OFFPRINT FROM *STRATEGIC OUTLOOK 8*

How we can protect Sweden's security-sensitive IT services

Anders Elfving and Anton Dahlmark, Fortifications Agency

Protecting vital societal functions is a significant element in the development of the Swedish total defence capability. This is why there is every reason to review the authorities' overall need for physically protected data centres used for security-sensitive IT operations. However, stringent demands in terms of physical protection against weapons effects and IT environment supply system uptime, for example, increase costs and extend lead times. Rebuilding existing facilities versus building new ones must also be weighed up. The changing global political and military dynamics and our insight into the vulnerability of our digitalised society mean that a national initiative in this field is a significant but necessary investment for society and its total defence.

Rapid digitalisation advances result in vulnerabilities

How the digitalisation of our day-to-day lives has made everything simpler and changed our behaviour can hardly have escaped anyone's attention. Large volumes of data generated are frequently processed in cloud services and stored in data centres over time. The fact that these services continue to work 24 hours a day, with no disruptions, is generally taken for granted. The debate regarding our increased vulnerability on account of the digital revolution is ongoing, and there is increasing insight into the potential gravity of the consequences in the event of any issues such as attacks or power failures.

While the topic of IT security is now high on Swedish authorities' agendas, physical protection of governmental IT services has also been identified as a crucial area for review if we are to strengthen the total defence. In 2017, the Swedish Post and Telecom Authority was commissioned by the government to prepare a proposal for a national management model for protected data centres. By no means all data requires enhanced physical protection, but some data is security-sensitive and needs to be protected to prevent both intrusion and military attacks.

Existing or new facilities for data centres?

Secure data centres may be located in protected underground facilities (in rock caverns) or secure buildings above ground. Protected underground facilities have the physical barriers to protect against weapons effects, but they also protect the data centre's functions and supply systems by offering what is known as fortified protection. Secure buildings above ground are designed with security-enhancing features that prevent or hinder damage to their functions, but they do not have the same fortified protection as underground facilities. The prior period of global stability and cuts in defence expenditure have resulted in protected properties such as underground facilities being taken out of service. This is why authorities in recent years have been asking whether these could be used for protected data centres, or whether it would be more appropriate to build new protected facilities.

One common perception is that implementation of protected data centres in underground facilities is straightforward. The usual arguments claim:

- That a large number of empty underground facilities are available that could be used as data centres following minor refurbishments – and that this would be relatively inexpensive to implement.
- That underground facilities automatically provide protection against weapons effects of all levels and types.
- That while retaining fortified protection, underground facilities can easily be adapted for data centres consuming ten megawatts (MW)

FOI Memo 6871
Project Nr I181103
Approved by Mike Winnerstig

October 2019

or more – equivalent to the power produced by about six wind turbines or the power required to run more than 4000 homes.

- That underground facilities are within 15 to 20 km of the geographical locations of current operations. That data centres can be established quickly at existing underground facilities.

Besides the above expectations, high uptime levels are frequently required as well. In other words, how much of the time facilities are operational and delivering the intended capability is also a factor to consider. Uptime is generally higher with redundant systems, but these are often more costly than anticipated.

However, the actual situation is not quite the same. Few of the underground facilities available would be suitable for use as data centres, and those that do exist are rarely located close to population centres. There is frequently a significant need for decontamination and extensive investment before the facility can be commissioned. Maintaining fortified protection while also devising a solution for the necessary cooling of the IT environment presents a major challenge. Although the rock has mostly been removed already, adaptation work takes longer than anticipated. However, the procurement time is considerably shorter when refurbishing a vacant underground facility, compared with constructing a new facility.

When building from scratch, the fact that the facility is designed for use as a data centre right from the outset is an advantage. There are economic benefits to be derived from coordinating physical protection, construction costs and operating costs when allowing multiple social stakeholders to share a single, physically protected facility. However, there are also negative aspects to sharing facilities. If only a small number of facilities are established as a result, there is a risk of them being viewed as more high-value targets from an attacker's perspective, compared with a large number of attack targets over a wide area. There is therefore a risk of more far-reaching consequences of an attack on a high-value target.

Power supply with no failures

The need for a robust power supply is another aspect to take into account. Data of significant importance to society must be stored and managed without power failures. IT services are also power-hungry applications requiring an efficient cooling infrastructure. Without cooling, IT equipment overheats – sometimes within minutes – and disables the facility's functions. The large amounts of energy that need to be dissipated from data centres mean that water cooling is deemed to be far more efficient than air cooling or geothermal cooling. It is therefore appropriate to select physical locations adjacent to large reservoirs or watercourses, which of course limits the number of potential locations.

Society is currently making a transition to fossil-free energy. At the same time, Svenska kraftnät[1] indicates that the need for auxiliary power supplies is increasing. A reliable, dependable auxiliary power supply is absolutely essential in a number of sectors of society that are crucial to maintaining a functional total defence: county councils, municipalities and voluntary organisations, for instance. Auxiliary power supplies at present usually involve diesel power stations, but these have a number of limitations: (i) major environmental impact, (ii) problems with fuel distribution during crises, (iii) dependency on imports from other countries, (iv) high thermal signature during combustion, making facilities easier to detect, and (v) noise.

All in all, therefore, it is necessary to test alternative new energy solutions to provide reliable auxiliary power supplies at vital societal facilities such as future data centres. Battery and fuel cell technologies are examples of areas where recent development has shown promising results from a robust societal perspective. For protected data centres, it is particularly important to ensure that the auxiliary power supplies of the future are not only robust, but also easy to maintain and inexpensive to run. Moreover, power and cooling supply intakes must be protected from the pressure waves caused by bombing attacks, threats from electromagnetic pulse and high-power microwave attacks and other threats. This may present a challenge, however, as these intakes often have to cover large areas.

Varying uptime and security requirements

The basic functioning of society is rarely dependent on a single stakeholder's ability to provide a service under difficult conditions. Electricity, data and telecommunications, financial services, transport, fuel distribution, food supply – everything is interlinked in various intricate chains of dependency. If the function offered by a social stakeholder fails – if an IT service is disabled, for instance – this may impact on all parties who are dependent on this service in their turn. Society would benefit from maintaining a holistic approach

[1] Svenska kraftnät is a Swedish state-owned electricity transmission system operator.

with regard to the uptime of individual subfunctions, along with selected levels of protection and the extent to which they merit protection.

Table 1. Assumed need for uptime and requirements for various IT services at different times

	Peacetime	Crisis	War
High uptime	Greater need	Average need	Reduced need
Secure building	Greater need	Greater need	Average need
Protected facility	Reduced need	Reduced need	Reduced need

Demands for higher levels of protection and uptime are very much cost-driven. The cost of a data centre increases rapidly depending on the level of uptime, potentially resulting in a highly costly undertaking. It is reasonable to assume that most vital societal IT facilities will have varying demands in terms of uptime and security, and that these may vary during peacetime, times of crisis and war. A more in-depth analysis is of course necessary, but a likely scenario is that a large number of IT services will have high uptime requirements in peacetime and considerably reduced requirements in wartime, when only the most essential functions are expected to be operational. In peacetime, the number of IT services requiring secure buildings above ground is likely to be considerably higher than the number of IT services requiring protected underground facilities. Table 1 shows a possible simplified description of this scenario.

According to what in Sweden is known as the responsibility principle for the crisis management system, individual stakeholders such as authorities have the same responsibilities in wartime as in peacetime and make their own decisions on what protection and uptime they need. Issues relating to which stakeholders' activities should constitute protected entities or issues of national interest, which physical threat levels should be addressed and which uptime levels must be achieved by each individual subfunction should benefit from being managed at an overall societal level. The Swedish Post and Telecom Authority's proposal for a management model in respect of protected data centres may present a useful starting point. This proposes the following:

- **Priority function.** A governmental function tasked with classifying and prioritising protection needs for the IT services of security-sensitive operations from an overall societal perspective.
- **Facility administrator.** An organisation that operates on the basis of the proposed administration model to manage the portfolio of protected data centres.
- **Facility owner.** An organisation that owns and manages the facilities that have protected data centres.
- **Occupants.** Practitioners running security-sensitive operations that need to house all or parts of their IT environments in protected data centres.

How could a governmental data centre concept be structured?

Protected facilities with a high useful power output – that is, a power output that is useful for the facility's functions – are expensive and take a long time to build. That said, facilities with high levels of fortified protection are necessary from a total defence perspective. One possible way of balancing the ratio of usage to risk and cost for a protected facility would be to implement less stringent requirements in terms of high useful power output. The outcome would be a less complex design at a lower cost. This would also accelerate the procurement process. Moreover, it may also make it easier to close in on the environmental quality objectives, as well as improving sales of surplus heat.

In most respects, secure buildings should be constructed so that they are as similar as possible to modern commercial data centres. Principles and experiences should be 'recycled' and developed in order to gain synergies as future expansion of a national data centre concept progresses; but they can largely be repeated in new locations in terms of design and capacity. However, one crucial difference between secure buildings and commercial reference properties is that physical security aspects will cost more, as it is imperative to protect the data centre from peacetime threat scenarios; including burglary, sabotage, suicide bombers, attacks with guns, car bombs and ramraiding.

It is likely that stringent demands are made of physical security for the vast majority of IT services that may be considered for a governmental data centre concept, but not on the level offered by protected underground facilities. On the one hand, a secure building that is specifically planned, positioned and designed for the purpose will probably offer an entirely satisfactory level of physical security for the majority of vital societal IT services. On the other, however, a wide

range of IT services also have to survive the demands of war by means of the fortified conditions offered by protected underground facilities.

Another way to derive benefit from a protected underground facility is to use it for storage and backup purposes; which is a less energy-intensive undertaking on the whole. Energy-intensive server services with more stringent uptime requirements can primarily be managed in secure buildings. This ensures that high capacity is available for vital societal IT services in terms of uptime and extensive security on a day-to-day basis, while the overall data volume is backed up regularly to a protected underground facility. In the event of adverse incidents, this concept means that data backed up to a protected underground facility will be inaccessible while it is being recovered to another secure building that is operational, but on the other hand it will be highly accurate in that no data will be lost and so it will be possible to recover it.

A PROTECTED UNDERGROUND FACILITY AND A SECURE BUILDING – AN ADVANTAGEOUS COMBINATION

There is need for further examination of protected underground facilities from a strategic perspective. Nevertheless, as a final example below, an overall national power requirement slightly in excess of 20 MW (enough power to run around 9000 homes) can be achieved by using 15 buildings and protected underground facilities (new sites and converted rock caverns) all over Sweden:

A conceptual regional data centre cluster could be distributed as follows:

- Two secure buildings, each with a power output of 2 MW (new sites)
- One protected underground facility with a power output of 0.5 MW (conversion of an existing underground facility that is currently not operational)

Five such regional data centre clusters throughout Sweden would therefore involve, in total:

- Ten secure buildings, each with a power output of 2 MW (new sites)
- Five protected underground facilities with a power output of 0.5 MW (conversion of existing underground facilities that are currently not operational)

As mentioned previously, the relatively long procurement time for a data centre concept is worth noting. Time-critical parameters that influence the production time for a new secure building include acquisition of land, environmental surveys and a secured contractor procurement process.

It is estimated that it takes about two years to procure a secure building. It is thought that conversion of a fortified underground facility will take four to five years. Several regional data centre clusters can be constructed by degrees or simultaneously. However, it is fair to assume that there will be no need for the full capacity in the short term, and that there will therefore be no need to implement the entire concept in parallel. Making the most of experiences from initial construction projects before taking unnecessarily large steps is also logical. It is therefore reasonable to assume that a total procurement time of about ten years may be required.

Overall, a potential data centre concept as described in the example above may offer high levels of uptime and redundancy at a more reasonable cost than if all the data centres are newly constructed fortified underground facilities. Although uptime for vital societal and security-critical IT services will be affected with the data centre concept, this will provide a high level of security in terms of both physical perspectives and accuracy, in the sense that no data will be lost even under extreme conditions such as times of crisis or war. However, a prerequisite for a procurement time of around ten years is that necessary strategic elements such as requirement specifications, funding and organisation have been established beforehand. Hence decision-makers should raise awareness of the needs and challenges associated with national coordination of protected IT services so that the process can commence. This is a vital and necessary investment in the future for the total defence of Sweden.