From the Outside In:
How I Uncovered the Cybersecurity Failures of Thousands of Companies

`Josephine Pfeiffer, 03/2023`

Some background on me

- Coding, tinkering around with Linux for ~9 years
- Most interested in (hybrid) cloud, SRE, cybersecurity
- Previously TPM, SRE at Sygnum
- Currently Cloud Native Consultant at Red Hat

`https://josie.lol`

Disclaimer

The opinions expressed in this presentation are solely those of the presenter and do not necessarily reflect the views or policies of the presenter's past, current or future employers.

The information presented is for educational and informational purposes only and should not be construed as professional advice.

The presenter takes no responsibility for any actions taken based on the information provided during this presentation.

Once upon a time…

I was bored on PTO waiting to start my new job.

Researching data breaches & started to wonder what is floating around in public…

…what about S3 buckets and `.tfstate` files?

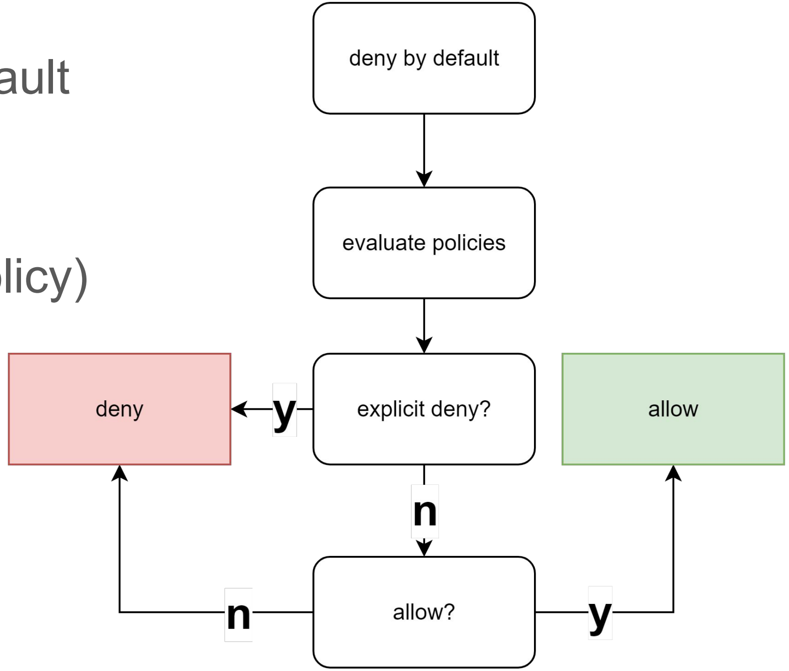
Some facts on S3 endpoints to begin with

- Be unique across all of Amazon S3
- Be between 3 and 63 characters long
- Not contain uppercase characters
- Start with a lowercase letter or number

```
https:\/\/[a-z0-9]([-a-z0-9]{2,61}[a-z0-9])?\.s3\.amazonaws\.com\/
```

S3 bucket access controls

- The bucket policy is always private by default
- One explicit “deny” policy trumps “allow” policies at other levels (IAM, S3 object/bucket ACL, S3 bucket policy)



Scraping endpoints (brief for legal reasons)

- Public? (y/N)
- Paths/prefixes make things more complicated
  - Traversing paths, recursively scan for filenames, extensions (`.tfstate`, `production`, `.env`, `secret.yaml`, etc.)

What I found

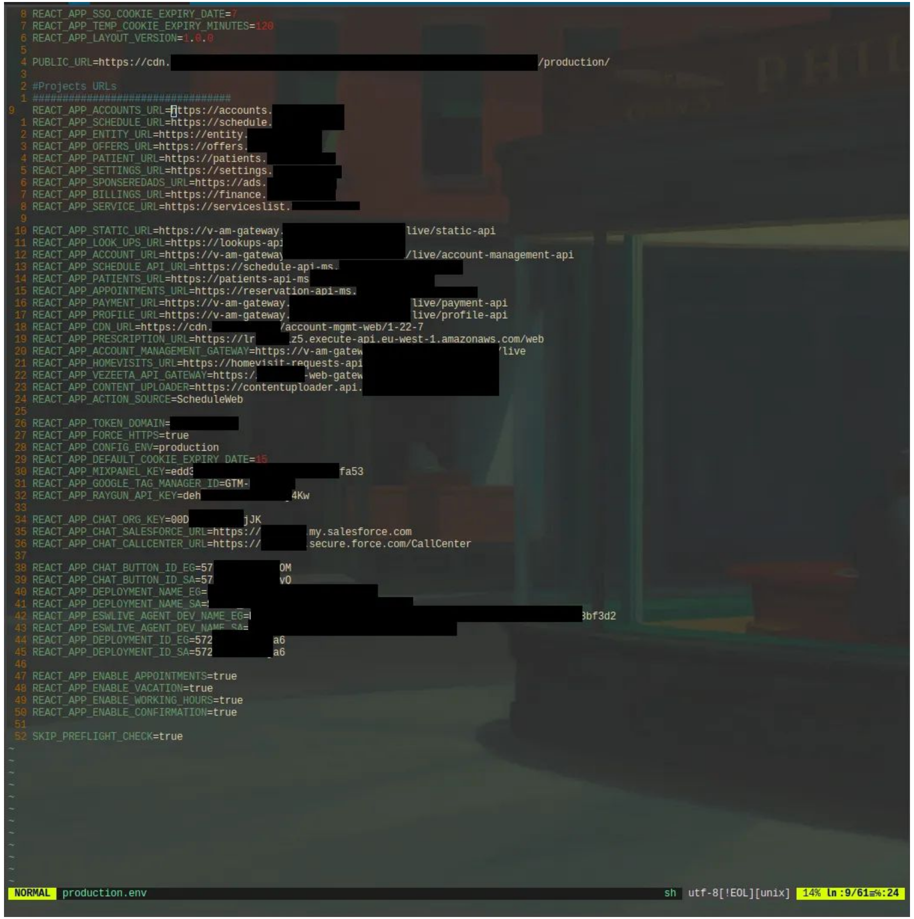
- In total, I scanned through ~308k AWS S3 buckets.
- Within only a few minutes, I could freely look through production secrets for thousands of large, international companies.

What I found

An international digital healthcare company from MENA

Stored a `production.env` file in a publicly accessible S3 bucket.

The file contained API keys and admin credentials to a CRM, patient database, and other internal systems.

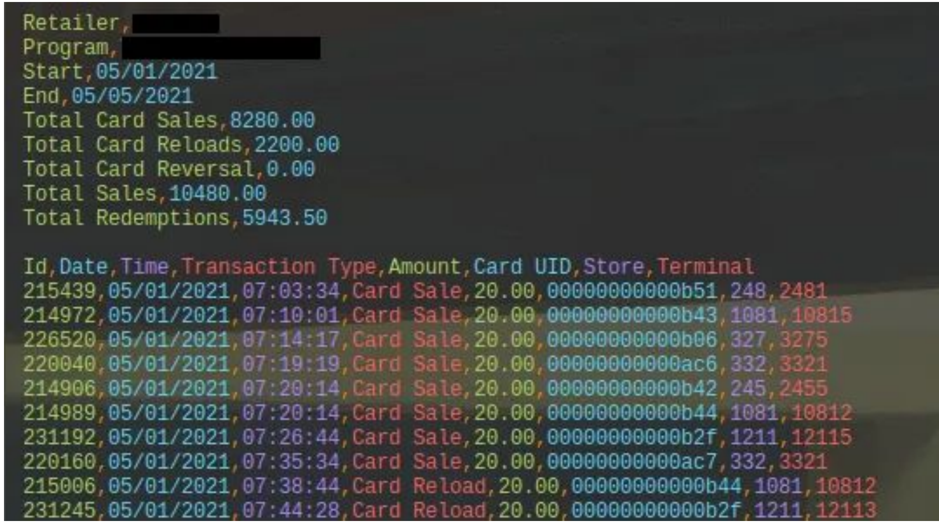


```
REACT_APP_SSO_COOKIE_EXPIRY_DATE=7
REACT_APP_TEMP_COOKIE_EXPIRY_MINUTES=120
REACT_APP_LAYOUT_VERSION=1.0.0

PUBLIC_URL=https://cdn.[redacted]/production/

#Projects URLs
##################################
REACT_APP_ACCOUNTS_URL=https://accounts.[redacted]
REACT_APP_SCHEDULE_URL=https://schedule.[redacted]
REACT_APP_ENTITY_URL=https://entity.[redacted]
REACT_APP_OFFERS_URL=https://offers.[redacted]
REACT_APP_PATIENT_URL=https://patients.[redacted]
REACT_APP_SETTINGS_URL=https://settings.[redacted]
REACT_APP_SPONSEREDADS_URL=https://ads.[redacted]
REACT_APP_BILLINGS_URL=https://finance.[redacted]
REACT_APP_SERVICE_URL=https://serviceslist.[redacted]

REACT_APP_STATIC_URL=https://v-am-gateway.[redacted]live/static-api
REACT_APP_LOOK_UPS_URL=https://lookups-api[redacted]
REACT_APP_ACCOUNT_URL=https://v-am-gateway[redacted]/live/account-management-api
REACT_APP_SCHEDULE_API_URL=https://schedule-api-ms.[redacted]
REACT_APP_PATIENTS_URL=https://patients-api-ms[redacted]
REACT_APP_APPOINTMENTS_URL=https://reservation-api-ms.[redacted]
REACT_APP_PAYMENT_URL=https://v-am-gateway.[redacted]live/payment-api
REACT_APP_PROFILE_URL=https://v-am-gateway.[redacted]live/profile-api
REACT_APP_CDN_URL=https://cdn.[redacted]/account-mgmt-web/1-22-7
REACT_APP_PRESCRIPTION_URL=https://lr[redacted]z5.execute-api.eu-west-1.amazonaws.com/web
REACT_APP_ACCOUNT_MANAGEMENT_GATEWAY=https://v-am-gatew[redacted]/live
REACT_APP_HOMEVISITS_URL=https://homevisit-requests-api[redacted]
REACT_APP_VEZEETA_API_GATEWAY=https:/[redacted]-web-gatew[redacted]
REACT_APP_CONTENT_UPLOADER=https://contentuploader.api.[redacted]
REACT_APP_ACTION_SOURCE=ScheduleWeb

REACT_APP_TOKEN_DOMAIN=[redacted]
REACT_APP_FORCE_HTTPS=true
REACT_APP_CONFIG_ENV=production
REACT_APP_DEFAULT_COOKIE_EXPIRY_DATE=15
REACT_APP_MIXPANEL_KEY=edd3[redacted]fa53
REACT_APP_GOOGLE_TAG_MANAGER_ID=GTM-[redacted]
REACT_APP_RAYGUN_API_KEY=deh[redacted]4Kw

REACT_APP_CHAT_ORG_KEY=00D[redacted]jJK
REACT_APP_CHAT_SALESFORCE_URL=https://[redacted].my.salesforce.com
REACT_APP_CHAT_CALLCENTER_URL=https://[redacted].secure.force.com/CallCenter

REACT_APP_CHAT_BUTTON_ID_EG=57[redacted]OM
REACT_APP_CHAT_BUTTON_ID_SA=57[redacted]vO
REACT_APP_DEPLOYMENT_NAME_EG=[redacted]
REACT_APP_DEPLOYMENT_NAME_SA=[redacted]
REACT_APP_ESWLIVE_AGENT_DEV_NAME_EG=[redacted]3bf3d2
REACT_APP_ESWLIVE_AGENT_DEV_NAME_SA=[redacted]
REACT_APP_DEPLOYMENT_ID_EG=572[redacted]a6
REACT_APP_DEPLOYMENT_ID_SA=572[redacted]a6

REACT_APP_ENABLE_APPOINTMENTS=true
REACT_APP_ENABLE_VACATION=true
REACT_APP_ENABLE_WORKING_HOURS=true
REACT_APP_ENABLE_CONFIRMATION=true

SKIP_PREFLIGHT_CHECK=true
~
~
~
~
~
~
~
~
~
~
~
~
NORMAL  production.env                                   sh  utf-8[!EOL][unix]  14% ln :9/61≡%:24
```

What I found

A US lottery company

Directly stored customer and transaction data as `.csv` files in a publicly accessible S3 bucket.

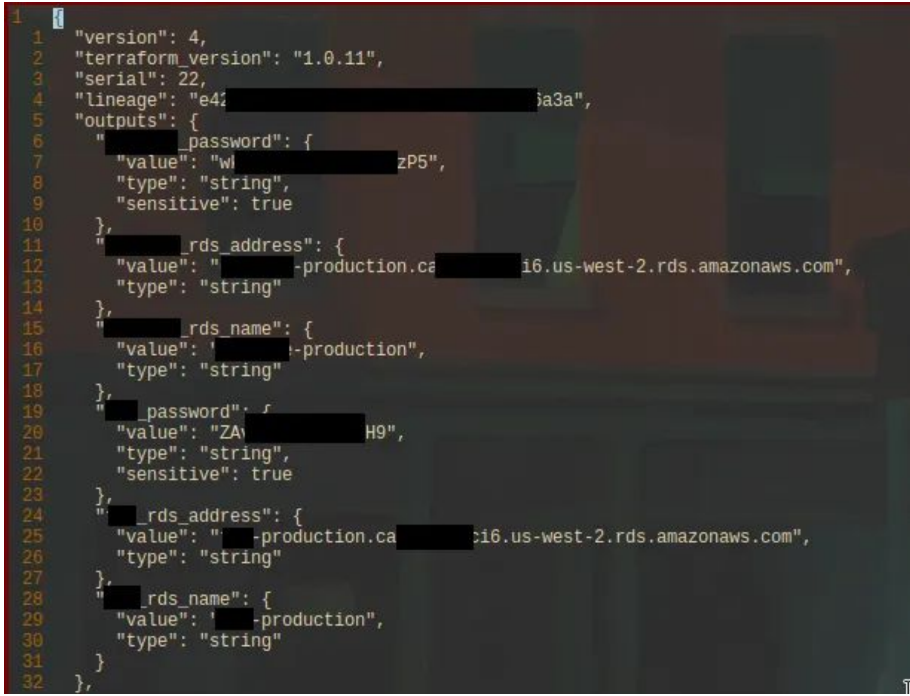


```
Retailer,[redacted]
Program,[redacted]
Start,05/01/2021
End,05/05/2021
Total Card Sales,8280.00
Total Card Reloads,2200.00
Total Card Reversal,0.00
Total Sales,10480.00
Total Redemptions,5943.50

Id,Date,Time,Transaction Type,Amount,Card UID,Store,Terminal
215439,05/01/2021,07:03:34,Card Sale,20.00,00000000000b51,248,2481
214972,05/01/2021,07:10:01,Card Sale,20.00,00000000000b43,1081,10815
226520,05/01/2021,07:14:17,Card Sale,20.00,00000000000b06,327,3275
220040,05/01/2021,07:19:19,Card Sale,20.00,00000000000ac6,332,3321
214906,05/01/2021,07:20:14,Card Sale,20.00,00000000000b42,245,2455
214989,05/01/2021,07:20:14,Card Sale,20.00,00000000000b44,1081,10812
231192,05/01/2021,07:26:44,Card Sale,20.00,00000000000b2f,1211,12115
220160,05/01/2021,07:35:34,Card Sale,20.00,00000000000ac7,332,3321
215006,05/01/2021,07:38:44,Card Reload,20.00,00000000000b44,1081,10812
231245,05/01/2021,07:44:28,Card Reload,20.00,00000000000b2f,1211,12113
```

What I found

A US lottery company

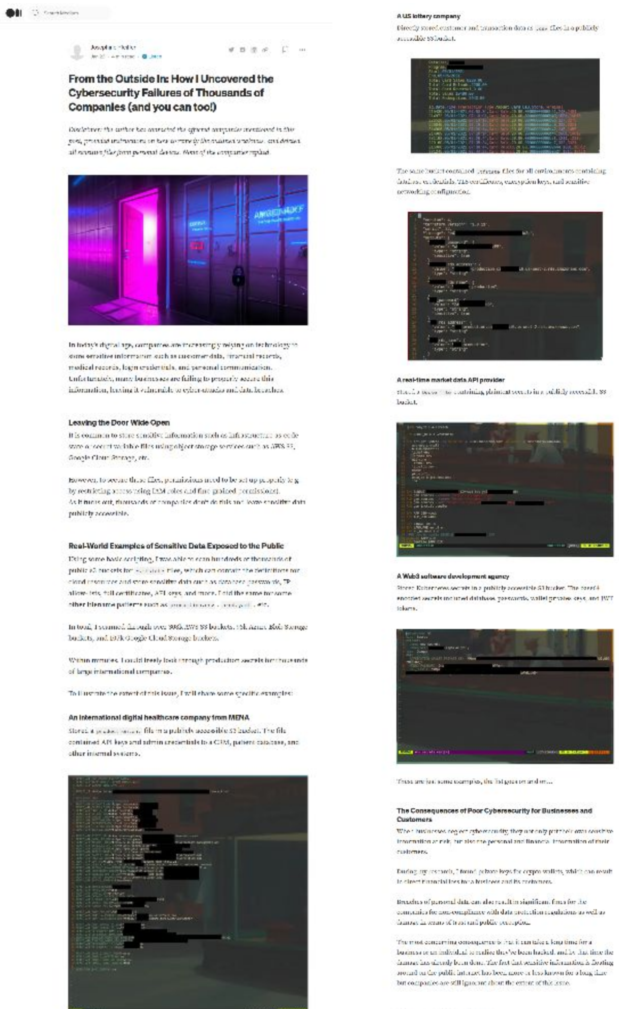
The same bucket contained `.tfstate` files for all environments containing database credentials, TLS certificates, encryption keys, and sensitive networking configuration.



```
{
  "version": 4,
  "terraform_version": "1.0.11",
  "serial": 22,
  "lineage": "e42[redacted]a3a",
  "outputs": {
    "[redacted]_password": {
      "value": "w[redacted]zP5",
      "type": "string",
      "sensitive": true
    },
    "[redacted]_rds_address": {
      "value": "[redacted]-production.ca[redacted]i6.us-west-2.rds.amazonaws.com",
      "type": "string"
    },
    "[redacted]_rds_name": {
      "value": "[redacted]-production",
      "type": "string"
    },
    "[redacted]_password": {
      "value": "ZA[redacted]H9",
      "type": "string",
      "sensitive": true
    },
    "[redacted]_rds_address": {
      "value": "[redacted]-production.ca[redacted]i6.us-west-2.rds.amazonaws.com",
      "type": "string"
    },
    "[redacted]_rds_name": {
      "value": "[redacted]-production",
      "type": "string"
    }
  },
```

Blog post

- Reached out to companies (none responded)
- Published blog post
- ~5k reads on medium
- Discussed at length on r/cybersecurity, and hackernews



[illegible]
Jan 23 · 4 min read · Listen

From the Outside In: How I Uncovered the Cybersecurity Failures of Thousands of Companies (and you can too!)

Disclaimer: the author has contacted the affected companies mentioned in the post, provided instructions on how to remedy the situation [illegible], and deleted all accessed files from personal devices. None of the companies replied.

In today's digital age, companies are increasingly relying on technology to store sensitive information such as customer data, financial records, medical records, login credentials, and personal communication. Unfortunately, many businesses are failing to properly secure this information, leaving it vulnerable to cyber-attacks and data breaches.

Leaving the Door Wide Open

It is common to store sensitive information such as infrastructure as code state or secret variable files using object storage services such as AWS S3, Google Cloud Storage, etc.

However, to secure these files, permissions need to be set up properly (e.g. by restricting access using IAM roles and fine-grained permissions). As it turns out, thousands of companies don't do this and leave sensitive data publicly accessible.

Real-World Examples of Sensitive Data Exposed to the Public

Using some basic scripting, I was able to scan hundreds of thousands of public S3 buckets for *.tfstate files, which can contain the definitions for cloud resources and store sensitive data such as database passwords, IP allow-lists, TLS certificates, API keys, and more. I did the same for some other filename patterns such as [illegible] , [illegible] , etc.

In total, I scanned through over 900k AWS S3 buckets, 15k Azure Blob Storage buckets, and 107k Google Cloud Storage buckets.

Within minutes, I could freely look through production secrets for thousands of large international companies.

To illustrate the extent of this issue, I will share some specific examples:

An international digital healthcare company from MENA

Stored a [illegible] file in a publicly accessible S3 bucket. The file contained API keys and admin credentials to a CRM, patient database, and other internal systems.

A US lottery company

Directly stored customer and transaction data as .csv files in a publicly accessible S3 bucket.

The same bucket contained .tfstate files for all environments containing database credentials, TLS certificates, encryption keys, and sensitive networking configuration.

A real-time market data API provider

Stored a .tfstate file containing plaintext secrets in a publicly accessible S3 bucket.

A Web3 software development agency

Stored Kubernetes secrets in a publicly accessible S3 bucket. The base64 encoded secrets included database passwords, wallet private keys, and JWT tokens.

These are just some examples, the list goes on and on…

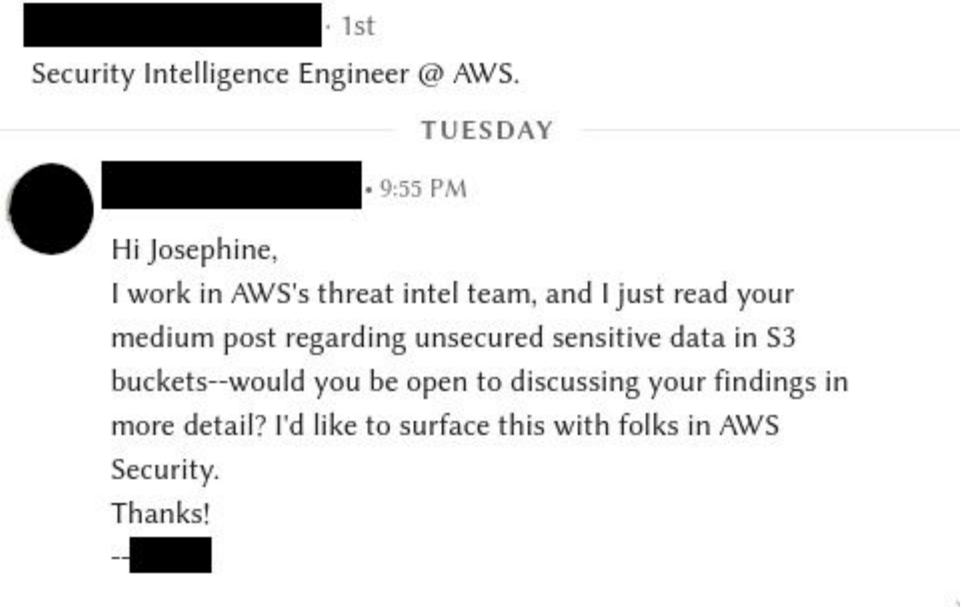
The Consequences of Poor Cybersecurity for Businesses and Customers

When businesses neglect cybersecurity, they not only put their own sensitive information at risk, but also the personal and financial information of their customers.

During my research, I found private keys for crypto wallets, which can result in direct financial loss for a business and its customers.

Breaches of personal data can also result in significant fines for the companies for non-compliance with data protection regulations as well as damage in terms of trust and public perception.

The most concerning consequence is that it can take a long time for a business or an individual to realise they've been hacked, and by that time the damage has already been done. The fact that sensitive information is floating around on the public internet has been more or less known for a long time but companies are still ignorant about the extent of this issue.



[redacted] · 1st

Security Intelligence Engineer @ AWS.

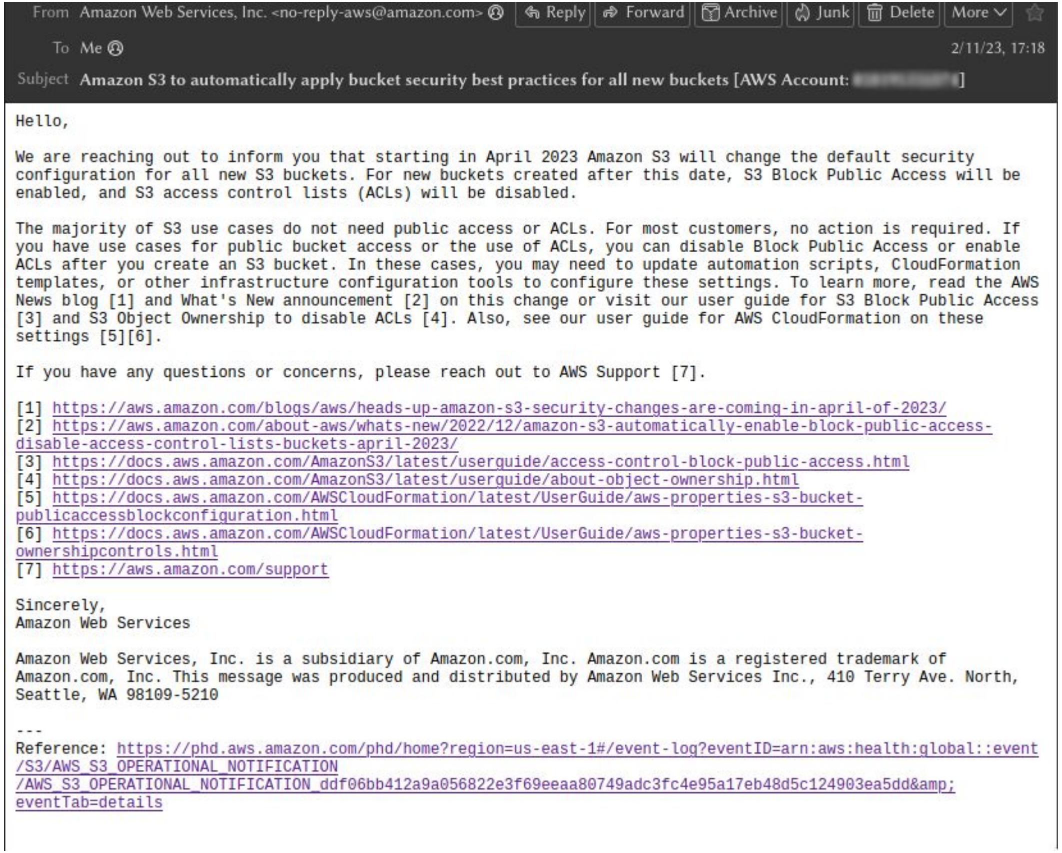
TUESDAY

[redacted]

[redacted] • 9:55 PM

Hi Josephine,
I work in AWS's threat intel team, and I just read your medium post regarding unsecured sensitive data in S3 buckets--would you be open to discussing your findings in more detail? I'd like to surface this with folks in AWS Security.
Thanks!
--[redacted]

Results



From Amazon Web Services, Inc. <no-reply-aws@amazon.com>

To Me

2/11/23, 17:18

Subject **Amazon S3 to automatically apply bucket security best practices for all new buckets [AWS Account: [redacted]]**

Hello,

We are reaching out to inform you that starting in April 2023 Amazon S3 will change the default security configuration for all new S3 buckets. For new buckets created after this date, S3 Block Public Access will be enabled, and S3 access control lists (ACLs) will be disabled.

The majority of S3 use cases do not need public access or ACLs. For most customers, no action is required. If you have use cases for public bucket access or the use of ACLs, you can disable Block Public Access or enable ACLs after you create an S3 bucket. In these cases, you may need to update automation scripts, CloudFormation templates, or other infrastructure configuration tools to configure these settings. To learn more, read the AWS News blog [1] and What's New announcement [2] on this change or visit our user guide for S3 Block Public Access [3] and S3 Object Ownership to disable ACLs [4]. Also, see our user guide for AWS CloudFormation on these settings [5][6].

If you have any questions or concerns, please reach out to AWS Support [7].

[1] https://aws.amazon.com/blogs/aws/heads-up-amazon-s3-security-changes-are-coming-in-april-of-2023/
[2] https://aws.amazon.com/about-aws/whats-new/2022/12/amazon-s3-automatically-enable-block-public-access-disable-access-control-lists-buckets-april-2023/
[3] https://docs.aws.amazon.com/AmazonS3/latest/userguide/access-control-block-public-access.html
[4] https://docs.aws.amazon.com/AmazonS3/latest/userguide/about-object-ownership.html
[5] https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/aws-properties-s3-bucket-publicaccessblockconfiguration.html
[6] https://docs.aws.amazon.com/AWSCloudFormation/latest/UserGuide/aws-properties-s3-bucket-ownershipcontrols.html
[7] https://aws.amazon.com/support

Sincerely,
Amazon Web Services

Amazon Web Services, Inc. is a subsidiary of Amazon.com, Inc. Amazon.com is a registered trademark of Amazon.com, Inc. This message was produced and distributed by Amazon Web Services Inc., 410 Terry Ave. North, Seattle, WA 98109-5210

Reference: https://phd.aws.amazon.com/phd/home?region=us-east-1#/event-log?eventID=arn:aws:health:global::event/S3/AWS_S3_OPERATIONAL_NOTIFICATION/AWS_S3_OPERATIONAL_NOTIFICATION_ddf06bb412a9a056822e3f69eeaa80749adc3fc4e95a17eb48d5c124903ea5dd&amp;eventTab=details

TL;DR

Default security config for new buckets now have public access disabled.

Key takeaways

- This is a really simple attack vector
    - Scary how easy it was: Awareness is important
- Impact can be catastrophic for businesses & customers
    - Financial, reputational, privacy violations
- Humans are the weakest link
    - The tech works fine, if used right
- A lot of companies are doing it right:
    - <1% of the scraped endpoints were public
      (excluding buckets that were obviously meant to be public – e.g. static website hosting)
- Cloud providers are taking steps
    - Making it even harder for users to leak data

Q&A