ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TU2BC001

# DATABASE'S DISASTER RECOVERY MEETS A RANSOMWARE ATTACK

M. Zambrano†, SKA Observatory
Jodrell Bank, Macclesfield, UK
V. Gonzalez, ALMA Observatory, Joint ALMA Office, Santiago, Chile

*Abstract*

Cyberattacks are a growing threat to organizations around the world, including observatories. These incidents can cause significant disruption to operations and can be costly to recover from. This paper provides an overview of the history of cyberattacks, the motivations of attackers, and the organization of cybercrime groups. It also discusses the steps that can be taken to quickly restore a key component of any organization, the database, and the lessons learned during the recovery process.

The paper concludes by identifying some areas for improvement in cybersecurity, such as the need for better training for employees, more secure networks, and more robust data backup and recovery procedures.

## INTRODUCTION

Hacking is now a multimillion-dollar industry where motivated people are looking for their next victims. Hospitals, schools, councils and observatories have also been targets of those groups. This threat will not disappear soon, and each organization should be prepared to handle a never-ending list of attacks. The average cost of a ransomware attack in the USA in 2023 was 4.45 million USD [1]. The median time to contain an attack is 73 days but this could be increased by 14 days when the disclosure is made by the attacker [1].

## A VERY SHORT STORY ON HACKING

Spring 1962 at MIT, Allan Scherr needed more than four hours to run his PhD simulation models of time-sharing systems [2]. He asked the system to print the password file on a punched card and he was able to steal credentials for the first time in history.

Sometime later in 1975 Kevin Mitnick started his hacking life by "traveling free" when he asked the driver of his bus where he bought the funny punch machine they use for a "school project". He continued his life by learning everything he could about the phone system. A friend gave him the phone number of a system used to develop an OS. Later he was searched by the FBI for intrusions at IBM, Sun and Motorola and four counts of wire fraud, two counts of computer fraud and one count of illegally intercepting a wire communication. He spent some 5 years in prison and returned as a "white hat" security professional consultant. He was the first to use "social engineering" to get what he wanted from people, and he continued to do so as a security consultant.

† mauricio.zambrano@skao.int

Hacking at this point started to be a for profit business with a lot of curious, talented young people with no guilt or remorse of doing what they were doing.

One of the main concepts that Mitnick exploited was the human factor. You can have the best security system in the world but if people are getting exploited, then you will be able to circumvent that system.

A new subculture emerged where breaking into other people's computers wasn't seen as bad. There was also a huge economic benefit of doing this and the perpetrators have no empathy for the real damage caused to real people. There is also an important cost of being caught by the authorities. If you "succeed" in this business, you might end up being prosecuted by the FBI. They now have a list of 119 individuals known as the Cyber's most wanted. Currently most of them are from Russia, North Korea, Iran and China and linked to their states and armies in some cases. A few of them have been captured when traveling abroad. Since then, they usually stay in their own countries. Some of the hacker activities also have had an important political impact for the targeted country.

The United Nations has been trying to define responsible cyber behavior for states with an Open-Ended Working Group negotiation but the talks, so far, have not reached an agreement. Russia has also submitted its vision for a Convention of the UN on ensuring International Information Security. Hence there is currently no change in the horizon for this kind of activity.

## BRAND NEW WORLD?

Hacking groups have existed since 1980. One of the first of them was named the 414s and was composed of "young, male, intelligent, highly motivated and energetic" from Milwaukie, Wisconsin. They met each other at an IBM youth program. By using information from the DEC manual to get the default password, they were able to breach systems at Los Alamos National Laboratories, a bank and deleted bill records of a hospital. Since then, groups have increased in complexity.

In each group there are specialized actors with different abilities with the final objective to get money from the victim.

There is also a trend to offer hacking as a service. In that case all the people involved share a portion of the ransom.

Most attacks start:

- With an innocent click on a phishing mail.
- Stolen credentials.
- Via an exploit on an exposed server. There is a lucrative market for zero-day exploits for all the different platforms.

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TU2BC001

- Via Managed Service Providers or an affected partner.
- By using free Wi-Fi in a public place.
- Downloading software

Once the infected content is on a victim's PC the attacker can wait until he has a foothold on the company's network. They will try to reach a valued target to access more resources and sensitive data.

The hackers will try to compromise more systems and launch a full-scale attack whenever they have collected enough information.

## A DATABASE RECOVERY PLAN

The design of the recovery plan should start with some definition of what is expected from the system. Two main parameters will define what architecture is needed to meet those requirements:

- Recovery Point Objective (RPO): the maximum amount of time of data loss before the incident.
- Recovery Time Objective (RTO): the maximum amount of time the system can be offline when recovering it.

Once established these 2 parameters will influence the backup requirements and the database architecture needed.

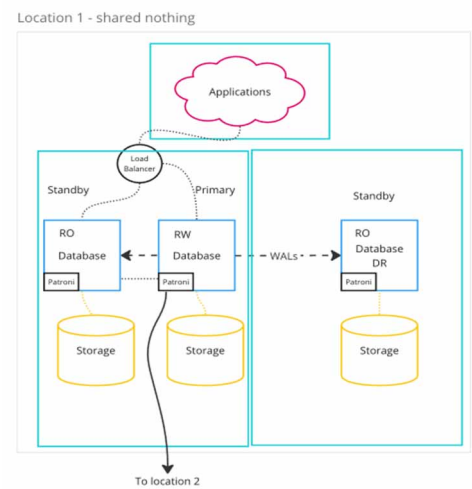


Figure 1: A shared nothing PostgreSQL architecture. The role of the database is controlled by Patroni in this case. Replication is done via PostgreSQL subscription mechanism. There is only one read/write node, the primary and the others are standby servers that can serve reads. The load balancer must send the right request to the specific node. Clients are unaware of the node's roles. MariaDB also offers a cluster solution called Galera where all the nodes can serve read and write.

Usually the Database Disaster Recovery Plan (DDRP) would consider having:

- One or more primary servers where the applications read and write happen. The more servers you add, the more resilient is your architecture against single server hardware failures at a higher cost. Depending on the database vendor you are using you might have a shared or a shared nothing storage as shown in Fig. 1 and Fig. 2. In the case of a shared storage, that component is a key one and a single point of failure. Its security and availability are very important to provide the level of service needed.
- One or more secondary read only servers. Replication between them is granted usually by a log replication mechanism specific to each database vendor.
- There could be an automatic or manual mechanism to enable role change between primaries and secondaries. In case of a node promotion, there should be a mechanism to ensure that the clients connect to the correct server. This could be achieved by a configuration at the client level or by using a proxy service that could detect the role of each server.

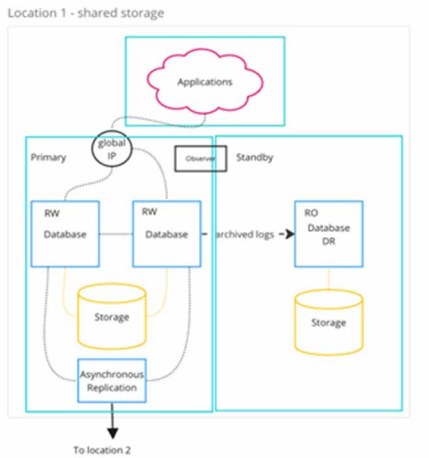


Figure 2: A shared storage Oracle architecture. This is a common architecture for proprietary systems. The storage is shared among the database nodes. Each node on the primary location can serve read and write. In this case the replication is done via DataGuard and asynchronous replication is served with GoldenGate. An Observer can decide when to change to switch the service to the standby server. The application must be configured to handle this configuration change.

- Your backup strategy must take in account the RPO and RTO. You might implement a weekly full backup and then an incremental one. You might also want to save logged transactions as soon as they are created on the primary. That would provide Continuous Data Protection (CDP) which is key to get zero data loss. The backups should be ideally stored on their own shared nothing storage and for improved security, moved or copied to a cloud.
- Primary and secondary must reside on different servers and each one should have their own storage. Usually, the secondary could be a standalone server with its own set of disks as storage, ideally on another datacentre close to the primary (less than 10ms away).

In modern database architectures, there will be replication to other data centers between the primary source and external destinations like regional centers. Having replication to a distant region or continent would provide an extra protection against a catastrophic failure in one of the

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TU2BC001



regions. In a worst-case scenario, those remote copies will be able to provide a backup.

An RPO 0 strategy implies that any transaction that arrives at a primary server must be first applied into one or more replicas then, when the primary receives the confirmation, it should be applied on the primary. This synchronous replication has a latency cost due to the cost of replicas transaction confirmation. RPO 0 is normally used in heavily regulated industries where no transaction can be lost.

Ransomware attacks will encrypt disks from affected systems. This encryption is done usually on a hypervisor system with external tools. If the system is rebooted, then the data access to the encrypted file is lost. This disaster case was usually not taken into consideration when planning for a disaster recovery. In the current situation of cyber security, you must also plan for this.

In case of an attack on the database you should assess the damage done. Contain the damage done by shutting down unaffected systems. The database recovery should be part of an integral recovery effort that should first start with the reconstruction of a brand-new network and infrastructure on new hardware.

## DATABASE BACKUP CONSIDERATIONS

Financial services have already in place or are starting to implement policies to prevent attacks or to reduce the recovery time after them. One of those requirements is to have off site Write Once Read Many (WORM) or offline backups. They should be stored in safe locations ideally outside of the organization's network or offline. In case of a successful attack this will ensure that at least one recent backup is available to start the recovery from there. In any case you should carefully consider where your backups are being stored. They should not reside on the same storage device where the production system is. It is important to coordinate with IT the details of the backup procedure and how they keep them for a longer period.

It is also a good idea to enable encryption on backups to reduce risks when losing them.

## DATABASE SECURITY DESIGN CONSIDERATIONS

When designing a database schema for a system, you should consider:

- Users needed by the system.
- Roles each user might need and their privileges.
- Ways to exchange data with other systems and network access needed by the database.
- Secure ways to store the credential and rotation mechanism.
- Securing the data access by preventing SQL injection.

After establishing the users that will access the database, you should map them in roles by following the principle of least privilege. For example, if a user has write and read access to some tables and others, like the one doing reports, can just have read access to the specific tables. You should control who has access to the tables and avoid using public schemas to host your tables. If you need to give access to tables, grant the appropriate rights to the specific roles and use views or synonyms to hide the full table schema.

Credential storage should never be in clear text and a strong hashing should be used. The user's credentials might be stored in a vault system. Those systems can manage time limited sets of credentials and rotate them at will. You might create one user per each client node connecting to the database. This way if a client node is compromised, you can isolate that specific node/user and revoke that credential without affecting the whole system.

User credentials and administrator credentials must be rotated regularly.

The database system should not have permissions to access any public IP. Network access should be granted to specific ips where the systems run and to specific partners ips where data exchange occur. Access to OS upgrades should be granted via a caching server or a secure proxy.

SQL injection is a known mechanism to execute malicious code on a database server when a query uses client provided information. An attacker could craft a specific SQL code that could run commands outside of the database. To prevent this flaw, the client code should validate each input given by the user making sure it contains only the expected data types and no escape character. Most modern SQL libraries or ORM will contain ways to prevent SQL injection. You can see a summary of mitigations on Table 1.

Table 1: Libraries & Features to Prevent SQL Injection

| Language | Library or feature | Database |
|---|---|---|
| java | jdbc prepared statement | Any with a jdbc |
| go | Use parameterized queries | Any with a go driver |
| python | psycopg2.sql - SQL string composition | PostgreSQL |
| python | MySQL cursor execute - use parameters | MySQL/MariaDB |
| Typescript | @nearform/sql | It is used with PostgreSQL, MySQL and mysql2 libraries |

You should also secure APIs by using sanitization libraries to check for malicious user input.

Many developers use for their simplicity sequence numbers as primary keys. This might allow an attacker to easily retrieve information from an API by just increasing one parameter. If you use instead of integer Universal Unique Identifiers (uuids) this won't be possible anymore. UUIDs are a 128-bit label with very low probability of collision.

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TU2BC001

## DATABASE BACKUP TESTING AND RECOVERY

The only usable backup is the one that has been tested. It is very important to fully document the recovery procedures for your database software and periodically test those backups. Multiple recovery scenarios must be envisioned. From the easiest one, recovering from the official backups to the hardest one, recover from a distant copy and recovery from other systems used in normal operations. In this case there were three backup sources. One on the IT system, other on the standby service that was receiving transactions from production with milliseconds delays and a testing backup. The testing system receiving transactions from production was tested two to three times per month. A snapshot image of the base system was mounted and deployed on a database for testing purposes [3].

The decision on where to recover the data depends on the timing and speed needed for the recovery. It could quickly promote a standby to a primary server, but you might face in the long-term performance issues, and you still must recover on the primary system. It might make more sense to recover on the primary database. The database recovery is probably part of a larger effort to recover all the IT systems and usually this is one of the latest parts before the applications recovery when suffering a ransomware attack.

## DATA REPLICATION AND ALIGNEMENT

Modern data architectures will require some kind of metadata replication with software from your database ecosystem. Implementations are specific to each vendor but usually you could have a replication middleware that copies committed transactions in the form of trail files, binary representations of compressed transactions, that are copied from one datacenter to another. That replication is most of the time asynchronous since it has a low impact on the performance of the source system. You could also have tools to check the status of the replication. These tools can have some limitations like for example not working on XML tables. For those cases, you will have to create your own set of tools. A nice way to reduce the amount of comparison is having a timestamp column with the latest change time. This way you can identify the time frame where you must do any data comparison and deduce which tables have missing rows or missing new information. You can compare the primary keys with a full hash of a concatenation of all the columns.

In other cases, replication might be achieved directly via subscription from source to target. You could also use the same alignment tools in this case.

It is very unlikely that the disaster hit two sites at the same time. You will have to carefully detect the latest transactions and whenever possible, use alignment tools to fix any misalignment to avoid any transaction lost.

## BE PREPARED

Data breaches have only augmented in recent years. All organizations must be prepared and ready to respond to such incidents. You should have:

- Spare hardware to reinstall from scratch servers, virtual machines and network appliances.
- Automate the IT infrastructure deployment as much as possible. This is probably easier if you deploy things in containers and Kubernetes.
- Evaluate the recovery time for each system.
- Have written plans on how to proceed.
- Test those procedures and the backups. The only good backup is the one that has been tested by a recovery.
- Detect any critical infrastructure that has no duplicates like storages. Make sure its access is well protected and has multi factor authentication.
- Follow recommendations from other more regulated industries.
- For example, in August 2023, the Security Exchange Board of India created Cyber security and resilience guidance [4]. They requested the implementation of 28 measures to improve the security for Market Infrastructure Institutions. Amongst the measure they recommend:
- Implement multi factor authentication as much as possible.
- Have encrypted offline backups, outside of your network.
- Having “gold images” to rebuild services.
- Having spare hardware where you could run full-service reinstallation outside of the primary and disaster recovery datacenter.
- Deprecate unused tokens periodically.
- Secure the Domain Controller and Active Directory services since they are used to spread ransomware.
- Network segregation to mitigate any attack impact.
- Explore the possibility to run systems under different architecture to prevent failures. This is very important for components like hypervisors. The popularity of a special product could make it an easy target for attackers.

Observatories do not have the resources of financial services and they are not regulated by authorities. In case of an attack, it is very unlikely that any ransom might be paid. That has not been the case now. Eventually attackers might be inclined in targeting more lucrative targets.

## CONCLUSIONS

Given the current environment it is hard to see an end to ransomware attacks. It is normal to expect an increase in the number of attacks to organizations.

Databases should be protected by taking into consideration this environment and continuous tests of a catastrophic recovery should be performed on a regular basis.

ISBN: 978-3-95450-238-7 ISSN: 2226-0358 doi:10.18429/JACoW-ICALEPCS2023-TU2BC001

## REFERENCES

[1] IBM, The cost of data breach report 2023; https://www.ibm.com/reports/data-breach

[2] D. C. Walden, and T. Van Vleck, "The compatible time sharing system (1961-1973): Fiftieth anniversary commemorative overview", (2011), https://studylib.net/doc/18434241/compatible-time-sharing-system--1961-1973--fiftieth

[3] M. Zambrano *et al.*, "Enabling faster ALMA software delivery by using containers", in *Proc. SPIE, Software and Cyberinfrastructure for Astronomy VI*, vol. 11452, Dec. 2020. doi:10.1117/12.2561935

[4] SEBI, Guidelines for MIIs regarding Cyber security and Cyber resilience, https://www.sebi.gov.in/legal/circulars/aug-2023/guidelines-for-miis-regarding-cyber-security-and-cyber-resilience_76056.html