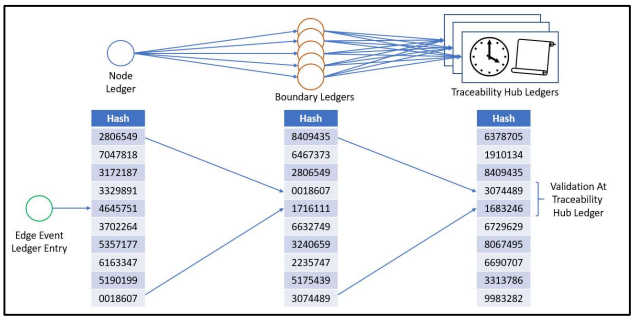

Figure 8. *The Node network is the cornerstone of trust. Hub ledgers provide a publication mechanism for external validation.*

4.4 Scalability

Figure 9 considers the scalability of the network. According to a recent forecast [10], sales of microcontrollers, the beating hearts of the Internet of Things, will reach 40 billion units per year by 2021. What would it take it implement the Time Fabric API on 40 billion edge devices?

Assuming each edge device connects to 5 Node devices running the Time Fabric API, and each API can connect to 10,000 edge devices. The API would be a small service

installed on 20 million servers within the existing computing infrastructure of edge device providers. Each Node would need to connect to three traceability time sources / ledgers ("boundaries"), distributed as software applications running on 10,000 servers in cloud data centres worldwide. At the top are three master traceability ledgers. The network is surprisingly scalable on a relatively low budget.

The design also has network effects. The more Nodes, the better the location assurance. Assuming the density of Node is proportional to population density, a city such as London would have around 12 Nodes, allowing place to be confirmed to the order of a kilometre.

Metric	Hubs	Boundaries	Nodes	Edge Devices
Number[1]	3 $^{\varepsilon}$	10,000 $^{\alpha}$	20 million	40 billion
Parent Count	–	3 $^{\phi}$	5	5
Children Count[2]	10,000 $^{\beta}$	10,000	10,000	–
Neighbour Count[3]	2	5-10	5-10	–
Hash Aggregation Ratio[4]	10,000	10,000 $^{\lambda}$	10,000	10,000
Sync Entry Rate[5]	1 per second	1 per second	1 per hour	1 per day
Event Entry Ratio[6]	75 $^{\sigma}$	25 $^{\delta}$	5	1

[1] Number of devices, *eg* $\alpha = \beta\varepsilon / \phi$

[2] The number of children that time is sent to and hash codes are aggregated from

[3] The number of neighbours with whom hash codes are exchanged and Proximity Delay measurements are made

[4] Number of ledger entries created before a hash is published to a parent (or otherwise published in the case of Hubs)

[5] Frequency of Traceable PTP and Proximity Delay measurements

[6] Ledger growth rate relative to rate of edge events, *eg* $\sigma = \beta\delta\phi / \lambda$

Figure 9. *Scalability calculations show that a large number of edge devices can be managed with a surprisingly small, resilient infrastructure employing a Node network facilitated by edge device vendors.*

4.5 Mapping the Virtual World

Many edge devices will know their location – or claim to know it, at least, for we cannot necessarily trust them. Edge devices that claim to know their location can advise the Node network; machine learning can then build a detailed map of the relationship between temporal proximity, *ie* round-trip time to neighbouring Nodes, and physical location.

The maps can then be used estimate location, advise on the accuracy of the estimate, and identify implausible location claims made by edge devices (figure 10).

Claimed Location	Proximity To Node 1	Proximity To Node 2	Proximity To Node 3		Estimated Location	Likelihood Radius, km	Claim Likelihood %
49°N, 157°W	156	124	265		49°N, 156°W	271	96
	269	56	89		30°N, 97°W	114	
44°N, 92°E	80	477	131	Machine Learning Models	43°N, 92°E	46	99
9°N, 64°W	53	109	300	→	9°N, 63°W	94	96
80°N, 20°W	238	96	50		61°N, 32°E	62	2
0°S, 31°W	261	262	171		1°N, 31°W	118	94
	169	54	85		3°S, 168°W	177	
	52	159	53		67°N, 163°W	456	
63°S, 3°E	51	74	85		61°S, 2°E	271	96
60°S, 130°W	89	80	182		61°S, 131°W	94	95

Not all edge devices know their location

Model of Location based on Proximity to Peer
Model of "standard deviation" of Estimated Location
Likelihood that Claimed Location is correct

Figure 10. *Using machine learning to map between round-trip time and physical location. Many devices will not know their location, in which case they will not claim one.*

5 Conclusions

The Time Fabric API (figure 11) provides a system for adapting the existing time metrology infrastructure to audit virtual events on identifiable edge devices, within a millisecond in time and within a kilometre in place. It is scalable, resilient, and provides us with the assurance we need that we can prove what happened to whom, where and when.

Consider the URSI-required text at top left on the first page of this paper: "*URSI GASS 2020, Rome, Italy, 29 August – 5 September 2020*". It identifies this matter of record in terms of who, where and when. The Time Fabric API achieves the same for the virtual world.

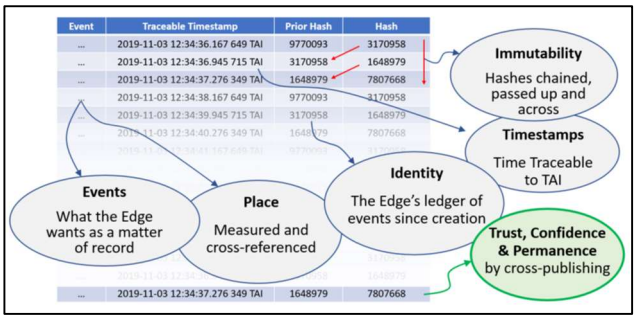

Figure 11. *The Time Fabric API combines traceable timing, location triangulation and hash ledger networks to create immutably watermarked data.*

References

[1] UK Information Commissioner's Office, "Update report into adtech and real time bidding 20 June 2019," ico.org.uk/media/about-the-ico/documents/2615156/adtech-real-time-bidding-report-201906.pdf (2019)

[2] Bureau International des Poids et Mesures, "Key products of the BIPM Time Department," www.bipm.org/en/bipm/tai

[3]* UK patent application GB1900789.7 *Method for Testing Time Distribution*, Hoptroff London Ltd

[4] IEEE Standard 1588-2008, Institute of Electrical and Electronics Engineers, ieee.org (2008)

[5]* UK patent application GB1622202.8 *Multi-GNSS Traceability Measurement*, Hoptroff London Ltd

[6]* UK patent application GB1911378.6 *Timestamping Events on Edge Devices*, Hoptroff London Ltd

[7]* UK patent application GB1614594.8 *Clock to Application Latency*, Hoptroff London Ltd

[8]* UK patent application GB1907677.7 *Watermarking Time, Place & Identity*, Hoptroff London Ltd

[9] D. Bayer, S. Haber, W.S. Stornetta, "Improving the efficiency and reliability of digital time-stamping", Sequences II: Methods in Communication, Security and Computer Science, pages 329-334 (1993)

[10] IC Insights, www.icinsights.com/news/bulletins/MCUs-Sales-To-Reach-RecordHigh-Annual-Revenues-Through-2022 (2018)

*If patent application is unpublished at the time of reading, contact the authors for a copy.