Holistic approach to urgent computing for flood decision support

Robert Brzoza-Woch¹, Marek Konieczny¹, Bartosz Kwolek¹, Piotr Nawrocki¹, Tomasz Szydło¹, and Krzysztof Zieliński¹

AGH University of Science and Technology, Department of Computer Science, Krakow, Poland

Abstract
This paper presents the concept of holistic approach to urgent computing which extends resources management in situation of emergency from computational resources to Data Acquisition and Preprocessing System. The layered structure of this system is presented in detail and its rearrangement in case of emergency is proposed. This process is harmonized with large scale computation using Urgent Service Profile. The proposed approach was validated by practical work performed under ISMOP project. Concrete examples of Urgent Service Profile definition have been discussed. Results of preliminary experiments related to energy management and data transmission optimization in case of emergency have been presented.

Keywords: Urgent computing, flood decision support, telemetry networks, wireless sensor networks, software reconfiguration, fog computing

1 Introduction
Recent research on urgent computing environments is oriented on improving resource authorization and allocation for time-critical applications and workflows using shared computing resources in Clouds and high-performance computing environments. Several examples of urgent computing applications and workflows have been proposed that include: severe weather forecasting workflows [8], simulation applications [7], storm surge modeling applications [3], and wildfire forecasting workflow [13].

The presented paper origins from research performed under the project ISMOP: a computerized levee monitoring and decision support system. One of the primary goals of this project is construction of a system to forecast danger of the loss of the levee structure stability and estimation of the levee breach risk leading to catastrophe.

The realization of this goal involves mathematical modeling of behavior of a levee exposed to a passing flood wave which can last from few hours to even several weeks. The water exposition of levee creates changes of their physical parameters which could be monitored on-line and used by mathematical models calculated on-demand in case of a flood emergency. In this case, the decision process support requires not only urgent computation and access to stored data but,
first of all, a real-time acquisition of a great amount of data (needed for calculations) from sensors installed over many kilometers of levees. This puts the management of sensors, control and measurement stations, and communication network as an integral part of flood decision support system operation in case of emergency.

The paper scientific contribution is the proposition of a holistic approach to urgent computing, which focuses not only on management of computational [12] and storage [5] resources in situation of emergency but also takes into account rearrangement of a data acquisition system in order to satisfy requirements of up-to-date data delivery to ongoing computations. This data should be transferred in real time on-demand to computing center to actualize or parameterize the existing decision models or satisfy the needs of modified computing processes. This paper does not consider semantic aspects of the decision support procedure [2], but it focuses on the data delivery aspect specifically.

The holistic approach key factor is the interoperation of the real-time data acquisition system with computations performed in the Cloud in urgent state. The paper presents this interoperation with focus on elements which characterize the urgent activity of the data acquisition system.

The paper is organized as follows. Section 2 overviews related works. Section 3 explains the holistic approach to urgent computing in more details with particular stress on the data acquisition system operation in urgent state. All elements of the proposed approach are illustrated by the implementation performed under Project ISMOP and presented in Section 4. The following Section 5 describes the preliminary experiment. Finally, Section 6 concludes the paper.

2 Related work

The wide literature of urgent computing research [8] [7] [3] [13] presents a variety of studies focusing on resource management issues of High-performance computing (HPC) or Grid computing infrastructures [12], [6]. In majority of the works, the urgent on-demand data delivery is taken for granted, in particular when internal storage resources of the infrastructures are used. The persistence of data resources and effective delivery channels are presumed to be guaranteed and reliable (e.g. Data Cloud in [8], Resource/Resource Access Layer in [3]). It is assumed that the usage of proper network protocols or distributed system techniques (e.g. Web-services as in [3]) solves the problem, but real-time measurement and transmission issues (i.e. maintenance and adaptation of communication channels) are left unnoticed.

The demand for further urgent data delivery research was recognized, as in LEAD environment [8], where the issue of real-time data flows was pointed out in plane of the complexity of managing data flows, dealing with multiple and changing data formats, and synchronizing complex data ingest. There are also other solutions, e.g. Urgent Data Management Framework (UDMF) [5] or robust data placement for SPRUCE [6]. However, they still refer to stored data, whereas real-time monitoring must be supported with ongoing up-to-date environmental measurements.

Sensor networks are a commonly used solution [9] for such real world monitoring (e.g. for wildland fires control [13], dikes condition governance [11] etc.). However, dealing with data streams received from, for instance, a number of sensors in embankments demands relevant solutions [14].

In this area, UDMF [5] made a step forward by pointing out additional capabilities not yet provided by urgent computing infrastructures, such as Quality of Service (QoS) management and monitoring mechanisms, data policy management tools etc. It provides workflows and users
with several capabilities of urgent storage and data management, including the configuration of QoS and Service Level Agreements (SLAs) for data services. However, the solution is applicable to cloud infrastructure data storages. We have not found works dedicated specifically to support emergency up-to-date data delivery to urgent computing systems from remote environmental sensor networks.

3 Holistic approach concept and requirements

The concept of the holistic approach to urgent computing is explained in Figure 1. The Large Scale Computing System (LSCS) e.g. Cloud or Grid data center is extended with Data Acquisition and Preprocessing System (DAPS). In case of emergency, both systems should cooperate in a coherent way to perform the requested urgent computation on time. The delay in data delivery could have the same catastrophic impact on a decision support system as a shortage of computational resources.

These two systems have to harmonize their operation in critical situation, so they should follow a global policy definition regulated by adaptation or reconfiguration processes. The studies presented in urgent computing literature propose different resource management techniques referring to central LSCS only. It is why in this section the DAPS behavior will be investigated at the first place.

Figure 1: The scope of holistic urgent computing approach.

DAPS has a multilayer structure built of three layers which perform the following operation:

**Measuring Layer** is composed of sensors measuring physical parameters of the environment. Values of these parameters are sent over a wired or wireless network to edge computing devices.

**Edge Computing Layer** is a collection of many distributed computing devices which control Measuring Layer operation, perform the measured data preprocessing such as data compression, encryption, filtering etc. In more advanced systems this is the place where event processing could be effectively deployed. This layer operation is referred to as Fog Computing [4].

**Communication Layer** provides bidirectional communication between Edge Computing devices and LSCS. This layer performs routing operation and selects the most suitable communication technology and routes to transfer preprocessed data to the central system. Taking into account the diversity of existing technologies, it opens a large space for optimization.
To manage DAPS effectively in the case of an emergency situation the Urgent Service Profile (USP) selected by LSCS should be deployed over system layers described above. This operation should be performed in parallel with the rearrangement of the central system resources. For each DAPS layer, such profile should specify the following set of parameters:

- SLA specification - a set of QoS parameters which values should be satisfied in a state of an emergency,
- configuration or adaptation policy specification in case of autonomous reconfiguration of the layer - a requested operational configuration in a state of an emergency,
- application related data delivery conditions and preprocessing requirements - a specification of data that should be delivered to LSCS to effectively support the decision processes,
- recovery procedure in case of failure detection - a specification of the layer activity in the situation of reduced availability of resources or breakdown of some components.

In general, USP may contain a parameter value or a policy describing the desired behavior. In the more advanced case, a layer may operate autonomously in order to satisfy (under the existing constraints) the goal set by the policy.

It is important to point out that DAPS has limited resources, thus it is designed to optimize the consumption of energy and IO channels during standard operation, i.e. uses Standard Service Profiles (SSP). In this general-purpose mode system resource budget is not violated. In contrast, USPs impose extended requirements of delivering up-to-date data to the LSCS (e.g. minimized latency, increased sampling frequency etc.). In this perspective, the goal of applying USPs is to assure the continuity of the computations by keeping the fluency of data transmission and avoiding data starvation (i.e. deficiency in input runtime data) at any cost, because in such case the necessary computations cannot be ran even when enough resources is allocated by LSCS. The process of switching between SSP and USP should be realized seamlessly, which means that parameters of the system are smoothly adjusted to avoid loss of data and additional latency.

On the one hand, not only are the DAPS resources limited, but during environmental disasters the real-time data sources within the catastrophe areas are prone to be affected (e.g. loss of connection, power depletion, hardware malfunction etc.). On the other hand, the fulfillment of the requirements setup by USP will inevitably result in the increased consumption of the limited resources.

Therefore, the system may sacrifice the cost and optimization of its operation (e.g. energy efficiency, security, economy etc.) at the expense of fulfilling additional needs. For example, measurements have to be performed more frequently, which consumes more energy. To balance this tradeoff, the edge computing layer can switch data encryption off, because it increases the data processing time. Communication layer can choose the fastest but cost-intensive connection to send the increased amount of data and lower the delays. The management of such behaviour requires constant adaptation and optimization techniques.

In the next section, we show how the DAPS for the levee monitoring system has been designed.

4 Holistic approach use case

We have developed a practical implementation of the prerequisites described in Section 3 which are specific to the ISMOP decision support system [1]. This system involves various data
analyses performed automatically or on-demand, including time series analysis of sensor measurements delivered by DAPS, prediction of the levee state, and flood threat assessment. The calculations involved in the computation of the threat level boil down to comparing suitable time series data sets that should be available on-line in urgent state. Hence, their accuracy strongly depends on DAPS operation in this state.

To present this operation, the abstract architecture of DAPS introduced in Section 3, is mapped to hardware and software elements of the ISMOP acquisition system. Figure 2 presents the system elements and information flow between all layers of the system. An example of a USP is presented in Table 1 and explained in following sections.

![Diagram of the system elements and information flow of the DAPPS implementation.](image)

**Figure 2:** The system elements and information flow of the DAPPS implementation.

### 4.1 Hardware and low-level software solutions

The developed control and measurement station is equipped with multiple hardware subsystems and software components which support urgent scenario data acquisition and transmission modes. The data collection hardware modules have been developed in such way that aging has little or no effect on them. This increases the data acquisition system’s overall reliability, especially in the states in which the system needs to provide data for urgent computations.

The main prerequisite for the power supply subsystem is that it has to be completely autonomous, mainly to utilize energy from locally available renewable sources. In our sample solution we use photovoltaic solar cells with electric double-layer capacitors (EDLC) also called supercapacitors (SC) and an elaborated custom-built power supply. The SC are far less sensitive to temperature variations and aging than standard rechargeable electrochemical cells and their number of charge-discharge cycles is virtually unlimited. The power supply is designed in such way that it supports large voltage variations, which are a completely normal phenomenon for SC-based energy storage. Secondly, the power supply supports an urgent profile implemented with an additional hardware that includes a boost converter and multiple management circuits. It can drain energy stored in the SC bank to extremely low levels. This, in turn, may cause the station to be even inoperable (SCs discharged almost completely), thus it should be used in critical situations only. Obviously, we have also implemented standard power-saving mechanisms with multiple power management options, i.e. the power supply to each important processing and communication subsystem can be independently enabled or disabled.

To provide even more flexible software adaptation mechanisms in the future, we considered
a possibility to reprogram in urgent state the embedded part of the control and measurement station with a new firmware by the local industrial computer. This feature can be easily added with a dedicated programming interface hardware or by using a common communication interface, e.g. a serial port.

4.2 Communication

In order to transfer data from the DAPS to the LSCS, it was necessary to develop an appropriate communication layer which uses a variety of wireless technologies (such as GPRS, EDGE, HSPA, LTE, Wi-Fi and XBee) for data transmission using MQTT (Message Queue Telemetry Transport), MQTT-SN or the power-aware variation of the MQTT protocol that we have developed during the course of our research. Many flood levees that require monitoring are outside urban areas where the infrastructure available for data transfer is not extensive enough.

In our system, that Wi-Fi access (via local ISPs) and 3G/4G network availability (e.g. LTE) are analyzed first. It is most often the case that the access to these data transfer technologies is possible only in built-up areas; however, wherever these networks are within range, they can be used to transmit data at high speeds of up to 50-100 Mbit/s. If using fast data transfer technologies is impossible, data is sent via GPRS modules. In some areas, it is possible to use the EDGE technology, which is an enhanced version of GPRS and allows you to achieve a transfer speed up to 236.8 Kbit/s.

As the wideband connections are not commonly available outside urban areas, the developed system is ready to use the GSM technology and the GPRS data transfer method as its most basic mechanism for transferring measurement data, which are available almost everywhere in Poland. If the wireless connection to any near Base Transceiver Station (BTS) is too weak to provide reliable GPRS data transmission, the station may choose the communication urgent profile and use the Short Message Service (SMS) as an alternative.

We assume that the GSM/GPRS network may be completely unavailable on any data acquisition station. This is a scenario with relatively low probability e.g. it happens during local severe weather conditions which lead to serious natural disasters. This may result that the entire infrastructure built by operators providing data transfer and text messaging services may be damaged. In this case, the station may run another communication urgent profile and use XBee technology to join MANET network consisting of a number of adjacent stations. The XBee version used in the ISMOP system (XBee 868LP for Europe) enables data transfer at the speed of 80 kb/s at the distance of up to 8.4 km. The data can be sent using the MQTT or MQTT-SN protocol to the nearest control and measurement station, and then transferred further via a wireless technology such as GPRS, Wi-Fi or LTE.

Moreover, the utilized XBee communication modules require low power which is also an important advantage in the presented solution.

4.3 Software reconfiguration and self-organization

In order to handle urgent scenarios properly, software applications responsible for the data acquisition process need to adapt to the new scenarios as well. Once the definition of USP is obtained, it will be applied to a station operating system as well as to all currently running applications. This adaptation process can be achieved in two ways: applications can handle changes themselves in their execution context (i.e. all actions are hardcoded in applications) or applications can be dynamically reconfigured in order to adapt to a new scenario (with help of external service). In our solution, we choose the second approach, because it gives more flexibility and allows us to improve management scalability of entire DAPS. The state
of each application is described by policies, compliant with promise theory. In this approach, applications declare promises regarding the functionalities they are expected to deliver. If, as the result of changes in context, the application will not be able to fulfill its promise, some reconfiguration actions may be taken. This actions may include changes in configuration files or applications modules update.

Dynamic switching between different communication media or encoding schemes requires changes in network configuration of operating system as well as it may require changes in application runtime (e.g. download of new drivers). In addition, required actions may vary for different fragments of embankments. For example, on some part of infrastructure the available Wi-Fi network may have different configuration options than on another. Controlling of all this parameters in the imperative way has limited applicability due to the complexity of the process. In our case, we are using declarative description of policies (declared in the USP), which are later applied autonomously by a control and measurement station to its operating system, as well as to all its applications.

Not only software performing edge computation needs to adopt during urgent data acquisition scenarios, but also LSCS processing might be affected [16]. In the proposed architecture data are sent to further processing via asynchronous messaging middleware. Because urgent computing proposes resource optimization on LSCS site in case of a critical situation, some services will not consume as much data as in regular scenarios (because they are preempted). It means that there is a serious threat that queues in messaging system may overflow or can be blocked by already delivered data. In order to address this issue, data important for urgent processing need to be prioritized. It can be achieved in two ways: (i) by setting proper priorities on messages sent by edge application or (ii) applications should send data to different higher priority queues (or messaging systems). Both approaches require software reconfiguration on LSCS and DAPS site.

USP may also affect the software responsible for handling Measurement layer. During urgent scenarios frequency of data readouts from sensors (all or selected ones) may increase. It is also possible that not all data are required, so the station can omit some measurement data by simply not sending them.

<table>
<thead>
<tr>
<th>Layer</th>
<th>Urgent Service Profile</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication</td>
<td>SLA: use communication channel with reliability above 75% [10]</td>
</tr>
<tr>
<td></td>
<td>Configuration: use in order: LTE, HSPA, EDGE, GPRS, SMS and XBee</td>
</tr>
<tr>
<td></td>
<td>Data delivery: disable messages aggregation</td>
</tr>
<tr>
<td></td>
<td>Recovery procedure: restart networking modules</td>
</tr>
<tr>
<td>Edge Computing</td>
<td>SLA: processing delay less than 0.5 sec</td>
</tr>
<tr>
<td></td>
<td>Configuration: switch off security features</td>
</tr>
<tr>
<td></td>
<td>Data delivery: enable messages prioritization</td>
</tr>
<tr>
<td></td>
<td>Recovery procedure: in case of low energy use boost converter</td>
</tr>
<tr>
<td>Measurement</td>
<td>SLA: measure five time per minute</td>
</tr>
<tr>
<td></td>
<td>Configuration: disable power saving modes on measurement network</td>
</tr>
<tr>
<td></td>
<td>Data delivery: increase accuracy temperature measurements to two digit points</td>
</tr>
<tr>
<td></td>
<td>Recovery procedure: force execution of e-calibration procedures</td>
</tr>
</tbody>
</table>
5 Preliminary experiment

We have conducted several tests of the developed control and measurement station. Most of them concern power requirements as they are perceived as one of the strongest constraints on the data acquisition system.

Supercapacitors (SCs) can be charged very quickly, but only if provided with large input currents. It is assumed that the SCs are charged with medium-sized photovoltaic cells which deliver 10 to 20 W of power. Such cells allow us to charge SCs at current levels of $\approx 500$ mA during normal operating conditions in daylight. A sample graph which shows the charging cycles of the SC batteries is shown in Figure 3 a). The batteries are built with 5 SCs of 200 F and 1200 F each. The graph shows that the system can be ready for operation in about 20 min. with the smaller battery or in less than 2 hours with the larger one. In daylight conditions, with full sun exposure, the system can be perceived as constantly ready for uninterrupted operation.

![Figure 3: Supercapacitor batteries charging cycles (a) and discharge cycle (b).](image)

The sample SC discharge cycle during system operation is shown in Figure 3 b). The graph shows how the battery with 5 SCs of 1200 F is discharged. At the end of the discharge cycle, the additional boost converter is enabled which is necessary during urgent data acquisition scenario. Notice the large current drawn from the battery during that time. The large current consumption is required to supply constant power to the system in low voltage conditions. Thus, the urgent data acquisition scenario should be used with caution as it can drain the SC energy to very low levels, leaving the control and measurement station inoperable when no renewable energy source is available.

We have also analyzed the communication protocol for transferring measured data from the control and measurement stations to the central part of the system [15]. In the ISMOP project, we have decided to use a message-oriented communication protocol MQTT-SN designed for sensor networks which is a modified version of the MQTT. Both protocols support QoS. The difference is that the MQTT uses TCP protocol, whereas the MQTT-SN uses UDP. Because of the fact that some data from the control and measurement stations can be prioritized in order to be sent using best effort method with QoS=0, the reliability introduced by the TCP is unnecessary. Moreover, it could consume energy for retransmitting unimportant data in urgent situations. In the MQTT-SN protocol, the acknowledgment of the data sent using QoS=1 is performed in application layer.

2394
Table 2: Comparison of MoM protocols over GSM network

<table>
<thead>
<tr>
<th>Protocol (layer 7)</th>
<th>Protocol (layer 4)</th>
<th>Data aggregation</th>
<th>Energy consumption for transferring 20kB</th>
<th>Delay introduced by communication protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>MQTT</td>
<td>TCP</td>
<td>in TCP</td>
<td>280Ws</td>
<td>0.2s</td>
</tr>
<tr>
<td>MQTT-SN</td>
<td>UDP</td>
<td>none</td>
<td>295Ws</td>
<td>0.0s</td>
</tr>
<tr>
<td>A-MQTT-SN</td>
<td>UDP</td>
<td>in A-MQTT-SN</td>
<td>220Ws</td>
<td>0.0s - 2s</td>
</tr>
</tbody>
</table>

Despite the fact that the MQTT-SN is better suited for our application, it consumes more energy than the MQTT protocol for data transmission. Table 2 shows the energy consumed by the GPRS modem when transferring 1000 messages of 20 B length. In the previous work, we have introduced the enhanced A-MQTT-SN protocol which aggregates the messages in order to decrease the power consumption. In Standard Service Profiles, control and measurement station can use the A-MQTT-SN protocol despite the fact that, in the worst case scenario, measurements might have significant delays (e.g. 2 s). In contrast, in urgent situations when prompt data is necessary, control and measurement station may switch to the MQTT-SN which results in smaller data delivery delay, though leads to higher energy consumption.

6 Conclusion

Disaster decision support systems require the holistic approach to urgent computing. It is due to the awareness of the real data acquired from environment for the computations performed by LSCS in case of emergency. The ISMOP project being a typical flood decision support system that illustrates this very well.

The DAPS set-up in case of emergency is rather challenging due to its distributed nature and very limited resources such as CPU power, transmission bandwidth, energy etc. This is rather complex process which has to take into account the existing trade-offs. This is why it opens space for adaptive applications and self-configuration systems rather than imperative centralized control.

Successful operation of holistic urgent computing systems requires suitable design of their architecture and implementation of mechanisms which allow their dynamic reconfiguration according to changing requirements. This is an area where a lot of efforts is still required.

Acknowledgments

The research presented in this paper was partially supported by the National Centre for Research and Development (NCBiR) under Grant No. PBS1/B9/18/2013 and by the Polish Ministry of Science and Higher Education under AGH – University of Science and Technology Grant 11.11.230.124 (statutory project).

References

Holistic approach to urgent computing for flood decision support


