Invited Paper: Perishable Data in Smart Communities

Glenn Ricart
US Ignite and U. Utah
Salt Lake City, USA
Glenn.Ricart@us-ignite.org

Abstract—Distributed smart and connected community applications of the present and future will continuously sense perishable data on which they make decisions. We say the data are perishable when they represent information sensed at a past point in time, possibly at a distant part of the distributed application, and for which the ground truth could have since changed. This paper discusses actions the sensing source, the network, and the controller can take to make the best possible decisions given that we know only a delayed version of the truth. Some of these actions to preserve freshness of data are based on known real-time techniques such as Time Sensitive Networking. To these we add Information Theory applied on a message-by-message basis and cross-layer coordination between content and network handling. Both the unexpectedness (surprisal) and timeliness (freshness) properties of perishable information can be encoded using a Shannon information entropy framework. Examples of smart community systems of the future for which these concepts are relevant are given.

Keywords—smart and connected communities, latency, perishable data, Shannon information entropy, surprisal, cross-layer networking

I. INTRODUCTION

Data management in support of smart and connected communities has largely centered upon acquiring and unifying such data across disparate owners, sources, geographies, formats, and purposes [1,2]. The goal is often to create data lakes for analysis, to use big data techniques to find previously unknown relationships, and/or to feed Enterprise Resource Planning (ERP) Systems [3,4].

However, a new category of smart and connected community data is beginning to get more attention: perishable data [5,6]. Perishable data is data whose primary value lies in its ability to track or predict current and possibly transient phenomena in the community—important things that are happening now or are predicted to happen in the near future. Perishable tracking data becomes less valuable for immediate decision-making over time as it increasingly describes something that has happened in the past and may no longer represent the current situation. Perishable predictive data becomes less valuable over time as the situation being predicted may be changing and the predicted event may already have happened.

It’s important to state at the outset that many types of smart community data will have both perishable and archival value. The perishable value may lie in the ability to optimize or tune community services to specific transient events happening right now. The archival value may lie in (a) documenting the history of the information available for the transient response, and (b) creating a database of these transient events which can be used to train artificial intelligence (AI) and other predictive systems to better understand the classes of events and inform the community’s reactions to similar events that may occur in the future. These archival purposes are important but this particular paper will not further address them.

This paper looks at present and future smart and connected community perishable data either from sensors, smart objects, or people, or headed to actuators, smart objects, or people and discusses how it best might be handled by sources, smart community networks, controllers, and actuators. The goal is to maximize the usefulness of perishable data in making decisions and closing cyber-physical system loops. In simple terms, how do we make sure the perishable data stay “fresh” and relevant to making good decisions?

We find that the source, the network, and the decision-maker of perishable data all have important roles to play in preserving freshness and value and in collaborating with one another. Because of the key role of the network between the source and the controller, and between controller and actuator, we find that communication between the network and the rest of the cyber-physical system can help make sure message handling is cognizant of message content and vice versa.

The mechanisms needed to avoid undue network delay are largely known but apply to entire flows regardless of the information value of specific messages. This paper couples the information value of the perishable data to actions that can be taken by the source, network, and controller or decision-maker.

Section II describes smart and connected community applications that do or will rely upon perishable data. Section III briefly discusses the response time requirements for these applications. Section IV discusses existing time sensitive network techniques and their applicability. Section V assesses the value of perishable information using a Shannon information entropy framework. Section VI discusses pragmatic tools and wisdom for managing the freshness of the most important information to the decision-making and action-taking processes.

II. SMART COMMUNITY APPLICATIONS OF PERISHABLE DATA

This section describes several current and future smart and connected community applications of perishable data.
A. Microgrid power coordination

Microgrids communicate within themselves attempting to match power available to power demands [7]. The main short-term management technique is often demand management since many power sources are mechanical and are slow to spin up and down [8]. Wind power and fossil fuels fall into this category since the response time to power demand changes is in seconds. However, solar energy can fluctuate within milliseconds as clouds cover and uncover solar panels [9], and some sort of inertia is usually provided (usually in the form of batteries) to smooth solar generation [10].

If one grid has extra energy and another grid needs energy, microgrids can coordinate an interflow between microgrids [11]; the microgrid providing power must match the provision of that power to the recipient microgrid’s ability to utilize that additional power in both time and amount [12]. If energy is provided that can’t be immediately used or stored, that energy is wasted. If energy is not available when needed, there can be brownouts (low voltage) and blackouts (energy cut off). To avoid these, source and demand must be synchronized on a millisecond basis [13]. Synchrophasors are used to report the status of power and report to controllers to optimize the microgrids [14]. Optimal decisions depend upon understanding system supply, demand, and transmission facilities in real time. Optimal decisions depend on having accurate and current information. The older the information, the less accurate the input data is to the grid optimization algorithms. Delayed information can result in inefficiency, wasted power, brownouts and blackouts.

The synchrophasor information is perishable because supply and demand can change extremely quickly and even a fraction of a second later the correct microgrid and intergrid optimizations may be quite different and require different flows.

B. Connected Vehicle Management

Pollution from vehicles can be drastically decreased if we don’t have them frequently stop for traffic lights [15]. Connected autonomous vehicles can be marshalled and platooned to more efficiently use pavement, avoid stops and starts, and better share with bicycles and pedestrians in our smart and connected communities [16]. It’s a puzzle where all the pieces are in constant motion. Ideally, the vehicle destinations, locations, speeds, and destinations are all known and predictively optimized, perhaps with artificial intelligence [17].

However, how does the connected vehicle management system deal with a vehicle executing an emergency braking operation due to detecting a child running into the street? Any platooning vehicles must also brake simultaneously and intersection scheduling may need immediate alteration.

It’s easy to see that information about an unexpected change in the vehicles and their roadways is highly important and needs top priority handling. It’s also very perishable. Within several seconds, positions and speeds are only of historical interest for training neural networks [18].

C. Managing Ischemic Stroke Patients

“Time is Brain” in the case of a stroke inhibiting blood flow to the brain. Every minute in which a large vessel ischemic stroke is untreated, the average patient loses 1.9 million neurons, 13.8 billion synapses, and 12 km (7 miles) of axonal fibers [19]. Fortunately, there’s a treatment that can save most patients: administering tissue plasminogen activator (tPA) to dissolve the clot. But that treatment can also kill the patient if they happen to have an intracranial hemorrhage, subarachnoid hemorrhage, or suspected/confirmed endocarditis. Determining whether tPA is lifesaving or contraindicated can be largely determined by examining a high-resolution CT scan of the patient.

Some smart communities are deploying CT scanners in ambulances designated as mobile stroke treatment units [20]. They go to the stroke victim, take a CT scan, and send the DICOM image to doctors at the nearby stroke center who examine it for contraindications to tPA. The patient is then either treated with tPA locally or taken to the nearest airstrip or helipad for transport to a stroke center for surgery.

At present, the entire image (about 100MB compressed) is sent over the cellular network taking about a half-minute with a good (downtown area) connection, or about the time it takes the patient to lose one million neurons. In rural areas, the patient might lose multiple millions of neurons during the transmission time.

The information itself is also perishable. The state of the blood flow on the image reflects the amount of the brain still alive at the time the image is taken. Delays in sending the image mean that doctors have to guess at the continued deterioration.

D. Earthquake Prediction and Notification

Accelerometers can distinguish the pressure wave (p-wave) of an earthquake from the shear wave (s-wave). Because the p-wave travels faster but the s-wave causes more damage, the detection of the p-wave can be used to proactively stop elevators at the next floor, shut down transit systems, and minimize damage form the earthquake’s coming s-wave [21,22,23].

Obviously, a p-wave detection is an event that needs time-sensitive handling to be of use because the information is quite perishable; by the time the s-wave arrives, the p-wave information is of no practical use. The s-wave data will correlate with damage, but the damage will already be in progress by the time s-wave data is received by a controller.

Given the rarity of p-wave warnings, it may not make sense to have a dedicated network to report them if an existing network will give top priority to this perishable data.

E. Facilitating choruses and orchestras in a COVID-19 era

Our last example is on a happier note: musicians are looking for ways to sing and play in ensembles, choruses, and orchestras in an era of COVID-19.

There is a clear danger of spreading COVID-19 through aerosols generated and expelled through wind instruments such as flutes, clarinets, bassoons, trumpets, French horns and tubas [24]. The musician must produce high wind pressures which may loosen mucus and release COVID-19 particles.

Because singers expel virus along with their voices at extended distances [25,26], the safest way to practice and play is to link singers with very low latency electronic paths and have singers in individual rooms.
If those paths include packet forwarding, the voice samples are very perishable; Delays of more than 11.5 ms of total delay from the singer to the ensemble listeners make it difficult for the ensemble to maintain a steady rhythm or “beat” [27]. The same is true for wind instruments.

Coarse sampling (fewer bits per sample or lower sampling rate) can help to reduce required channel bandwidth and is desirable, musically, to help maintain ensemble playing if there is insufficient bandwidth to sustain higher sampling rates and more bits per sample.

Note that 11.5 msec. means that as a practical matter, ensembles are still only viable within limited geographic areas such as well-interconnected metropolitan areas unless very special and currently costly low latency networking techniques are used [28].

III. DETERMINISTIC RESPONSE TIME REQUIREMENTS

The response time requirements for cyber-physical control loops for the preceding smart and connected community applications are discussed in this section. While response times will vary, overall system design for cyber-physical systems usually depends on the response time that can be guaranteed or which is deterministic.

Applications that will appear to be instantaneous to human beings need to respond in the 5-35 ms range or less. Trained musicians can detect 10-20 ms time arrival differences between instruments or voices, and virtual reality headsets can cause a seasickness in some people if the response time between head motion and the update of the visualized image is more than 5 ms [29]. Therefore, simultaneous music performance needs total response time (not just network response time) of the previously stated 11.5 ms [27].

Vehicles moving at 100 kph change their position by almost 3 cm/ms. Microgrid fault detection and localization via synchronphasor requires a 10 ms response time and a time accuracy of 32 μs [30].

IV. TIME SENSITIVE NETWORKING

There are already mechanisms defined for express delivery of perishable packets but rarely implemented in the interconnected Internet present in most smart and connected communities. The Internet Engineering Task Force (IETF) began looking at the deterministic networking problem in 2015 resulting in a October 2019 problem statement RFC 8557 [31]. They set out use cases in RFC 8578 [32], and outlined an architecture for addressing it in RFC 8655 [33] (October 2019). The IETF looks at deterministic networking as involving: (a) time synchronization to better than 1 μs, (b) resource reservation for critical data streams, (c) extraordinarily low packet loss, and (d) guaranteed end-to-end latency.

IEEE working group 802.1 has a Time-Sensitive Networking (TSN) working group and standards [34] and considers multiple stream flows and latency that meet control loop frequency requirements. For example, a time-critical flow may have guaranteed time slot assignment, routing rules implemented by a fast programmable logic controller, and priority over other traffic [35]. Real-time systems are explicitly considered [36]. A significant portion of the research has been focused on implementations as part of wireless 5G [37].

Another approach is to use edge / local cloud / fog computing to manage the types of cyber-physical systems in smart and connected communities discussed in Section II. This is another very good tool for ensuring appropriate response times and is highly recommended. However, care must be taken to provide for seamless interconnections between the edges / local clouds / fog domains with other such systems and many of the synchronization and prioritization issues discussed in this paper re-emerge in that context. We are also beginning to see inter-related and inter-dependent smart community operating systems comprised of cyber-physical systems where these same issues apply.

The network portion of response time / latency requirements for packet-based networking can be broken down into: (a) speed-of-light in medium, (b) processing time by network elements (NICs, switches, etc.), and (c) queueing time (at NICs and switches) due to other traffic being present.

The first factor, the speed-of-light, is largely dependent upon the wire or fiber lengths over which the traffic must pass. However, the superior speed of light in air means that radio (e.g., microwave) communications can propagate faster than the speed of light in fiber. In addition, fiber may take twists and turns while radio radiates in straight lines. Smart and connected communities can analyze their rights of way to design shorter fiber paths for their distribution networks, and give those advantages to multiple competitive private providers by providing open access to those distribution networks. Advanced wireless can be used to leap barriers to fiber links.

The second factor, processing time by network elements (NICs, switches, etc.), consists of obtaining access to the medium (where applicable), serialization/deserialization of the packet (which depends upon the ratio of the packet size to the bit rate), and any packet overhead (such as synchronization preambles or postambles). This time can be minimized by using fast NICs and switches and minimizing the number of switching points (which double as possible failure points).

The third factor is any delay due to other traffic being present. That factor can be minimized by designing networks with large “headroom,” the unused-by-design bandwidth of the link capable of absorbing surges and unusual stochastic events. This technique is inexpensive on new builds. However, there are also known approaches to favor critical traffic based on prioritization, timeslot scheduling (such as in 5G), and deterministic networking / time sensitive networking.

Since error/retransmission and especially timeout/retransmission can add unacceptable latency, response time critical applications must either be able to deal with missing data and/or improve the odds of receiving that information by such techniques as forward error correction and sending duplicate information over disparate paths.

Time-sensitive networking is largely aimed at expediting critical time-sensitive messages at the flow level. This paper suggests that the value of the content in these flows varies over time and should also be considered when deciding how to prioritize and expedite messages.
V. INFORMATION THEORY AND CONTENT VALUE

In a stable cyber-physical system, perishable data delivered by sensors may result in small changes in operational parameters; if one of these sensor readings is lost or not delivered on time, it’s likely that the system will remain stable and compensate for the missing sensor reading when the next one is received. Stated another way, expected or close-to-expected sensor readings have fairly low information value because they only confirm what we already expect to be the situation. However, surprising or unexpected data (e.g., a child has run into the street, an earthquake p-wave has been detected, the choral director has just given the signal for a vocal cut-off) can have a large and time-sensitive impact on the system. The surprising or unexpected has high value.

However, that high value is perishable if the information is not delivered on time. If vehicles are not quickly notified to brake, or elevators are not told to stop, or a singer keeps singing, much of that high value has quickly decayed.

The information content (amount of surprise or sometimes surprisal or Shannon entropy) of receiving a perishable information message about an event E is a function that decreases as the probability p(E) of an event E increases [38].

\[ S(E) = - \log p(E) \]

The properties that you’d expect surprise to follow directly lead to this equation [39,40].

Some readers may note the relationship to Shannon Information Entropy [41] where a single event E is replaced by a set of i symbols. The Shannon entropy H (eta) of an information source is defined as:

\[ H = - \sum p_i \log_2(p_i) \]

Where \( p_i \) is the probability of occurrence of the i-th possible value of the communicated symbol. By using \( \log_2 \), the equation expresses the results in bits per symbol or in shannons [42].

We take the event of receiving a sensor reading to be represented by either the event E or the set of i symbols which encode E. The probability of an unexpected symbol/state is low, and its logarithm is highly negative, increasing the value of surprisal or S or the corresponding entropy H.

Now let’s insert perishability. If the information is not delivered quickly, two things happen. First, we know that we’ve lost valuable reaction time and must devise systems that can tolerate higher reaction times. And second, we know the system being measured is most likely continuing to change. Over time, the probability that the received event closely represents the current physical state decreases and probability it represents what are now adjacent states increases. Probabilistically, the system may currently be in one of a wide range of nearby states when the perishable reading arrives.

Models for information decay depend upon the information. The event “a child is in the street” could be replaced over time by “several children are in the street,” or “oops, it was a lawn chair blown by the wind, not a child,” or it could remain “there’s still a child in the street.” The event “a p-wave earthquake was detected” probably stands by itself and information about the cessation of the p-wave would not change the resultant action. The event “the conductor has conducted a vocal cut-off,” might be replaced with “everyone else has stopped singing,” or with “it turned out to be a double cut-off but not much harm done since others are still humming.”

Regardless, we do know that things could have changed since the original event E was reported. We know that the passage of time creates additional uncertainty and that additional uncertainty introduces additional entropy. Mathematically, one can compute the conditional entropy of the two events X (the random variable representing the current state) and Y (the random variable representing the state as reported some time ago) as

\[ H(X|Y) = - \sum_{x_i, y_j} p(x_i, y_j) \log \frac{p(x_i, y_j)}{p(y_j)} \]

where \( p(x_i, y_j) \) is the probability that \( X = x_i \) and \( Y = y_j \) and the divergence between X and Y in our case is time dependent [43]. The distributions of the random variables depend upon the specific application and prior mutually shared information.

To summarize, the value of the information received is related to both its surprisal (information entropy) and its ability to help us understand the state of the current system even though we only have delayed information (its joint entropy with how state might have changed). The result is that a single concept, entropy, can encode both surprisal and current usefulness considering time-dependent information decay (perishability).

VI. TOOLS AND WISDOM

So far, tools for dealing with perishable data have included using time-sensitive networking for minimizing network latency and using information theory to represent the value of message content based on surprisal and delay. Let’s look at a few more tools and wisdom that apply when response time is critical.

A. Don’t use TCP or other re-transmission protocols blindly

All parties involved would rather have a new reading rather than a repeat of an old one.

B. Dynamic system properties

Given that information delay in a closed loop can easily cause instability or oscillation in a closed-loop dynamic system, think about putting an appropriate amount of damping into the decision-making model. The amount of that damping is often a function of the maximum amount of delay in the full cyber-physical loop, so minimizing the delay can help stabilize the system.

C. Time-stamping perishable data

All the devices in a distributed system have a somewhat delayed view of what everyone else knows. If, however, there is a degree of relative or absolute accuracy to time-stamps attached to data, it is easier to understand the possible information decay and/or to use the data received to better model or predict the uncertain current state at another place in the distributed system [44].
D. Leverage mutually-known information

If a sensor knows that previous information it has sent has been received correctly, it can compress future information. For example, instead of sending a full latitude/longitude, it could send the difference in latitude/longitude since the last report. The reduced packet size will spend less time serializing/deserializing on the networking media and be a smaller roadblock to other queued data. It may take less energy to send.

Further, if the sensor knows the information it currently has will not change what the controller (decision-maker) will do, it need not send the information at all.

E. Cross-layer communication between network and content

If the network can be told by the sensor or source the amount of information entropy in the message (as it will be viewed by the controller), the network can give differential service to messages with high surprisal. Perhaps some header bits visible to the network should encode the message surprisal. Reciprocally, if the network can tell the sensor or source the current state of network queueing, the sensor can adjust reporting rates or the precision of readings to better fit into available network bandwidth, taking into account the amount of surprisal in the readings themselves.

F. Controller to smart object collaboration

If the controller needs more frequent and accurate data to execute, for instance, one connected autonomous vehicle passing another, it may ask both the passing and to-be-passed vehicles to use more deterministic and more frequent communications during the passing maneuver. Both vehicles may increase surety of delivery by, for example, adding stronger forward error correction.

G. Network-managed overtaking information

If packets have headers that tell the network the type of message they contain, a new message with updated information could be used to replace an earlier message if the earlier message is still being buffered in the network. This could be quite common for lower surprisal messages being given low priority. When one is delivered, it should be the most recent one available.

H. Discard the least valuable messages

If messages are coded for both surprisal and current delay, discard those with low surprisal and high accumulated (or anticipated) delay.

I. Big picture before details

Up-to-date “big picture” messages are usually more valuable than delayed messages with lots of detail. The big picture may contain the bulk of the surprisal value. Sending details with less delay drastically in these situations.

K. Leverage forward-looking research networking facilities

US Ignite recommends that communities work with their academic partners, regional optical networks, Internet2, FABRIC [45], and forward-looking industry partners because they may have access to emerging network capabilities friendly to perishable data.

VII. Conclusion

Several forward-looking and future smart and connected community applications will generate perishable data—data that represent the current state of a portion of their system—but, in a distributed system, may be out-of-date by the time they arrive.

Without implementing an entirely new low-latency smart community data network, existing time sensitive networking capabilities can be added to existing networks. A low-latency local Internet exchange can also help to avoid delays caused by out-of-area peering points.

Not all smart community data need deterministic networking delivery, but for those which do, this paper has set forth an information theory framework for identifying the data which will have the most effect on the control models—have the most surprisal value or Shannon information entropy—and hence deserve to get priority treatment (among the priorities allowed to a given application).

Delaying any such messages means the believability of the information is likely to be eroded because the data now represent a reading from the past from a sensor or system that is continuing to evolve and gather new readings. That erosion can be measured in terms of information entropy as well using a conditional entropy based on the probabilities the reported state has moved to any subsequent state.

Finally, the paper has cataloged a set of tools and wisdom that may be useful to minimize the decay of perishable data in practical smart and connected community settings.

ACKNOWLEDGMENT

The author wishes to thank Tamer Nadeem for assistance in the selection of the topic.

REFERENCES


Li, Qing, Dong Li, and Xinbo Sun. "Key Technologies of Time-Sensitive Networking use cases." draft-finn-detnet-problem-statement-05 (work in progress) (2016).


