Gazzang

ApacheCon North America

Portland, Oregon 26th – 28th February 2013

Reining in Security Sprawl

ApacheCon NA 2013
Wednesday, February 27, 2013

Dustin Kirkland
CTO, Gazzang

Health Record Compromises

- US Department of Health and Human Services
  - 480 reported breaches
    - between September 2009 - September 2012
  - 21 million American health records exposed
  - Of the 480 incidents, the reasons given:
    - **55% - Theft of device or physical media**
    - 26% - Hacking or unauthorized access
    - **12% - Lost devices, disks, tapes, drives, media**
    - **5% - Improper disposal of media**
    - 3% - Other / unspecified
  - 55% + 12% + 5% = **72%**
    - trivially solved by comprehensive data encryption at rest

Student Record Compromises

- PrivacyRights.org
  - Since **2010**, more than **three million student records** have been compromised due to attacks, lost, stolen or missing files
  - In **2012** alone...
    - 23,000 SSN’s breached at the University of North Florida
    - 16,000 SSN’s, birth dates and student ID’s breached from Eugene, Oregon school district
    - 650,000 records breached from University of Nebraska
    - 350,000 records from UNC Charlotte

What is different about security in the cloud?

- Security gets outsourced, distributed, consolidated
  - Cloud means giving up usual physical and logical barriers
- Automation and orchestration
  - Non-password, non-interactive, automated authentication
- Numbers game
  - Exponential increase in the sheer number of system instances and data stored in the cloud
- Randomness
  - Some current cloud practices render virtual machines inherently predictable
    - *See my other presentation about lack of entropy**

What is different about security in big data?

- Big Data is exploding
  - Petabyte scale!
- Keys and certificates along with it
  - Exponential increase in keys, certs, and tokens
- Diverse projects and organizations
  - Any one company might have a dozen big data projects
- Hence, security sprawls
  - Policies in place to guide the security of these projects?
- Case studies across verticals
  - Health care, education, finance, SaaS vendors

Sensitive data proliferates

- More OS environments means more...
  - Private keys
    - SSH, SSL, Kerberos, GPG
  - Configuration files
    - /etc/*
  - Log files
    - /var/log/*
  - Application data
    - /var/lib/*
  - User data
    - /home/*
  - Password hashes
  - Machine DNA

What you can do

*Always **encrypt local** data at rest*

*Always **encrypt network** data in motion*

*Always **protect keys** and certificates*

*Always **monitor and log** meticulously*

Always Encrypt Local Data at Rest

- Unfortunately, disks do disappear
  - Refer back to the health record and student compromises
- By law, some are required to encrypt data at rest
  - Heath, finance, academic
- Encryption at rest
  - File level
    - eCryptfs
  - Block level
    - dm-crypt
- Performance concerns?
  - Leverage AES-NI for hardware acceleration

** Commercial support for **zNcrypt** available from Gazzang*

Always Encrypt Network Data in Motion

- Many Apache projects have native SSL/TLS support
  - But too often, SSL is not enabled!
- Certificate management is actually pretty hard
  - Expensive, if you buy commercial CA-signed certificates
    - Which doesn't scale economically to cloud or big data size
  - Self-signed certificates can still be secure
    - If you can pre-share your own CA at deployment
  - PKI is essential
  - Ensure you have sufficient entropy to generate high quality, secure certificates

** Commercial support for **zTrustee** available from Gazzang*

Always Protect Keys and Certificates

- Encrypted data is only secure if keys are stored somewhere else
    - Separate from the encrypted data
- Avoid storing keys and certificates
    - On disk
    - Hard coded within applications
- Retrieve keys and certificates dynamically at time-of-use
    - Store in secure memory
    - Discard when done
    - Institute policies that guard retrieval

** Commercial support for **zTrustee** available from Gazzang*

Always Monitor and Log Meticulously

- Monitor and log your applications
  - Meticulously!
  - Some services have native monitoring/logging tools
  - For others, you'll need third party applications
- Audit those logs
  - Live, in real time
  - Run analytics, after-the-fact
- Alert on aberrant behavior
  - Live, in real time

** Commercial support for **zOps** available from Gazzang*

Gazzang Protects Customers Running

- Apache Projects
  - Accumulo
  - Cassandra
  - Couch
  - Hadoop, Hbase, Hive, Pig
  - Tomcat
  - Solr
- As well as
  - GlusterFS
  - MongoDB
  - MySQL
  - PostgreSQL
  - Riak

Customer Case Study #1

- Global mobile device company
  - Streaming and storing
    - User credential data
    - Includes personally identifiable information (PII)
    - Device usage patterns
  - Suite of applications analyze and operate against this data
    - Backup/restore device state
    - Maps, traffic patterns, targeted advertising
  - Company uses Gazzang to protect all data at rest
    - Huge Cassandra and Hadoop cluster
    - On-premises data center

Customer Case Study #2

- SaaS identity management provider
  - Massive security sprawl without management or policy
    - Passwords
    - Operating system data
    - Keys and certificates
  - Company uses Gazzang within their SaaS application
    - Encrypt sensitive information in their big data storage
    - Centrally manage their keys and certificates
      - Linked Gazzang's zTrustee Java API
    - Public cloud application

Customer Case Study #3

- Health care SaaS vendor
  - Archive and retain huge data sets
    - Patient records
    - Demographic information
    - Billing information
  - Used to run analytics
    - For physicians
    - For health plan administrators
  - Gazzang helps this organization
    - Meet HIPAA and PCI requirements
      - Data encryption at rest
      - Key management
    - Hybrid public/private cloud environment

Gazzang Thanks the Apache Foundation for

- Apache Cassandra
- Apache Commons
    - lang, collections, io, pool, beanutils, logging, codec, digester
- Apache Hadoop
- Apache HTTP Server
- Apache Maven
- Apache Tiles
- Apache Whirr

Thank you to all of the contributors
and supporters of these projects

About Gazzang

- Who we are
  - Headquartered in Austin, TX
  - Funded by Austin Ventures and Silver Creek Ventures
  - Active in various open source communities
- What we do
  - Encryption, key management, and diagnostics for Cloud and Big Data applications
    - Packaged solutions for Hadoop, Cassandra, MongoDB, OpenStack, Apache, MySQL, PostgreSQL
- Why we do it
  - Help our customers secure sensitive data at rest
    - HIPAA, FERPA, PHI, PCI DSS, PII, etc.

Contact Information

- Dustin Kirkland
  - **CTO** at Gazzang
  - **Author** and **maintainer** of open source projects
  - **Twitter** - @dustinkirkland
  - **Blog** - http://blog.dustinkirkland.com
  - **LinkedIn** - http://www.linkedin.com/in/dustinkirkland
  - **Launchpad** - https://launchpad.net/~kirkland
  - **Github** - https://github.com/dustinkirkland
  - **GPG Public Key**
    - 4096R/F1529469
    - E2D9 E1C5 F9F5 D592 91F4 607D 95E6 4373 F152 9469