Big Data Solutions on Tactical Infrastructure

Rune Langleite, Eirik Anette Flynn Opland and Frank T. Johnsen
Norwegian Defence Research Establishment (FFI)
Kjeller, Norway

Abstract

In this paper, we discuss the need for explainable artificial intelligence (AI) in defense systems. Further, we elaborate on the need for Big Data solutions to support AI on tactical infrastructure, and discuss an architectural approach to address this need. Finally, we present our proof of concept implementation of this architecture, instantiated to support the human concepts of fast and slow thinking. The proof of concept was built using free and open source software to allow the solution to be shared, and our approach to be repeatable for others.

1 Introduction

The drive towards increased digitalization[1] is pervasive in both the civilian and defense sectors. For defense, digitalization may increase both Information and Communication Technologies (ICT) operations and combat operations efficiency[2] and effectiveness[3]. Situational awareness, *the perception of the elements in the environment within a volume of time and space, the comprehension of their meaning and a projection of their status in the near future* [1], is of paramount importance both for planning and when executing missions. Today, with dynamic situations and need for rapid decisions, the actor with *information superiority*[4] gains the operational advantage [7]. Investigating new approaches to digitalization and novel ICT solutions may facilitate continued information superiority.

Artificial Intelligence, or *AI* for short, is *the science of making computers do things that require intelligence when done by humans* [8]. AI is being used extensively for civilian applications, and in recent years has also seen an increase in military applications as well. Indeed, with the rapidly increasing amount of data, automating analysis tasks becomes more and more necessary to reduce information overload in a defense context [2]. It seems evident that AI may play a major role in the future of situational awareness and maintaining information superiority. One strength of AI is that computers may work and analyze high-volume and high-velocity data much more efficiently than a human operator can.

Kahneman [4] launched the theory that humans have two basic systems of thinking to guide decisions:

1) Thinking Fast
2) Thinking Slow

Here, *System 1* implies an approach that, while being fast, is instinctive, unconscious, imprecise, and sensitive to bias.[5] Conversely, *System 2* implies a slow approach, that is logical, conscious and rational. Indeed, the two human systems are analogous to machine learning and logical reasoning, respectively, and likely the combination of these two approaches will prove necessary to achieve desired properties of an AI-based system [6]. Due to this, we aim for an architecture in our work that can support a combination of fast and slow thinking processes.

We can envision a large number of sensors, both a combination of military and civilian (e.g., Internet of Battlefield Things (IoBT) [9]) in the modern battlefield. The plethora of sensors leads to Big Data, which would overwhelm any human operator. Gartner [3] defines *Big Data* as *high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation*. Indeed, there is a need

1. Note that *digitization* involves converting from analog to digital, whereas *digitalization* is when data from throughout the organization and its assets is processed through advanced digital technologies, which leads to fundamental changes in business processes that can result in new business models and social change [54].

2. Efficient (adj.) — Performing or functioning in the best possible manner with the least waste of time and effort.

3. Effective (adj.) — Adequate to accomplish a purpose; producing the intended or expected result.

4. Information Superiority is the operational advantage derived from the ability to collect, process, and disseminate an uninterrupted flow of information while exploiting or denying an adversary's ability to do the same [7].

5. Bias is the inclination or prejudice of a decision made by an AI system which is for or against one person or group, especially in a way considered to be unfair [61].

for a next generation of Communications Information Systems (CIS) concepts and solutions that can leverage AI-based applications for Command and Control (C2).

Different sensors can support situational awareness and C2 processes in different ways. By investigating a proof of concept solution supporting human thought processes, we contend that this will be a necessary component in future tactical platforms.[6] This motivates our design and implementation, as described throughout this paper. Initially, we pursue setting up a single node as a first step, which can later be part of a network of nodes. The premises we set for our work are that:

- AI is needed in the tactical domain to achieve information superiority through digitalization and automated analysis.
- Digitalization involves speeding up manual processes, and so the AI system should support the two basic systems of thinking (i.e., fast and slow).

Our hypothesis is that *we can realize a technical architecture using open source products that support both basic systems of thinking, and that this may be used in a stand-alone tactical node*.

Working towards realizing this hypothesis, our contribution is a functioning technical solution based on open source software, that can support both basic systems of thinking. Further, we establish the technology requirements for a minimum baseline. Here we identify the minimum needed with respect to computational resources that must be set aside in a tactical node to install such a system.

The remainder of this paper is organized as follows: Section 2 presents related work. The motivation and scope is elaborated on in Section 3. Our design and architectural approach considerations are discussed in Section 4. Section 5 covers our proof of concept implementation. We discuss technology trends in light of our proof of concept in Section 6. Finally, Section 7 concludes the paper.

2 Related work

AI and machine learning are used for an increasing number of applications. Here, we give a few examples of defense applications.

One example is in tactical policy routers, where policies allow assigning a specific portion of the available network capacity do different services. Recent developments with this approach includes aspects of AI and machine learning, e.g., [10] proposes an architectural concept for the use of decentralized, machine learning based reinforcement agents to improve communications in tactical networks. Another example is within Internet of Things (IoT) [25], and its military counterpart, IoBT [9]. The possibilities of IoBT are best exploited through highly automated systems, and there is a need for well-developed analysis modules. Though analysis may be performed by simple statistical means for some applications, e.g., a soldier wearable, in [11] we observe that more advanced AI-based approaches to handle the Big Data would be the preferred approach in large, complex systems.

Explainable AI is a set of processes and methods that allows human users to comprehend and trust the results and output created by machine learning algorithms [12]. Explainable AI is deemed a necessity in defense systems that leverage this technology, since it is needed to establish trust in the system for operators and soldiers, as well as for ethical reasons. This is necessary not only nationally, but also in coalition forces, where *Neuro-Symbolic AI*[7] technology has been identified for addressing explainability and managing trust in data across the coalition [13]. Additionally, Neuro-Symbolic AI can support several different applications like detecting different IED (improvised explosive device) threats [13], and routing convoys to minimize risk in complex attack scenarios [16].

Crowdsensing[8] is much used as a data source for civilian applications, but may also be leveraged for military operations. In segregated networks involving multiple actors, collaborative sensing may prove helpful to fill information gaps on lower levels, i.e., between collaborating nations in a coalition force, or otherwise disconnected units due to hierarchical structures or technical barriers [26]. In Norway, crowdsensing experiments have been conducted as part of a Home Guard field training

6. In this paper, we define tactical platform to be the ICT platform used in the tactical domain, i.e., hardware, software, and the security profile allowing said platform to host services and process classified data.

7. Neuro-symbolic artificial intelligence is a novel area of AI research which seeks to combine traditional rules-based AI approaches with modern deep learning techniques [14]. For defense systems, it is claimed that *neuro-symbolic learning is necessary for complex event processing* [15].

8. Crowdsensing is short for Mobile Crowdsensing (MCS). MCS is a new sensing paradigm that empowers ordinary citizens to contribute data sensed or generated from their mobile devices, aggregates and fuses the data in the cloud for crowd intelligence extraction and people-centric service delivery [59].

exercise, where the soldiers were equipped with Android phones installed with an app for situational awareness [50].

Many tools, both commercial and open source, exist to aid in building AI systems. A full survey of such systems is beyond the scope of this paper, but prominent examples include Apache Hadoop, Apache Spark, and Apache Flink, which are commonly used frameworks for Big Data analysis [55]. Further, distributed message systems like Kafka [56] are often used to enable sensor integration with AI systems. Kafka provides what is called an immutable "log" of incoming data. However, for Big Data processing, a database that can be structured efficiently, read from, written to and also modified, is needed. Classical relational databases may be used, but so-called graph databases are often used instead, due to their superior performance for Big Data handling [57]. There exist many graph databases, ArangoDB, Neo4j, Oracle Spatial and Graph, IBM System G Native Store and OrientDB are a few of them. For our work, we opted for Neo4j [58] because it is easy to learn, and supports the important properties of protecting data integrity while providing fast reads and writes. Neo4j exists in both an open and a closed source version, where we used the open source.

3 Motivation and scope

Technology increases the complexity and pace of operations [60]. For military decision-makers this means that situational awareness as well as a functioning, effective C2 system, will be crucial to emerge victorious from the war of the future. This is already the case today, but in the course of the next 10–20 years, we contend that this will become even clearer. When an opponent is able to leverage units across domains, and employs new and disruptive technologies from both the civilian and military sectors, it becomes challenging to understand what is happening and to know how to respond. The uncertainty that occurs can be exploited and an adversary can create confusion and lead to paralysis of action. To counter this effect, it is important to investigate new communications and computing paradigms.

In a few years, access to sensor information, data and intelligence will increase significantly for the Armed forces. With the phasing in of new platforms, the Armed forces must be able to utilize these to the fullest extent. Hence, data processing and analysis of this information will be very important. The possibilities that have been introduced with new information technology can even today assist with this. The ability to automate multi-source analysis using Big Data and AI can have positive effects over the next years for the operational performance of the Armed forces. The technology will be able to drastically reduce the time needed to analyze complex situations, and enable the access to more sensors to be utilized better.

3.1 Fast/slow "thinking" and the applicability of AI

The appropriateness of computation and AI as analogous to human thinking, has been discussed extensively [5], including theorems about practical limits of computation [45] and extensive arguments about whether digital computers can be conscious, e.g. [47].

According to the Penrose–Lucas argument [46], the human mind must arise from quantum processes, and a revolution in physics is required in order to scientifically explain the human mind. Whether the argument holds, is contentious.

Such deep considerations have their place, but are far beyond the scope of this paper. However, the biological plausibility of the division into fast and slow thinking is worth some attention.

Biological analogies have lead to several important developments in AI, such as neurons and synapses of the brain giving rise to machine learning based on artificial neural networks. Convolutional neural networks (CNN) [48], inspired specifically by the visual cortex, is an example of such inspiration that has been hugely successful in image recognition. The analogy of AI to fast and slow thinking has been made [5], and we aim to identify any practical limitations on the analogy.

Machine learning is an important field of AI, basically covering all systems designed to improve themselves over time. AI is much broader, covering any artificial systems that would require intelligence if the task were performed by humans.

As discussed above, we base our work on System 1 (thinking fast) and System 2 (thinking slow) being analogous respectively to machine learning and logical reasoning. However, these are not absolute correlations. The learning phase of machine learning is usually pretty slow, for example. However, once a model has been trained, exploiting the trained model is often fast. Therefore, creating an online learning system, one that learns continuously as the data arrives, may be slower but also more up-to-date on the latest data and therefore less sensitive to bias. So, in that case, it may be closer to the thinking slow concept. It depends on the problem one is trying to solve.

Some machine learning systems are less sensitive to bias than human agents and their logical approaches. However, once a new machine learning system has been developed, their type of bias is often different than that of a human analyst. For example, a machine learning system may outperform humans on image recognition, successfully distinguishing faces that appear really similar, even to an experienced human eye. However, then it might fail on a case that seems simple to humans. Which type of error is more harmful is not necessarily obvious. However, human errors may be more likely to be socially accepted, because they seem more understandable to humans than machine errors.

This is one reason why explainable AI is important. Somebody needs to be responsible for deploying an intelligent computer system and be able to defend that decision if a failure occurs. Obviously, as a decision maker, one wants to optimize and secure some success criteria. Being able to characterize a level of certainty that the criteria will be met, is a good start. However, one wants to also be able to explain why the criteria will be met with a given level of certainty and in the event of failure, still be able to explain why trusting the computer system was still a good decision in the first place.

Logical reasoning systems have an advantage over many machine learning systems, in that the latter are so-called black box approaches: the way they produce results is not directly explainable, or only explainable at a high level of abstraction. Such black box approaches may still be explainable at some level. For example, it may be possible to prove a certain probability or frequency of failure. It may be sufficient to explain why the probability holds, as opposed to explaining exactly how a system is going to behave, in order to satisfy the need for explainability.

Explainability may come at a cost, because requiring explainability in a certain form, may rule out methods with a higher performance. Therefore, it's important to be conscious of what level of explainability one requires and understand any costs involved. For example, performance statistics may be sufficient evidence that an algorithm works as it should.

The thinking fast and slow paradigm keeps the more explainable methods in the top level[9] decision making, while less explainable methods serve more immediate purposes. In addition to any technical advantages of this approach, it may also be most politically acceptable, because it mimics human behavior. Humans, similarly, often base short term decisions on heuristics and intuition, which, while successful, cannot necessarily be proven rationally to be optimal. We contend that if such failure to give a detailed account of how one makes a decision is accepted of a human, then it may also more easily be accepted by a computer system. Even then, computer systems may be held to a higher standard and so it is worth erring on the side of a more explainable system.

3.2 Use case description

An example of a use case in the tactical domain is illustrated in Figure 1, which shows a joint force comprising a deployed headquarter (HQ) at a fixed location, multiple ships, operational experts (not deployed) and several deployed mobile tactical units (e.g., combat vehicles, drones and dismounted soldiers). The figure also shows the links between the force's main components, including some characteristics and functional area services. Since the tactical domain is characterized as a Disconnected, Intermittent connectivity and Low-bandwidth (DIL) environment [28], tactical links typically lead to issues when systems need to communicate. Due to this, CIS in the tactical domain typically must be specially tailored to cope with the DIL characteristics. This is, of course, a well known problem, and has been studied extensively in NATO research task groups, such as IST-118, which worked towards creating a tactical SOA[10] profile [27]. This group, and its successor IST-150 [29], typically targeted optimizing the communication going across the tactical links, to make the most out of the performance of deployed services. Due to disconnections, units must occasionally work totally disconnected from any deployed infrastructure. This means that one group of units may be able to communicate among

9. By top level, we mean operational level decisions that use aggregated knowledge.

10. SOA, or service-oriented architecture, defines a way to make software components reusable and interoperable via service interfaces. Services use common interface standards and an architectural pattern so they can be rapidly incorporated into new applications. This removes tasks from the application developer who previously redeveloped or duplicated existing functionality or had to know how to connect or provide interoperability with existing functions [44].

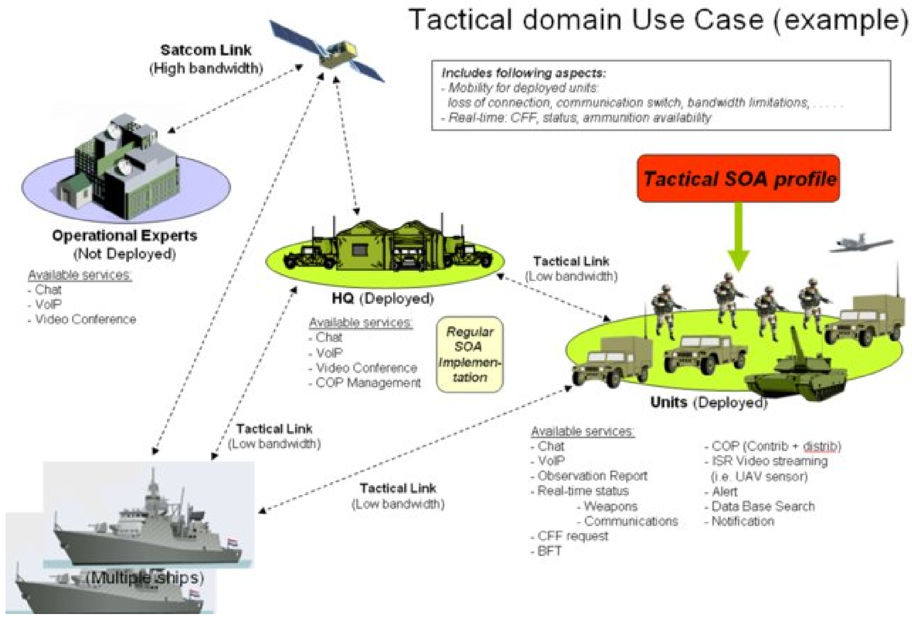

Fig. 1. Tactical domain use case example (from [27])

themselves, but cut off from other groups of units. Or, in the extreme case, that a unit has to rely only on its own resources, being fully cut off from other units. This lack of communications capability may remain for some time, in which case it is a disconnection, or be for a shorter time frame, in which case we do not consider it a full disconnection but rather an intermittent connection. Supporting traditional user-facing services like voice, video, and blue force tracking (BFT) in conditions like these is definitely challenging, even more so in current operations where there is an increasing number of sensors being deployed, which leads to an increase in data needing to be processed, and ultimately shared, among units in the field.

Introducing IoBT, with sensors being deployed in the field, on soldiers, on manned and unmanned vehicles, there comes a large increase in the amount of data that needs to be processed. One approach to optimizing information processing could be to attempt to process as much of this data as possible near where the data originates, and so only allow certain identified "events" to propagate across the tactical network. The analogy here would be the civilian approach of so-called *Edge computing*[11], which we contend will be the main enabler for coping with Big Data in the tactical domain. One can anticipate that such an approach would be beneficial in tactical networks, since the inherent volume of data will likely be too much for the network to handle. As for *cloud computing*[12], one cannot expect to have continuous access to the cloud, from the tactical battlefield. Hence, cloud computing resources must be provided near the tactical edge (e.g., in the deployed HQ, like we did in our previous work on coalition cloud [32]) or processing must be done on the mobile units themselves (e.g., vehicles, soldiers) in the battlefield. In this paper, we consider the latter case, that is, the Edge computing aspect of processing data near/on the unit that needs the information.

3.3 Edge computing tactical node

Pursuing the idea of Edge computing, we aim to support Big Data analysis in a stand-alone tactical node. Our current national tactical platform comes with several different profiles. Considering Figure 1, you would find the most capable profile realized in the deployed HQ. This profile encompasses high performance servers and storage solutions that you would find deployed in the HQ. Our biggest ships can also carry such servers. There is also a less

11. Gartner defines *Edge computing* as *a part of a distributed computing topology in which information processing is located close to the edge—where things and people produce or consume that information* [30].

12. NIST defines *cloud computing* as *a model for enabling ubiquitous, convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction* [31].

capable profile, for servers that can be installed in lighter units, like the manned land vehicles shown in the figure. Finally, the least capable profile targets rugged laptops for field use, which can be brought onboard vehicles or carried by a soldier. For the future, we foresee further automation possibilities, by introducing further CIS diversity. For example, Edge computing could be supported on all tactical units by leveraging small form factor computers, that could be integrated into both soldier equipment and unmanned vehicles. The aim of the prototype we're developing, is to support Big Data analysis on the least capable tactical platform profile.

Glancing at Figure 1 again, we can anticipate the need for a Local Operational Picture (LOP) for each unit. Further, for each group of units, there needs to be a common understanding, they need the Group Operational Picture (GOP). Finally, the HQ builds the Common Operational Picture (COP), across the units and groups of units that are part of the operation. Here, we understand that the amount of data that needs to go into a COP is larger than that of a GOP or LOP. Due to the DIL characteristics of tactical networks, sensors capacities may come and go in the network. This means that being able to process data near the source makes a lot of sense to optimize communications capabilities. Further, data can be turned into information at the edge, and so reducing complexity by only transmitting valuable information to other units and higher levels. We aim to support digitalization of these processes by using AI.

Processing Big Data requires both CPU and RAM, and the higher the data volume is, the larger the requirements for these resources become. For the sake of this paper, we do not consider the most capable servers that would be able to use traditional approaches to Big Data handling, like MapReduce. MapReduce has been shown to work in "tactical clouds" [35], and so this approach would be suitable in, e.g., a HQ or onboard a ship where you find the most capable tactical platforms. However, for the least capable tactical profile, we need a less resource demanding approach. Hence, we pursue a *Lambda architecture* in this paper, since it has been shown to be a cost-effective approach to supporting both batch and speed layer processing [36]. The Lambda architecture is presented in the following section.

4 Architectural approach

4.1 Lambda architecture

As described in [24], the Lambda architecture specifies two parallel data processing pipelines, called the *batch* and *speed* layers, which produce views that should answer questions pertaining to the business logic, made accessible through the *serving* layer. The batch views are computed using all existing data (consistent and complete, high latency), while the speed view is computed using current data only (low latency, lacking in consistency and completeness). Drawing the parallel to the two human basic systems of thinking, this means that the *speed* layer implements *fast* thinking. Conversely, the *batch* layer represents *slow* thinking. A high-level illustration of the Lambda architecture is shown in Figure 2.

4.2 Software component stack

Following the recommendations from [23], we want to build knowledge graphs which represents the LOP for a single unit. Thus, we chose to utilize a graph database which naturally translates entities from the real world to a graph structure. In this particular context, the root node is the unit itself identified by its callsign, which has a relationship to two logical sets of views, namely batch and speed, further separated by the type of data (e.g., weather information).

The message queue layer should be as easy as possible for any given sensor to communicate with, as they should be loosely connected to the data processing layer of the Big Data infrastructure. This is due to the inherently complex and challenging environment battlefield sensors reside in, and they could potentially be leaving and joining the network at random.

At the processing layer, which in this case encapsulates both the speed and batch layers, resource consumption is a restriction in itself, as outlined in Section 3.2. Simultaneously, processing speed is of utmost importance in order to achieve timely delivery of potentially mission-critical data. The speed layer must react within seconds, whereas the batch layer may digest data for hours.

Finally, the serving layer needs only be demonstrative for this project, as integration with existing systems such as analysis tools and Battle Management Systems (BMS'es) is a task for future work. Thus, we chose to build a simple RESTful web API to fulfill this purpose.

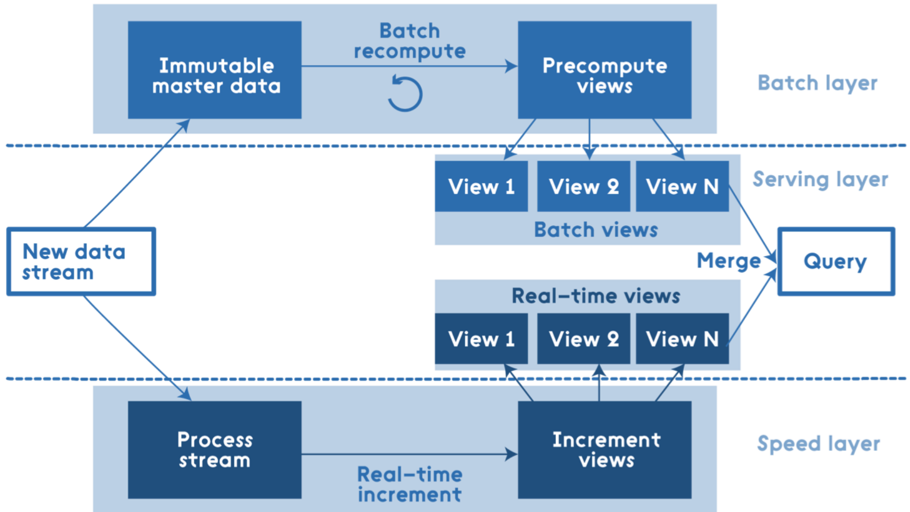

Fig. 2. Lambda architecture high-level perspective (from [37])

The prototype Big Data implementation was built using the following open source software components:

- Kafka [17] server, a distributed event-processing platform
- Zookeeper [18], Kafka dependency for service synchronization and naming registry
- Spark [20] Master and Worker nodes, data processing cluster
- Neo4j [21] graph database, storage for view calculations

Kafka was chosen based on its flexibility to support both streaming and more classic publish/subscribe patterns. In addition, the data ingested to the architecture is stored in Kafka's immutable, append-only log, which served as the immutable master data.

Furthermore, Spark was chosen to serve as the processing engine in both the speed- and batch layers, as it can support both paradigms using the native Kafka connector for data ingestion and proper offset configurations [43]. Using Spark for both paradigms implies that the two AI components realizing speed and batch layer functionality, respectively, will both submit their jobs to the same Spark cluster.

Finally, Neo4j was chosen to serve as the storage component for views produced by both processing layers, which could be made queryable using a number of tools. In addition, Spark can be easily configured to ingest data to Neo4j using its Spark connector [33], further simplifying development of the data pipeline.

5 Implementation

The implementation was built with the open source software components outlined above, where we used simple weather sensors to extract, process, and store data for the purpose of establishing a proof of concept. During execution, we collected some run-time metrics which we provide later in this paper, establishing the minimum baseline needed to deploy this software stack on a computer.

5.1 Proof of concept

The proposed solution architecture is shown in Figure 3.

5.1.1 Sensors

The sensors are, for the proof of concept, diverse IoT weather sensors which pertain to the unit's LOP with respect to the local environment. Based on the Raspberry Pi 3, we have added sensor hardware that each provides a stream of JSON-encoded weather data. This data is published from the sensor to the message queue layer, by publishing the data on a Kafka topic. The data is processed in the batch and speed layers, and eventually the information is included in the LOP. The sensors we have are

- Enviro pHAT [51]
- Pioneer 600 [52]
- Sense HAT [53]

An example JSON-encoded reading from one of these sensors is given in Listing 1.

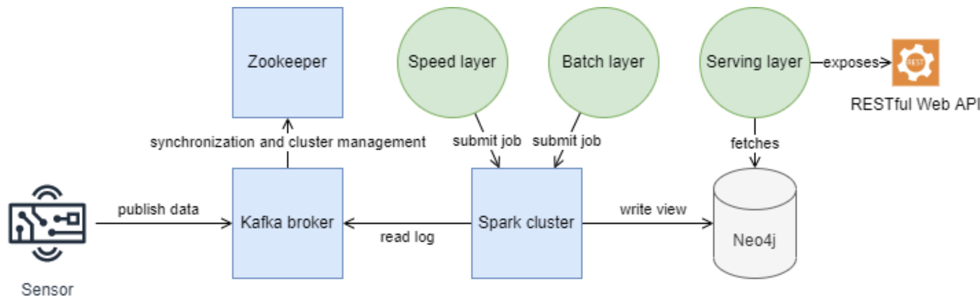

Fig. 3. Technical solution architecture

Listing 1. Pioneer600 sensor reading

```
{'host': 'pioneer', 'sensor': 'Pioneer600',
'ts': '2022-05-18T11:31:23.118751',
'temperature': 31.6, 'pressure': 1008.27,
'altitude': 41.03}
```

In Listing 1, the "host" identifies the node on which the sensor is mounted, here simply called "pioneer". Next, the sensor type is identified, in this case it is the "Pioneer600". The sensor reading has an associated ISO standard timestamp ("ts") showing the time and date of this particular measurement. Following this, the measurement values are included, here the "temperature" (Celsius), the "pressure" (in hectopascal), and finally the "altitude" (in meters). Different sensors give different readings, for example, the Enviro pHAT can also give values for the light level (in lux), whereas the Sense HAT provides measures of relative humidity. Relative humidity is expressed as a percentage. For any given air temperature there is a maximum amount of water vapour that it can suspend. Relative humidity is the percentage of actual water vapour present, compared to the maximum possible amount.

These data are all fed into our tactical node via Kafka, where the data may be further used by the batch and speed layers.

5.1.2 *Batch and speed layers*

At the message queue layer, Kafka and Zookeeper create the entry point and storage. By enabling Kafka to keep data forever by configuring log retention, we can access all producer data from the beginning by using the earliest offset. This is realized in the batch layer where, following the principle of conducting re-computation on the entire data set, all data on a given topic is considered when computing the batch view. Similarly, the "latest" offset configuration is used in the speed layer to produce the speed view. This approach also removes the need for an additional storage layer, by utilizing the data already existing in the Kafka log.

At the speed and batch layer, two separate Java applications connect to the Kafka broker by using Spark's native Kafka connector and the above-mentioned offset configuration to extract data for the processing logic in the two layers. As we are only working with simple weather data, the batch layer calculates an average value based on all existing data, while the speed layer simply extracts the current temperature value. Only statistical functions and no trained models were utilized at this point in time. However, for future work, we want to extend the logic in both layers with more advanced AI-capabilities for the purpose of for instance anomaly detection and predictions.

Both layers write to a local instance of Neo4j graph database, which uses the structure illustrated in Figure 4. The blue node represents the local unit on which a LOP is being generated. For the LOP, a set of nodes representing a given named view, further segregated by layer logic, are created. In this example, views named "LOP_weather" are generated with both a batch- and a speed view label, annotated by the red color. Finally, the actual view data is represented by data nodes, which each hold a descriptive name, the current value, and a timestamp for when the view was last calculated.

5.1.3 *Serving layer*

Finally, at the serving layer, a simple RESTful web API was built using Jersey [22] which queries the graph database using Cypher [34] and outputs the value of the above-mentioned data node, as shown in Listing 2.

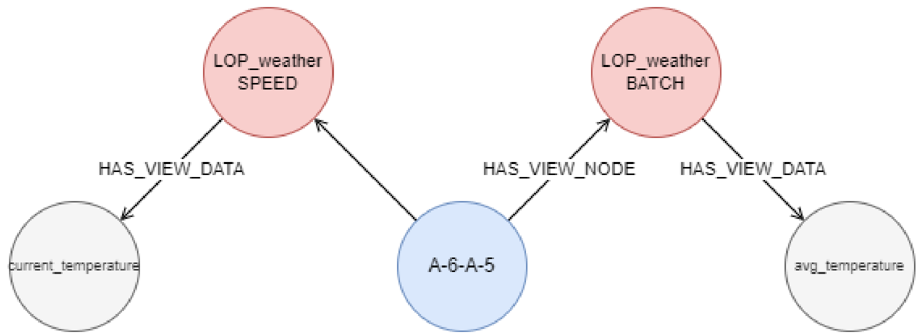

Fig. 4. Graph database structure

Listing 2. Serving layer output using curl

```
curl -XGET http://localhost:8888/weather | jq
{
  "viewDataSet": [
    {
      "name": "vd.updated_timestamp",
      "value": "2022-05-30T08:39:07.132"
    },
    {
      "name": "vd.value",
      "value": 31.88495635986328
    }
  ]
}
```

5.2 Open source products resource use

To establish the baseline resource use, we measured how the different software components used in our prototype impacted the node. The node was set up with Ubuntu 20 LTS using a standard desktop computer in our lab:

- HP ELITEDESK 800 G3 SFF
- 16GB RAM
- Intel i7-6700 CPU @ 3.40GHz (8 CPU cores)
- 1TB hard disk drive

On this node, we instantiated 1x Zookeper, 1x Kafka, 1x Spark with 3x Spark workers, and 1x Neo4j. The software used 3.5GB of the total RAM, 11GB was still available, and the remainder was used by the operating system and supporting libraries. A deeper analysis of each component, with individual (idle) CPU and RAM usage, is shown in Table 1.

From these findings, we can see that the baseline can easily be accommodated by common off-the-shelf hardware. The desktop used has, by today's standards, modest specifications and performance, and so the same capabilities can easily be found in rugged laptops suitable for use in the field. Hence, we deduce that this software stack is definitely viable at the tactical edge, and can be accommodated even in the least capable of our future tactical platform nodes. Note that we do not list any data of the system under load here, where naturally resource use will increase. The specific resource use will depend on the AI model supported, the nature of connected sensors and volume of sensor data being analyzed. However, our findings do support our hypothesis that *we can realize a technical architecture using open source products that support both basic systems of thinking and may be used in a stand-alone tactical node*. For future work, we foresee testing many different sensor and application types, where such a deeper analysis could be performed.

Component name	CPU usage	RAM usage	Of RAM total
Kafka	1.10%	356.7MB	2.25%
Zookeeper	0.16%	117.0MB	0.74%
Spark	0.13%	200.7MB	1.26%
Spark-worker-1	0.14%	195.1MB	1.23%
Spark-worker-2	0.13%	204.3MB	1.29%
Spark-worker-3	0.14%	202.3MB	1.27%
Neo4j	3.39%	1.427GB	9.20%

TABLE 1
Resource use when the entire system is up, but idle.

6 Discussion

We provide some insights into how we foresee the bigger picture of AI in conjunction with other technologies for the future. Following that, we discuss limitations in our current approach, and how we may develop the proof of concept further.

6.1 AI and other technology developments

Trend analyzes from NATO [2] (and also the research community; both universities and defense institutes) are relatively consistent in identifying technology areas that will be of great and increasing importance for military operations in the future. These are the technology areas where progress is constantly being made, but where maturity and real-world applications vary today. However, they have in common that they are all technologies with great potential added value for the Armed forces and social security.

Developments in advanced electronics and computing are important because almost all platforms, systems and services contain a programmable element. Quantum computers are an immature technology, and there is a long way to go before military exploitation. There is a lot of research in the field because there is a great deal of civilian demand, since a breakthrough could have a major impact on complex and hyper-fast data processing. For military purposes, quantum computers combined with AI have the potential to process vast amounts of data and thus solve problems that today appear to be unsolvable, such as breaking the crypto solutions that are considered secure today [38].

The next generation of sensor technology will be able to detect signals with increasing sensitivity, detect more signal types and thus improve the overall sensing abilities. New sensors will contribute data sets from IoT which increase situational awareness. Much of the development takes place in the civilian sector, but some areas will be reserved for military development [39]. The military should focus on adapting and integrating the sensors in platforms [40]. Due to the commercial drive behind IoT, the sensors are becoming lighter and cheaper, and the use of modules and open architecture makes it easier to change sensors as new technology becomes available.

AI, machine learning and Big Data are in a rapid development for a plethora of civilian applications, and this development is also applicable to and affects the defense sector. AI, as we have mentioned earlier in this paper, is about getting machines to perform tasks that normally require human intelligence, such as interpreting speech, translating between languages or recognizing objects in images. The progress in AI is mainly due to the fact that we have gained access to a lot of good data that the machines can train on, as well as fast enough machines. Among other things, this will help to establish a superior understanding of the situation, and provide a more robust basis for quick and good decisions.

Increasing investments are made in autonomous systems, both in the civilian and military sectors. Technology is moving in the direction of concepts where autonomous systems and humans complement each other [49]. Military applications of autonomous systems include surveillance, reconnaissance, transport tasks, logistics, but also swarms of armed, unmanned systems. The technology implies that demanding tasks can be solved faster, better and with lower risk when using unmanned systems. Developments in AI and autonomous systems, however, raise legal and ethical questions about how far the systems should be allowed to make decisions themselves [41].

When different technologies work together and reinforce each other, it can fundamentally change future operations. An example of a grouping of technology areas that we contend will have such an effect in the future is AI, Big Data, autonomy and sensors. In the interplay between these technologies, there is a significant potential for innovation that may radically change the future of military operations. Such advances may have great added value for a range of defense activities, from optimizing the performance of military equipment, reducing costs, and improving ways of conducting military operations.

The proof of concept implementation we have described in the previous section, supports this vision for the future in that we have investigated one initial building block of the future battlefield. Investigating the technology that needs to go into one tactical node, to enable Big Data analysis. By having this functionality in the tactical domain, the data can be processed where it occurs. Thus, we may say that Edge computing is a technology approach that supports the vision of an edge organization.[13]

6.2 State of the proof of concept

Recall that our aim in this paper was to address a stand-alone tactical node. Although the solution proposed in this paper shows promising results, further discussion and development is required in order to provide an operational Big Data solution at the tactical edge, which ultimately should provide decision makers with the necessary information to improve the operational efficiency of own forces. For instance, a complete overview of what kind of data that can be acquired through sensing devices that is of interest for military commanders at all levels is currently lacking.

In addition, the time of delivery requirements for such intelligence is also not specifically set, and it can be argued that various data may have varying time of delivery requirements. For instance,

13. Edge organizations are organizations where everyone is empowered by information and has the freedom to do what makes sense [42].

detection of Chemical, Biological, Radiological, Nuclear (CBRN)-agents should be alerted immediately, while we may tolerate some delay for weather data. Another application would be Radio Frequency (RF) spectrum data analysis, which is another use of AI that we think would be beneficial for future work.

Furthermore, not all data may be suitable for processing using the technology stack used in this paper, which either requires an additional layer for source-specific data ingestion, processing, and storage, or an additional interface which could convert the raw data into something that can be understood by the components used in this project.

It should also be emphasized that a tactical infrastructure such as the one proposed in this paper is not the sole source of intelligence. Rather, it is a supplement to existing procedures and capabilities already present in modern Armed forces. To succeed in becoming an integral part of military operations, however, further development involving distributed processing and autonomous information exchange for the purpose of building GOP and COP in disadvantaged networks and IT infrastructure is required. In addition, it can be argued that such a system must be as autonomous and user friendly as possible, requiring minimal interaction of users operating at the front lines.

7 Conclusion

7.1 Summary

When the Armed forces can establish situational awareness quickly and with high credibility, this will lead to a need for decisions to be made at the same pace. This means that explainable AI will be needed, both to automate analysis of the information flow, and to support the decision making. In a complex, multi-domain battlespace, the analysis needs to be distributed and the resulting information products need to be disseminated efficiently. In this paper, we have started the journey towards such capabilities, by investigating open source software for supporting Big Data analysis at the tactical edge. The continued research of robust communication and secure ICT solutions will therefore remain very important in parallel with surveying and leveraging the long-term technological development and with the possible introduction of new disruptive capabilities.

7.2 Future work

For future work, we aim to investigate further applications of the technology stack we set up. We discussed some of these applications in this paper, which include CBRN sensor data and RF sensor data. We also plan to expand the proof of concept by investigating a network of collaborating tactical AI nodes. The idea is to leverage parallel data processing and also introduce high availability of the AI capability in the tactical domain. There is also the federation aspect to explore, how this AI capability can be used in a network involving multi-national forces and collective C2.

References

[1] Neville A. Stanton, Peter R.G. Chambers and John Piggott, "Situational awareness and safety", In: Safety science 39.3, pp. 189–204, 2001.

[2] D. Reding and J. Eaton, "Science & technology trends 2020-2040", NATO Science & Technology Organization, Office of the Chief Scientist, Brussels, Belgium, 2020.

[3] Gartner. "IT Glossary — Big Data". https://www.gartner.com/it-glossary/big-data (accessed 18 January 2021)

[4] Daniel Kahneman, "Thinking, fast and slow", New York: Farrar, Straus and Giroux, 2011.

[5] G. Booch, "Thinking Fast and Slow in AI", Proceedings of the AAAI Conference on Artificial Intelligence 2021, 35(17), 15042-15046

[6] Lex Fridman, "Daniel Kahneman: Thinking Fast and Slow, Deep Learning, and AI", https://lexfridman.com/daniel-kahneman/ (accessed 18 January 2021)

[7] D. S. Alberts, J. J. Garstka, and F. P. Stein, "Network Centric Warfare — Developing and Leveraging Information Superiority", ISBN-10:9385505777, ISBN-13:978-9385505775, January 2016.

[8] J. Copeland, "What is Artificial Intelligence?", http://www.alanturing.net/turing_archive/pages/Reference\%20Articles/What\%20is\%20AI.html, (accessed 10 June 2022)

[9] S. Russell and T. Abdelzaher, "The Internet of Battlefield Things: The Next Generation of Command, Control, Communications and Intelligence (C3I) Decision-Making", 2018 IEEE Military Communications Conference (MILCOM), 2018, doi: 10.1109/MILCOM.2018.8599853

[10] T. Möhlenhof, N. Jansen and W. Rachid, "Reinforcement Learning Environment for Tactical Networks", 2021 International Conference on Military Communication and Information Systems (ICMCIS) (2021): 1-8.

[11] R. Langleite et al., "Military Applications of Internet of Things: Operational Concerns Explored in Context of a Prototype Wearable", ICCRTS 2021 (virtual)

[12] IBM, "What is explainable AI?", https://www.ibm.com/watson/explainable-ai (accessed 19 January 2022)

[13] A.D. Preece et al., "Coalition situational understanding via explainable neuro-symbolic reasoning and learning", in 'Artificial Intelligence and Machine Learning for Multi-Domain Operations Applications III', Vol. 11746, International Society for Optics and Photonics, 2021.

[14] Zachary Susskind, Bryce Arden, Lizy K. John, Patrick Stockton and Eugene B. John, "Neuro-Symbolic AI: An Emerging Class of AI Workloads and their Characterization", https://arxiv.org/abs/2109.06133v1, 2021. (accessed 19 January 2022)

[15] U.S. Army DEVCOM Army Research Laboratory Public Affairs, "AI research helps soldiers navigate complex situations", https://www.army.mil/article/240917/ai_research_helps_soldiers_navigate_complex_situations, 2020 (accessed 19 January 2022)

[16] T. Xing et al., "DeepCEP: Deep complex event processing using distributed multimodal information", in '2019 IEEE International Conference on Smart Computing (SMARTCOMP)', IEEE, pp. 87–92., 2019.

[17] Apache Foundation, "Kafka", https://kafka.apache.org/ (accessed 3 June 2022)

[18] Apache Foundation, "Zookeeper", https://zookeeper.apache.org/ (accessed 3 June 2022)

[19] Confluent Inc., "Schema Registry", https://docs.confluent.io/platform/current/schema-registry/index.html (accessed 3 June 2022)

[20] Apache Foundation, "Spark", https://spark.apache.org/ (accessed 3 June 2022)

[21] Neo4j Inc., "Neo4j", https://neo4j.com/ (accessed 3 June 2022)

[22] Eclipse Foundation, "Eclipse Jersey", https://eclipse-ee4j.github.io/jersey/ (accessed 3 June 2022)

[23] B. Hansen et. al., "Stordata og avansert analyse", FFI-Rapport 21/02647, 2022 (In Norwegian)

[24] J. Warren and N. Marz, "Big Data: Principles and best practices of scalable realtime data systems", 2015

[25] P. Fraga-Lamas et al., "A review on internet of things for defense and public safety", In: Sensors 16.10, p. 1644. ISSN: 1424-8220, Oct. 2016.

[26] M. Pradhan et al., "Leveraging Crowdsourcing and Crowdsensing Data for HADR Operations in a Smart City Environment", In: IEEE Internet of Things Magazine 2.2, pp. 26–31, ISSN: 2576-3180., 2019.

[27] Christoph Barz et al., "IST-118 SOA Recommendations for Disadvantaged Grids in the Tactical Domain - SOA Experiments on Wireless Broadband Mobile Networks in the Tactical Domain", International Command and Control Research and Technology Symposium (ICCRTS), CCRP publication, USA, June, 2015.

[28] Keith L. Scott et al., "Robust communications for disconnected, intermittent, low-bandwidth (DIL) environments", 2011 Military Communications Conference (MILCOM), pp. 1009–1014, 2011.

[29] Norman Jansen et al., "NATO Core Services Profiling for Hybrid Tactical Networks", STO TECHNICAL REPORT. Published March 2021. ISBN 978-92-837-2328-8

[30] Gartner. "IT Glossary — Edge Computing", https://www.gartner.com/en/information-technology/glossary/edge-computing (accessed 3 June 2022)

[31] NIST, "The NIST Definition of Cloud Computing", Special Publication 800-145, September 2011.

[32] Hallvard M. Andersen et al., "NATO Federated Coalition Cloud with Kubernetes: A National Prototype Perspective", ICCRTS 2021 (virtual)

[33] Neo4j Inc., "Neo4j Connector for Apache Spark", https://neo4j.com/docs/spark/current/ (accessed 6 June 2022)

[34] Neo4j Inc., "Cypher Query Language", https://neo4j.com/developer/cypher/ (accessed 6 June 2022)

[35] J. George et al., "Hadoop MapReduce for Tactical Clouds", 2014 IEEE 3rd International Conference on Cloud Networking (CloudNet), 2014, pp. 320-326, doi: 10.1109/CloudNet.2014.6969015.

[36] M. Kiran et al., "Lambda Architecture for Cost-effective Batch and Speed Big Data processing", IEEE International Conference on Big Data, 2015.

[37] T. Matyashovsky, "DZone - Big Data Zone - Lambda Architecture with Apache Spark", https://dzone.com/articles/lambda-architecture-with-apache-spark (accessed 3 June 2022)

[38] M. Strand, "A Status Update on Quantum Safe Cryptography", 2021 International Conference on Military Communication and Information Systems (ICMCIS), 2021, pp. 1-7, doi: 10.1109/ICMCIS52405.2021.9486413.

[39] M. Pradhan, "Interoperability for Disaster Relief Operations in Smart City Environments", PhD Thesis, University of Oslo, Faculty of Mathematics and Natural Sciences, Department of Technology Systems, April 2021.

[40] M. Pradhan et al., "Concepts and Directions for Future IoT and C2 Interoperability", IEEE MILCOM 2021.

[41] M.K. Orten, "Machine Ethics for Autonomous systems", Master's thesis, University of Oslo, Faculty of mathematics and natural sciences, Department of Informatics, May 2021.

[42] D.S. Alberts and R.E. Hayes, "Power to the Edge: Command and Control in the Information Age", CCRP Publication Series, 2003, ISBN 1-893723-13-5.

[43] Apache Foundation, "Spark Structured Streaming – Kafka Integration Guide", https://spark.apache.org/docs/latest/structured-streaming-kafka-integration.html (accessed 6 June 2022)

[44] IBM Education, "SOA (Service-Oriented Architecture)," https://www.ibm.com/cloud/learn/soa (accessed 2022-06-05), published 2021-04-07.

[45] A. M. Turing, "On Computable Numbers, with an Application to the Entscheidungsproblem", 1937, https://doi.org/10.1112/plms/s2-42.1.230

[46] R. Penrose, "The Emperor's New Mind: Concerning Computers, Minds and The Laws of Physics", Oxford University Press, 1989, ISBN 0-19-851973-7.

[47] J. Searle, "Is the brain's mind a computer program?", Scientific American, 1990, https://www.cs.princeton.edu/courses/archive/spr06/cos116/Is_The_Brains_Mind_A_Computer_Program.pdf.

[48] IBM Education, "Convolutional Neural Networks", https://www.ibm.com/cloud/learn/convolutional-neural-networks (accessed 2022-06-05), published 20 2020-10-20.

[49] L. Burski et al., "Human-Autonomy Teaming: Supporting Dynamically Adjustable Collaboration", Technical report of Task Group HFM-247, March 2020, 10.14339/STO-TR-HFM-247.

[50] F.T. Johnsen et al., "SMART II: Android apps, cloud computing and mobile device management as enablers for efficient operations", 24th International Command and Control Research and Technology Symposium (ICCRTS), October 29-31 2019, Laurel, Maryland, USA.

[51] Pimoroni, "Enviro pHAT", https://github.com/pimoroni/enviro-phat (accessed 2022-06-07)

[52] Waveshare Wiki, "Pioneer600", https://www.waveshare.com/wiki/Pioneer600 (accessed 2022-06-07)

[53] Raspberry Pi, "Sense HAT", https://www.raspberrypi.com/products/sense-hat/ (accessed 2022-06-07)

[54] J. Prause, "Digitization vs digitalization", https://www.sap.com/insights/digitization-vs-digitalization.html (accessed 2022-06-13)

[55] B. Akil, Y. Zhou, and U. Rohm, "On the usability of Hadoop MapReduce, Apache Spark & Apache Flink for data science", 2017 IEEE International Conference on Big Data (Big Data), December 2017, 10.1109/BigData.2017.8257938.

[56] J. Kreps, N. Narkhede, J. Rao, "Kafka: a Distributed Mes-

saging System for Log Processing", NetDB workshop ’11, 2011

[57] M.A. Rodriguez and P. Neubauer, "The Graph Traversal Pattern", chapter in Graph Data Management: Techniques and Applications, eds. S. Sakr, E. Pardede, 2011

[58] F. Lopez and E. Cruz, "Literature review about Neo4j graph database as a feasible alternative for replacing RDBMS", Industrial Data, 2015, 18. 135. 10.15381/idata.v18i2.12106.

[59] B. Guo, Z. Yu, X. Zhou, and D. Zhang, "From participatory sensing to mobile crowd sensing", in IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), Mar 2014, pp. 593–598.

[60] D.C. Billing, G.R. Fordy, K.E. Friedl and H. Hasselstrøm, "The implications of emerging technology on military human performance research priorities", Journal of Science and Medicine in Sport, Volume 24, Issue 10, October 2021, Pages 947-953

[61] E. Ntoutsi et al., "Bias in data-driven artificial intelligence systems—An introductory survey", WIREs Data Mining and Knowledge Discovery, Volume10, Issue3, May/June 2020