FLAWS.CLOUD DATASET - COMPLETE ANALYSIS SUMMARY

📋 TABLE OF CONTENTS

1. What Scott Piper Created
2. Example Data Structure
3. Summary of Our Analysis Journey
4. Key Discoveries
5. Identity Progression Explained
6. Detailed Identity Breakdown

1️⃣ WHAT SCOTT PIPER CREATED

flaws.cloud - An AWS Security CTF (Capture The Flag)

Creator: Scott Piper (security researcher at Summit Route)
Launch Date: February 2017
Purpose: Educational - teach people about common AWS security mistakes
Format: 6 progressive security challenges

The CTF Design:

```
Level 1: Public S3 Bucket Discovery
├─ Find that flaws.cloud is hosted on S3
├─ List bucket contents publicly
└─ Find secret file with next level URL

Level 2: Authenticated AWS Access
├─ S3 bucket allows ANY authenticated AWS user
├─ Use your own AWS account to access
└─ Find secret file

Level 3: Git Repository Secrets
├─ S3 bucket contains .git directory
├─ Search git history for AWS credentials
└─ Find Level3 access keys in old commits

Level 4: EC2 Snapshot Forensics
├─ Public EBS snapshot contains credentials
```

```
├─ Copy snapshot to your account
├─ Mount and extract Level5 credentials
└─ Find access keys in setup scripts

Level 5: EC2 Metadata SSRF
├─ Vulnerable web proxy allows SSRF
├─ Access EC2 metadata service (169.254.169.254)
├─ Steal EC2 instance role credentials
└─ Get Level6 permissions

Level 6: Over-Privileged IAM User
├─ Level6 has SecurityAudit policy (read-only to everything)
├─ Enumerate all AWS resources
├─ Find Lambda functions, API Gateways
└─ Discover final secret
```

AWS Infrastructure Setup:

```
AWS Account: 811596193553
Domain: flaws.cloud
CloudTrail: Enabled (logging all API calls)

IAM Users Created:
├─ backup (high-privilege account)
├─ Level5 (from snapshot)
├─ Level6 (SecurityAudit policy)
└─ flaws (root account - owner)

IAM Roles Created:
├─ flaws (for AWS services)
├─ Level6 (can be assumed)
└─ level5 (for privilege escalation)
```

Key Point: This is an EDUCATIONAL environment. Credentials were INTENTIONALLY leaked as part of the challenge!

2️⃣ EXAMPLE DATA STRUCTURE

CloudTrail Event Structure:

```
json
```

```
{
  "eventTime": "2017-02-12T19:57:06Z",
  "eventName": "ListBuckets",
  "eventSource": "s3.amazonaws.com",
  "awsRegion": "us-east-1",
  "sourceIPAddress": "255.253.125.115",
  "userAgent": "[S3Console/0.4]",
  "userIdentity": {
    "type": "Root",
    "principalId": "811596193553",
    "arn": "arn:aws:iam::811596193553:root",
    "accountId": "811596193553"
  },
  "requestParameters": null,
  "responseElements": null,
  "errorCode": null
}
```


Key Fields We Analyzed:

Field	Purpose	Example Values
eventTime	When the action happened	2017-02-12T19:57:06Z
eventName	What AWS API was called	ListBuckets, RunInstances, AssumeRole
eventSource	Which AWS service	s3.amazonaws.com, ec2.amazonaws.com
userIdentity.type	Type of identity	IAMUser, Root, AssumedRole, AWSService
userIdentity.userName	Username (if IAMUser)	Level5, Level6, backup
userIdentity.principalId	Unique identifier	AIDAI4VF... (user), AROA... (role)
requestParameters	Action details	Contains role ARN for AssumeRole
errorCode	If action failed	AccessDenied, UnauthorizedOperation
sourceIPAddress	Origin IP	255.253.125.115

Dataset Overview:

Total Events: 1,939,207
Time Period: Feb 12, 2017 → Oct 7, 2020 (1,333 days / 3.6 years)
File Size: 20 compressed .json.gz files
Total Data: ~1.2 GB compressed

3️⃣ SUMMARY OF OUR ANALYSIS JOURNEY

Phase 1: Initial Exploration

✅ Loaded 1.9M CloudTrail events
✅ Examined data structure (14 fields per event)
✅ Identified key fields for analysis
✅ Extracted username from userIdentity

Phase 2: User Distribution Analysis

✅ Found 117 unique users/identities
✅ Identified main players:
- backup: 915,834 events (47.2%)
- Level6: 905,082 events (46.7%)
- Level5: 39 events (0.002%)
- flaws: 3,372 events (0.17%)

✅ Discovered identity types:
- IAMUser: 94.14%
- AWSService: 2.97%
- AssumedRole: 2.18%
- Root: 0.57%

Phase 3: Understanding Identity Confusion

✅ Clarified difference between:
- IAM Users (permanent credentials)
- IAM Roles (temporary credentials)
- Same names can exist as both!

✅ Example: "Level6"
- Level6 IAM User: 905,082 events (direct credential use)
- Level6 IAM Role: 5,333 role assumptions (temporary access)
- These are DIFFERENT things!

Phase 4: Graph Analysis (AssumeRole)

✅ Analyzed 79,322 AssumeRole events
✅ Successfully parsed 62,381 edges (78.6%)
✅ Found 21 unique source→target combinations

Key Edges:
- Unknown → flaws role: 44,315 times (AWS services)

- Unknown → Level6 role: 5,333 times
- backup → NULL: 11,603 times (failed, no parameters)
- Level6 → NULL: 5,268 times (failed, no parameters)

❌ Limited role-based privilege escalation
✅ Most activity uses DIRECT IAMUser credentials!

Phase 5: Temporal Analysis

✅ Established chronological timeline
✅ Identified time gaps between user appearances
✅ Discovered monthly activity patterns
✅ Found August 2019 explosion (70% of dataset!)

Phase 6: The August 2019 Discovery

🔥 BREAKTHROUGH FINDING:
- August 21-23, 2019: 1.3M events in 3 days
- 96.4% were RunInstances attempts
- 97.5% failure rate
- Both Level6 and backup equally active
- Sudden start, sudden stop

4️⃣ KEY DISCOVERIES

Discovery 1: Levels 1-4 Are Missing

❌ Level1: 0 events (external public access, not logged)
❌ Level2: 0 events (external accounts, not in this account)
❌ Level3: 0 events (credentials barely used)
❌ Level4: 0 events (snapshot access, minimal logging)
✅ Level5: 39 events (first appearance in logs)
✅ Level6: 905,082 events (MASSIVE activity)

Why? Levels 1-3 happen OUTSIDE the AWS account, so they don't appear in CloudTrail.

Discovery 2: IAMUser Activity Dominates

94.14% of events are IAMUser type
This means: People are using DIRECT access keys, not assuming roles

Level6 IAMUser: 905,082 events
backup IAMUser: 915,834 events

Level5 IAMUser: 39 events

These are PERMANENT credentials being used directly!

Discovery 3: backup Appeared FIRST

Timeline:
Feb 12, 2017: backup first activity
Feb 19, 2017: Level5 first activity (7 days later)
Feb 26, 2017: Level6 first activity (6 days later)

This means backup is NOT the escalation target!
backup is either:
1. Infrastructure/management account
2. Separate parallel compromise

Discovery 4: The "flaws" Root Account Mystery

flaws (Root) first appeared: September 18, 2017
That's 218 days AFTER the CTF launched!

Theories:
1. Root account wasn't used until then (good security practice)
2. CloudTrail logging for root was enabled later
3. Root events are in a different log file
4. Scott Piper only needed root for specific tasks

Discovery 5: "Unknown" = AWS Services

Unknown events: 57,617
├─ 99.998% are AWSService type
└─ 0.002% are truly unknown

These are NOT CTF players - they are:
- AWS Config (security scanning)
- Lambda functions
- CloudTrail itself
- Other AWS infrastructure

Discovery 6: August 2019 Explosion

🔥 THE SMOKING GUN:

Aug 1-20, 2019: Normal (~500-3K events/day)
Aug 21, 2019: 421,560 events ⚡ EXPLOSION STARTS
Aug 22, 2019: 591,688 events ⚡ PEAK
Aug 23, 2019: 301,547 events ⚡ CONTINUES
Aug 24-31, 2019: Back to normal (back to ~100-8K/day)

What happened:
- Level6 & backup both active (50/50 split)
- 96.4% RunInstances (trying to launch EC2 instances)
- 97.5% failure rate (AWS blocked them)
- 3-day burst, then stopped

Theory: Automated security scanner or testing tool

Discovery 7: High Error Rate

Overall error rate: 77% of all events failed

Why?
- AWS rate limiting (too many requests)
- Permission errors (SecurityAudit = read-only)
- Resource quotas exceeded
- Service capacity limits

This indicates AUTOMATED MASS SCANNING!

5️⃣ IDENTITY PROGRESSION EXPLAINED

The Intended CTF Path:

THEORETICAL PROGRESSION (What Should Happen):

Level 1 → Level 2 → Level 3 → Level 4 → Level 5 → Level 6
↓ ↓ ↓ ↓ ↓ ↓
Public Access | External AWS User | Git Creds | Snapshot Forensic | EC2 SSRF | Security Audit

The ACTUAL Progression (What We See in Data):

REAL TIMELINE:

Feb 12, 2017: backup account active
├─ First activity: ListBuckets

├─ Identity: IAMUser
├─ Role: Unknown (infrastructure? or compromise?)
└─ Activity: 915,834 events over 3.6 years

↓ 7 days gap

Feb 19-23, 2017: Level5 reconnaissance
├─ First activity: DescribeLoadBalancers
├─ Identity: IAMUser (using access keys directly)
├─ Events: 39 total over 4 days
├─ Actions: IAM enumeration, policy inspection
│ - ListPolicies
│ - GetUser
│ - ListAttachedUserPolicies
│ - GetPolicyVersion
│ - ListFunctions (Lambda)
│ - GetStages (API Gateway)
└─ Behavior: MANUAL reconnaissance (human typing)

↓ 6 days gap

Feb 26, 2017 - Oct 7, 2020: Level6 explosion
├─ First activity: GetUser
├─ Identity: IAMUser (using access keys directly)
├─ Events: 905,082 over 1,318 days (3.6 years)
├─ Peak: August 21-23, 2019 (700K events in 3 days)
├─ Actions:
│ - StartInstances: 778,426 (86% of Level6)
│ - DescribeImages: 19,554
│ - DescribeSnapshots: 6,709
│ - AssumeRole: 5,268 attempts
└─ Behavior: AUTOMATED mass scanning

Sep 18, 2017 onwards: flaws (Root) appears
├─ First activity: DescribeLoadBalancers
├─ Identity: Root (AWS account owner)
├─ Events: 3,372 over 1,114 days
├─ Actions: Management tasks
│ - GetBucketLocation
│ - DescribeAccountAttributes
│ - GetBucketAcl
└─ Role: Scott Piper infrastructure management

The Gap Analysis:

MISSING LEVELS (Why they don't appear):

Level 1: Public S3 Access
├─ No authentication required
├─ Happens OUTSIDE the AWS account
└─ CloudTrail doesn't log external public access
Result: 0 events ❌

Level 2: External AWS Account
├─ Users use THEIR OWN AWS accounts
├─ Might appear as "AWSAccount" type (2,628 events)
└─ CloudTrail logs show external account IDs, not usernames
Result: Possibly in "AWSAccount" type ⚠️

Level 3: Git Credentials
├─ Level3 IAMUser credentials found in git history
├─ Likely revoked quickly or never widely used
└─ Some usage might be in "Unknown" or other identities
Result: Very limited evidence ❌

Level 4: Snapshot Access
├─ Public snapshot contains Level5 credentials
├─ DescribeSnapshots calls happen
└─ But actual snapshot access happens in attacker's account
Result: Only Level5 activity visible ⚠️

Level 5: Visible! ✅
├─ 39 events from Feb 19-23, 2017
├─ Clear reconnaissance behavior
└─ Led to Level6 discovery

Level 6: Visible! ✅
├─ 905,082 events over 3.6 years
├─ Massive automated activity
└─ Clear exploitation

6️⃣ DETAILED IDENTITY BREAKDOWN

A. "backup" - The Infrastructure Account

PROFILE:

```
├─ Type: IAMUser (permanent credentials)
├─ First seen: Feb 12, 2017 21:15:12
├─ Last seen: Oct 7, 2020 21:03:30
├─ Active period: 1,332 days (3.6 years)
├─ Total events: 915,834 (47.2% of dataset)
└─ Error rate: 81.1% (742,447 failures)

BEHAVIOR SIGNATURE:
├─ RunInstances: 640,767 attempts (69.97%)
│  └─ 81% failed (rate limiting, permissions)
├─ DescribeSnapshots: 99,301 (10.84%)
├─ GetBucketAcl: 36,186 (3.95%)
├─ AssumeRole: 11,603 (1.27%) → All failed (NULL parameters)
└─ Pattern: Automated mass scanning

TIMELINE MYSTERY:
├─ Appeared 14 days BEFORE Level6
├─ Same scanning pattern as Level6
└─ Likely NOT the escalation target

THEORIES:
1. Infrastructure Account (Most Likely)
   ├─ Scott Piper's admin/management account
   ├─ High privileges for CTF setup
   └─ Name "backup" suggests administrative role

2. Parallel Compromise
   ├─ Credentials leaked separately
   ├─ Same automated tools used
   └─ Similar behavior to Level6

3. Honeypot
   ├─ Intentional trap for attackers
   └─ Monitoring who tries to use it
```

B. "Level5" - The Scout

```
PROFILE:
├─ Type: IAMUser (permanent credentials)
├─ First seen: Feb 19, 2017 23:38:17
├─ Last seen: Feb 23, 2017 04:09:21
├─ Active period: 4 days
├─ Total events: 39 (0.002% of dataset)
└─ Error rate: 7.7% (3 errors)
```

COMPLETE ACTIVITY LOG (All 39 Events):

Day 1 (Feb 19, 23:38-23:40):
├─ DescribeLoadBalancers (checking for load balancers)
├─ ListFunctions (Lambda enumeration)
└─ GetPolicy (checking Lambda policies)

Day 2 (Feb 20, 00:27-02:09):
├─ ListDistributions (CloudFront check)
├─ IAM Reconnaissance (15 events):
│ ├─ ListPolicies (multiple times)
│ ├─ ListAccessKeys
│ ├─ GetUser (checking own permissions)
│ ├─ GetAccountSummary
│ ├─ ListAttachedUserPolicies (key discovery!)
│ ├─ GetUserPolicy
│ ├─ GetPolicy
│ └─ GetPolicyVersion (reading full policy)
└─ API Gateway Investigation:
   ├─ GetResources → AccessDenied
   ├─ GetStages (successful)
   └─ GetRestApis → AccessDenied

Day 5 (Feb 23, 03:59-04:09):
└─ Repeat reconnaissance (verification)

BEHAVIOR ANALYSIS:
├─ Manual reconnaissance (time gaps = human typing)
├─ Systematic permission checking
├─ IAM policy inspection (looking for SecurityAudit)
├─ Service enumeration (Lambda, API Gateway, CloudFront)
└─ Found Level6 permissions → escalated

WHY ONLY 39 EVENTS?
├─ Careful exploration, not brute-force
├─ Human operator (CTF participant)
├─ Once found SecurityAudit policy → moved to Level6
└─ This is TEXTBOOK CTF behavior

C. "Level6" - The Mass Scanner

PROFILE:
├─ Type: IAMUser (permanent credentials)

├─ First seen: Feb 26, 2017 23:13:33
├─ Last seen: Oct 7, 2020 14:52:39
├─ Active period: 1,318 days (3.6 years)
├─ Total events: 905,082 (46.7% of dataset)
└─ Error rate: 73.6% (665,899 failures)

BEHAVIOR BREAKDOWN:

Phase 1 (Feb 26 - Jul 2019): Steady Scanning
├─ 2,000-40,000 events/month
├─ Mix of manual and automated
└─ Credentials shared in CTF community

Phase 2 (Aug 21-23, 2019): THE EXPLOSION 🔥
├─ 700,148 events in 3 days (77.4% of Level6 activity!)
├─ 96% RunInstances attempts
├─ 97.5% failure rate
└─ Automated tool gone rogue

Phase 3 (Aug 24, 2019 - Oct 2020): Sustained Scanning
├─ 10,000-30,000 events/month
└─ Continued automated testing

TOP ACTIONS:
├─ StartInstances: 778,426 (86.00%)
│ └─ Trying to launch EC2 instances
├─ DescribeImages: 19,554 (2.16%)
├─ DescribeSnapshots: 6,709 (0.74%)
├─ GetBucketAcl: 5,571 (0.62%)
├─ AssumeRole: 5,268 (0.58%) → Failed (NULL parameters)
└─ DescribeInstances: 5,201 (0.57%)

WHAT THIS MEANS:
Level6 has SecurityAudit policy = READ-ONLY access
├─ Can DESCRIBE/LIST everything
├─ Cannot CREATE/MODIFY/DELETE
└─ StartInstances fails because it's a WRITE operation

WHY SO MANY RunInstances ATTEMPTS?
├─ Automated security scanner
├─ Testing permission boundaries
├─ Misconfigured script
└─ 86% failure rate = hitting AWS limits

CREDENTIAL LIFECYCLE:
Feb 26, 2017: First use (legitimate CTF)
↓
Credentials shared in CTF community
↓
Posted on GitHub/forums/tutorials
↓
Automated tools pick them up
↓
Aug 21-23, 2019: Mass exploitation
↓
Oct 7, 2020: Last recorded activity

D. "flaws" - The Root Account

PROFILE:
├─ Type: Root (AWS account owner)
├─ First seen: Sep 18, 2017 22:53:01 (218 days after CTF launch!)
├─ Last seen: Oct 7, 2020 18:32:38
├─ Active period: 1,114 days (3.0 years)
├─ Total events: 3,372 (0.17% of dataset)
└─ Error rate: Low (management tasks)

TOP ACTIONS:
├─ GetBucketLocation: 304
├─ DescribeAccountAttributes: 289
├─ GetBucketAcl: 189
├─ GetBucketEncryption: 119
├─ GetTrailStatus: 117
└─ Pattern: Infrastructure management

ROLE:
├─ Scott Piper's account owner access
├─ Used for CTF infrastructure setup/maintenance
├─ NOT part of the attack chain
└─ Good security practice (root rarely used)

WHY IT APPEARED LATE:
Theory 1: Root wasn't needed until Sept 2017
Theory 2: CloudTrail logging enabled later
Theory 3: Different log files for root

E. IAM Roles (The Temporary Credentials)

```
THREE KEY ROLES:

1. "flaws" Role
   ├─ Assumed 44,315 times
   ├─ Source: Unknown (AWS services)
   ├─ Purpose: AWS infrastructure automation
   ├─ NOT used by CTF attackers
   └─ Example: AWS Config, Lambda execution

2. "Level6" Role
   ├─ Assumed 5,333 times
   ├─ Source: Unknown (external users)
   ├─ Purpose: Temporary Level6 access
   └─ Different from Level6 IAMUser!

3. "level5" Role
   ├─ Assumed 1 time (very rare!)
   ├─ Source: Unknown
   └─ Possible CTF progression path

NULL TARGETS (Failed Assumptions):
   ├─ backup → NULL: 11,603 attempts (all failed)
   ├─ Level6 → NULL: 5,268 attempts (all failed)
   └─ These are failed AssumeRole calls (no parameters logged)
```

🎯 THE COMPLETE STORY

What Scott Piper Created:

An educational AWS security CTF to teach people about cloud security mistakes.

What Actually Happened:

1. **Feb 12, 2017:** CTF launches with intentionally vulnerable AWS account
2. **Feb 19-23, 2017:** First CTF participants find Level5 credentials
3. **Feb 26, 2017:** Level5 → Level6 progression (7-day gap = methodical)
4. **Mar 2017 - Jul 2019:** Steady CTF activity, credentials shared
5. **Aug 21-23, 2019:** 🔥 EXPLOSION - automated tool uses credentials
6. **Aug 24, 2019 - Oct 2020:** Sustained automated scanning continues

7. **Throughout:** backup account also heavily used (infrastructure or parallel attack)

The Key Insight:

This dataset shows the LIFECYCLE of leaked cloud credentials:

```
Manual Discovery → Community Sharing → Automation Adoption → Mass Exploitation
   (Level5)        (Early Level6)     (Mid 2019)          (Aug 2019)
  39 events        2-40K/month        Growing usage       1.3M in 3 days
```

Research Value:

This is the ONLY publicly available, real-world dataset showing:

- ✅ Cloud credential lifecycle
- ✅ Temporal progression of attacks
- ✅ Transition from manual to automated
- ✅ Impact of credential leakage at scale
- ✅ Behavioral signatures of different attack phases

📊 DATA SUMMARY

```
Total Events: 1,939,207
Time Span: 1,333 days (Feb 12, 2017 - Oct 7, 2020)

Key Players:
├─ backup: 915,834 events (47.2%)
├─ Level6: 905,082 events (46.7%)
├─ Unknown (AWS Services): 57,617 events (3.0%)
├─ flaws (Root): 3,372 events (0.2%)
└─ Level5: 39 events (0.002%)

Identity Types:
├─ IAMUser: 94.14%
├─ AWSService: 2.97%
├─ AssumedRole: 2.18%
├─ Root: 0.57%
└─ AWSAccount: 0.14%

Top Actions:
├─ RunInstances: 1,421,193 (73.3%)
├─ DescribeSnapshots: 105,198 (5.4%)
```

```
├─ GetBucketAcl: 42,135 (2.2%)
├─ DescribeImages: 20,088 (1.0%)
└─ DescribeInstances: 18,489 (1.0%)

Error Rate: 77% overall
Peak Activity: August 21-23, 2019 (1.3M events in 3 days)
```

🎯 NEXT STEPS FOR RESEARCH

Based on this analysis, the recommended research directions are:

1. **Credential Leakage Detection**
   - Temporal anomaly detection (Aug 2019 spike)
   - Behavioral change analysis (manual → automated)
   - LLM-based narrative generation
2. **Attack Evolution Analysis**
   - Phase characterization (recon → exploit → automation)
   - Tool fingerprinting (user agents, patterns)
   - Credential lifecycle modeling
3. **Cloud Forensic Timeline Reconstruction**
   - LLM-enhanced event summarization
   - Natural language attack narratives
   - Analyst Q&A systems

End of Summary