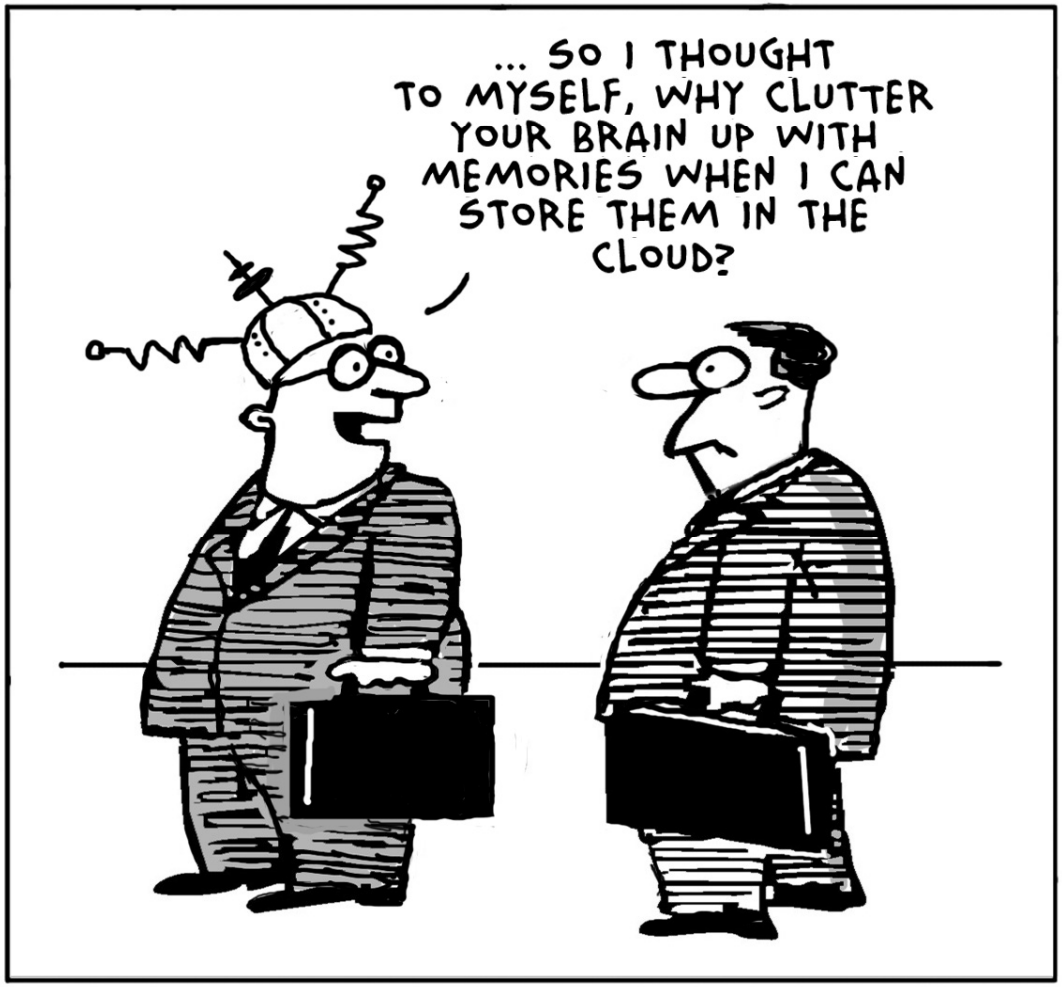
SAND2012-8497C

Exploring Security Implications of the Cloud

Vince Urias, David Zage
Sandia National Laboratories
Cyber Analysis R&D Solutions



Presentation Overview

- 30 Second Overview of Cloud
  - Platforms: *aaS
- Current Threats and Challenges to security professionals
  - Contemporary examples
  - What keeps security professionals up at night
- Deployment challenges and solutions
- Current Research Areas
  - Cloud SA, Full-Stack Analysis
  - Cloud Storage
  - Deployment Strategies

What is Cloud Computing?

- Cloud computing "is a model for enabling convenient, on-demand network access to a shared pool of configurable computing resources (e.g., networks, servers, storage, applications, and services) that can be rapidly provisioned and released with minimal management effort or service provider interaction". - ***NIST***
- Cloud computing push to prominence
  - Executive branch CIO → adopt a "cloud first" policy (25 Point Implementation Plan to Reform Federal Information Technology Management)
  - White House Office of Science and Technology Policy → Big Data Research and Development Initiative
  - DHS → Roadmap for Cybersecurity Research

Is it Really that Popular?

Traditional DFS

- Lustre
- PVFS
- GPFS
- Ceph
- Gluster
- Tahoe
- HDFS

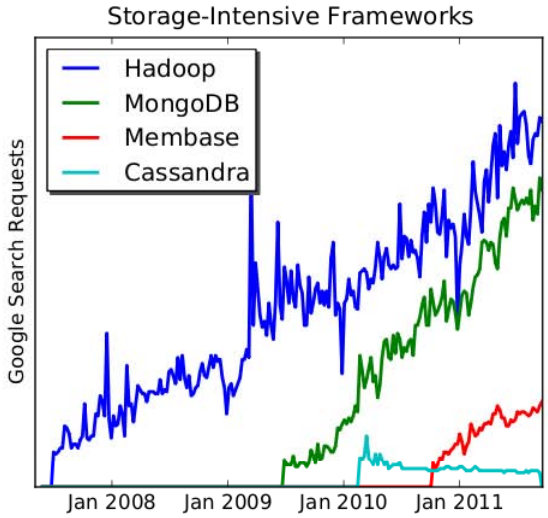
Information Stores

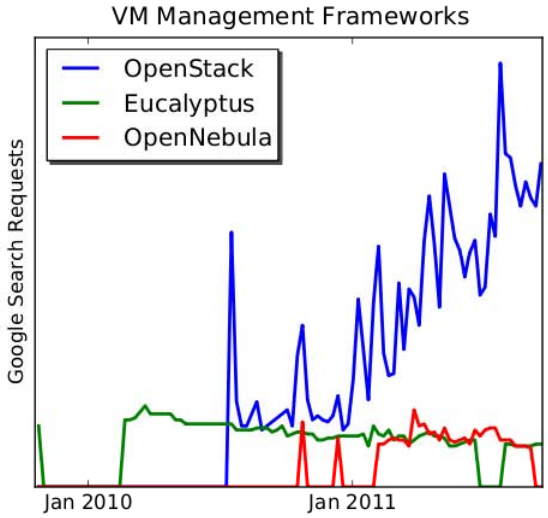
- Cassandra
- MongoDB
- Voldemort
- Membase
- Accumulo
- HBase

VM Stores

- Eucalyptus/Walru
- OpenStack/Swift

Google Trends

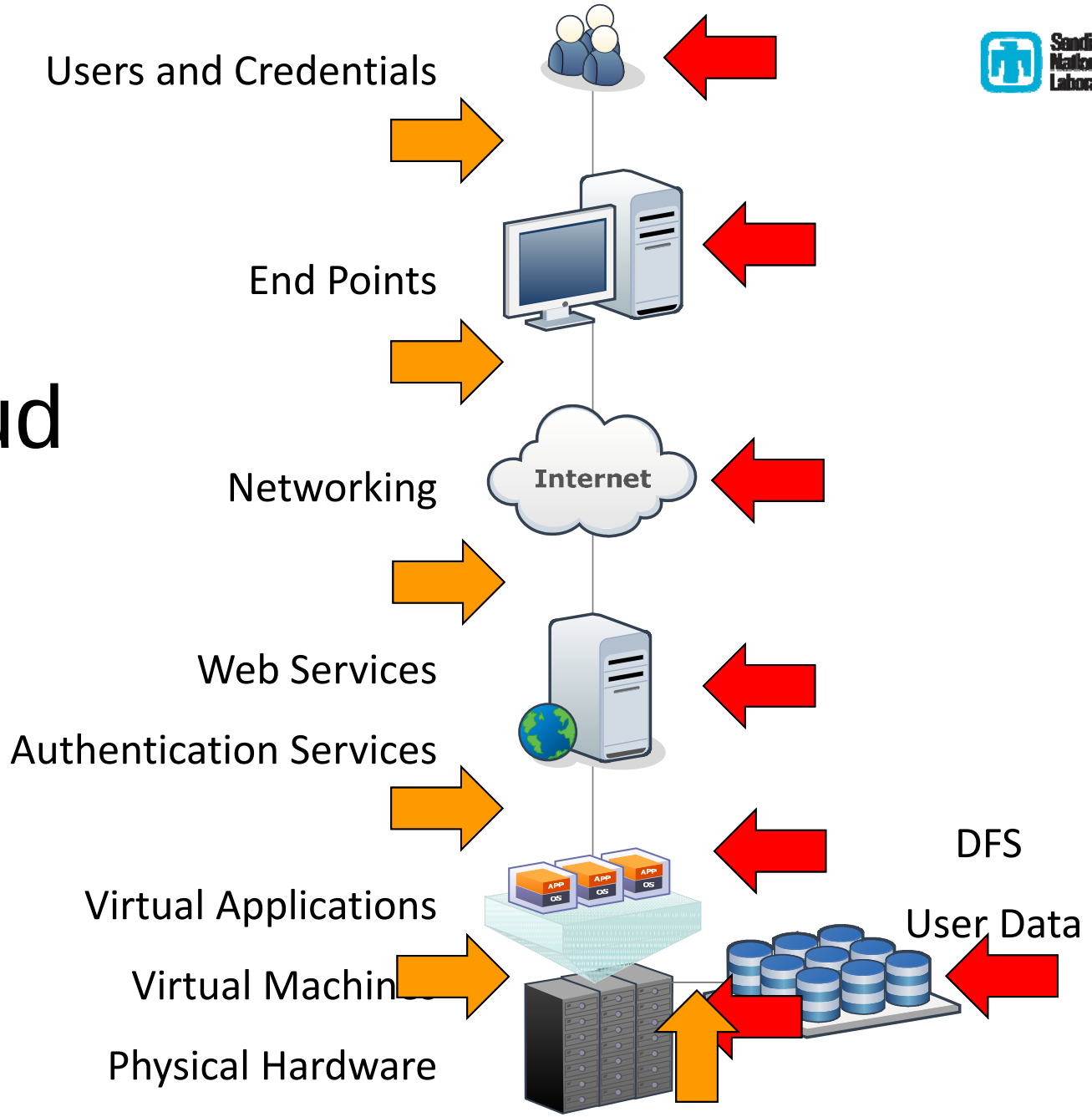




Cloudscape

- Cloud Computing is both immature and decades old
  - Cloud has many links back to early mainframe and grid computing architectures of the early 60s, 70s and 80s
  - Current implementations (Eucalyptus, Openstack, ec2, etc.) are new and going through growing pains
    - Scale, new software, etc.
    - Lots of old vendors and new vendors merging and getting into the business
- Consequence: field is ripe for vulnerabilities and a playground for malicious actors
- Security is still an afterthought
  - Most analysis is on ROI for companies not security
  - Classical paradigm still used: RBAC, encryption
  - It's not enough: scale, heterogeneity, diversity and it's ubiquity are all new and emerging issues for the cloud

The Cloud Stack



Software
Division
The Global Voice of Quality™

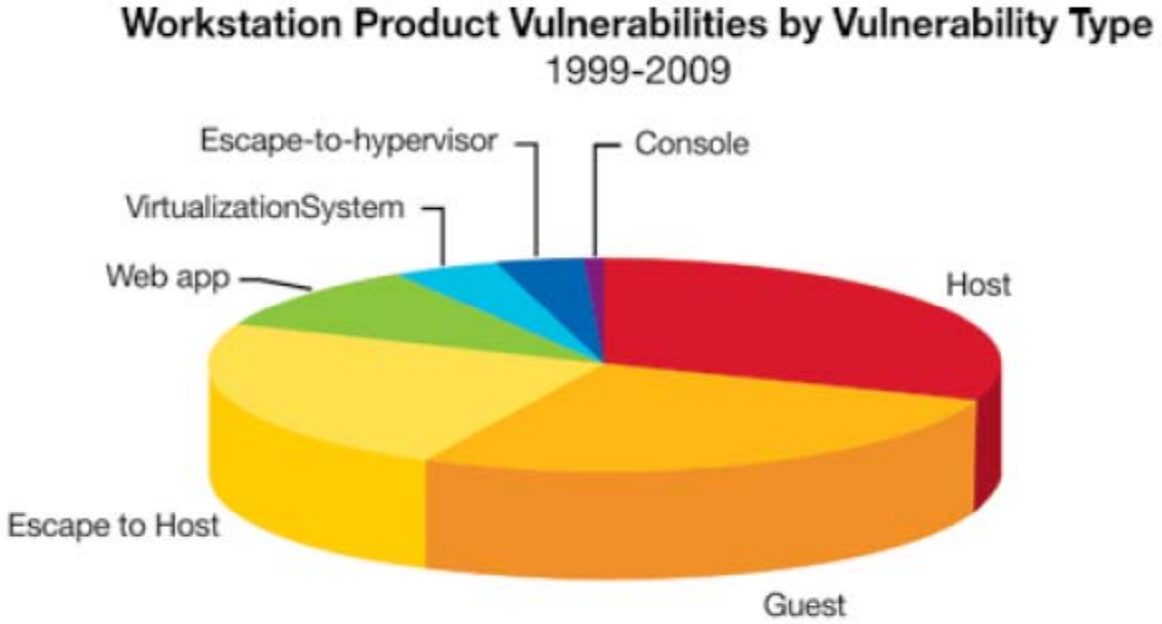
Threats, Vulnerabilities, and Fears… Oh My

Open Source Threats

- So... most of the threats aren't new; they are old, explored, known threats
  - The difference is the impact of the vulnerability
  - AND the pervasiveness the technology has
  - AND the access to all the technologies
- Overview of a variety of threats that the community has faced over the last 3 years
  - Only a sample
  - Provide some context to what corporations (and you as government, DOE) are facing
- These are threats that have been exercised in the wild today

Open Public Threats: VM Focused

- "Hey, you, get off of my cloud" (CCS 2009)
- Cloudburst (Black Hat 2009)
  - Leverage same physical hardware and then do a cross-VM attacks



IBM X-Force® 2010 Mid-Year Trend and Risk Report (373 vulnerabilities)

Open Public Threats: Datastorage Focused

- Linkup (2008)
  - Deleted active customer data
- Nokia (2009)
  - OVI deleted all user data for a 1 month period. Users lost images, contacts, messages, etc.
- Sidekick Data Loss (2009)
- Dropbox Vulnerabilities (2011)
- Megaupload (2012)
  - Megaupload host wants out
    - Carpathia pays $9000/day to keep data

Open Public Threats: Application Focused

- Nothing really new
  - Lots of vulnerability in applications APIs
  - Security is getting better but users still use unsecure APIs for convenience
  - Except that there is an immense reliance on web based APIs
  - Predominant method for accessing and interaction of resources
- Examples:
  - OpenEMR (XSS, SQL Injection), EyeOS (unauthorized access), Google API (XSS), etc…

Hey Where's My Data?

- Data Replication
  - Cross-coast & cross-continent, little insight into location
  - Data migration
  - Data backup
- Data Removal
  - No standard procedures
  - How long does temporary data persist?
- Collocation Issues
  - Lots of companies, users data is intermixing
  - One location to penetrate
- Mosaic problem
  - Can get lots of critical information and user information from the cloud

And I Want Access Now!

- Putting all storage in “the cloud” could have detrimental effects
  - Lots of outages recently including Amazon and Azure where services went down for hours to days which equates to millions in lost revenue
- Examples:
  - VMware causes second outage while recovering from first
    - Cloud Foundry (PaaS)
  - Windows Azure Leap-Year Glitch Takes Down G-Cloud
  - Amazon says it's getting a handle on EC2 outage
  - The Hidden Risk of a Meltdown in the Cloud

Other Areas of Concern

- Lack of visibility
  - Lack of understanding of internal practices
- Credential Stealing
  - Moving the target from the data itself to the endpoint
- Vendor Lock In
  - If the system is out of business, how do you get the data?
  - Fate of data held by Megaupload up in the air
- Cyber Criminals and the Cloud
  - Exploit As A Service
  - DDoS
- Insider Threat

Software
Division
The Global Voice of Quality™

Did Someone Say Solutions?

Better Infrastructure Deployment and Management

- Cloud and it's applications are hard to deploy
  - Tens of services running
  - Tens numerous layers of abstraction between the applications and the administration
  - Tens of choke points that could be failures
  - Systems hard to debug and understand
  - System administrators pray that it just works at all
- Clouds tend to be easy from the users perspective
  - The consequence is it's hard for administrators to support, secure the versatility and flexibility of the cloud
- Need to streamline the process

Streamlining The Process

- Leveraging several open-source initiatives, it is possible to create a streamlined, logged automated install of IaaS platforms
  - From bare metal -> net install -> operating system install -> service install (Openstack) -> application configurations -> user interaction w/ system
- From instantiation of the cloud, leverage and create an environment to use and effectively manage and gather the state of the cloud
- Take out the guess work and ad-hoc configurations of the cloud
- Create standard secure builds that have been vetted, tested, and repeatable

Security and the Cloud

- The cloud has been great for monetizing computing capabilities to cents per compute cycle
  - Industry has a monetizing availability of these systems
  - However other parts of security go unanswered, there is limited incentive for them to invest more into confidentiality or integrity of the files and systems used
- The goal of our research and the laboratory is to augment the capabilities, help prototype and look at next-generation solutions, not reinvent the wheel
- Need new ways of incentivizing and testing ***Confidentiality*** and ***Integrity*** of the systems

Full Stack Analysis

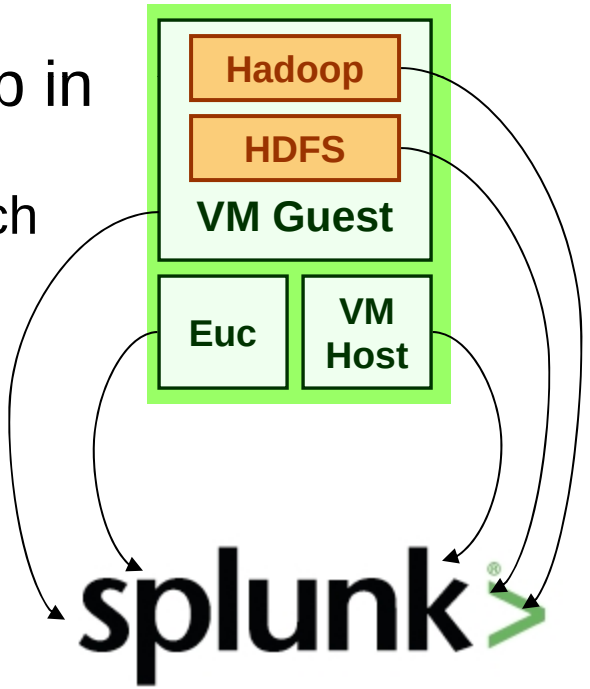
- Cloud has interesting consequences for the CND analyst
  - Previously the analyst had limited to no control of the env
    - Differing hardware, software, packages, switches, routers, etc
  - Caused different tools to used, different techniques for collection for limited insight into the space
- NOW the cloud, the CND analyst has full control over everything
  - The link, the machine, the hardware its all running on
- CND analysts MUST leverage this insight
  - However tools, analytics must be developed

One Solution Being Explored

- The entire enterprise comes down to packets and logs
  - Packets for the network connections and communications that have been created
  - Logs from all the systems in the eco-system
    - The bare metal operating system hosting the cloud services
    - Applications of the cloud
    - Virtual machine manager
    - Virtual machines (and OS) running in the cloud
    - Application being hosted by the virtual machine
- Each of these logs and packets can be synthesized and analyzed to provided
  - Context of the applications
  - Situational Awareness
  - Detection
  - Prevention

About the problem

- How do we detect hostility in the cloud and its storage?
  - Clusters are complex systems, hard to decipher activities
- Eucalyptus/Openstack Cluster w/Hadoop in VMs → Splunk
  - Capture: Hadoop, VM Guest, VM Host, Switch
  - Splunk server stores all, interrogates
  - Dashboards enable easy user queries
- Next: Automated analysis
  - Leverage machine learning
  - Outlier identification tools
  - Splunk or other DFS data to outlier tool's input form



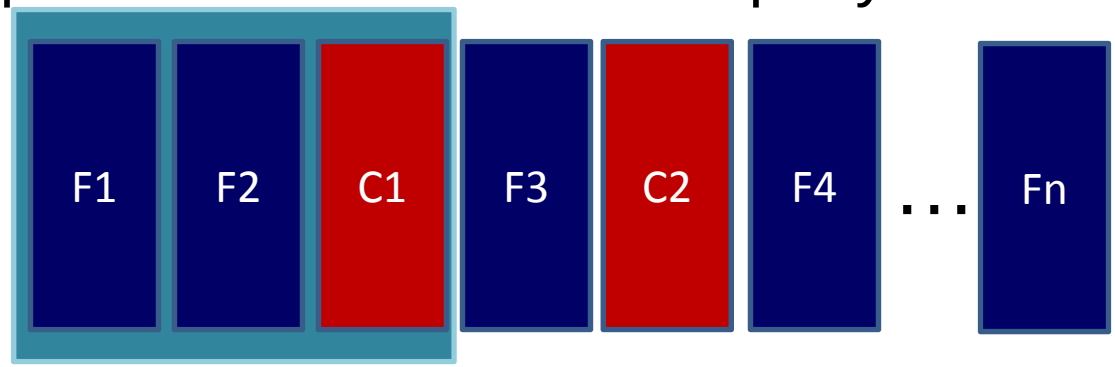
Software
Division
The Global Voice of Quality™

Wheat and Chaff - What Does Farming Have to do With Clouds?

Sandia
National
Laboratories

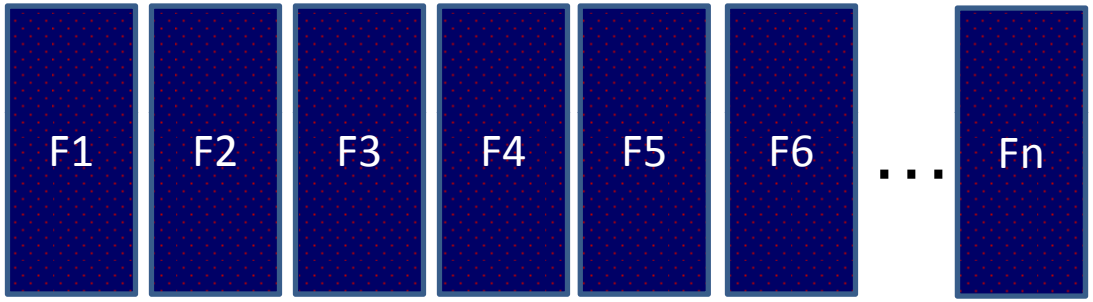
Using Algebraic Subspaces - Wheat and Chaff (W&C)

- How do I know that my Service Level Agreement (SLA) is being followed by the provider?
  - Challenge/response, store hashes, etc.
- Another option: Store a percentage of legitimate-looking bogus data (chaff) in conjunction with the good data (wheat)
  - Expect % of chaff with each query



Improved W&C

- Naïve solution ends storing many pieces of chaff
- Better: Incorporate chaff into the file

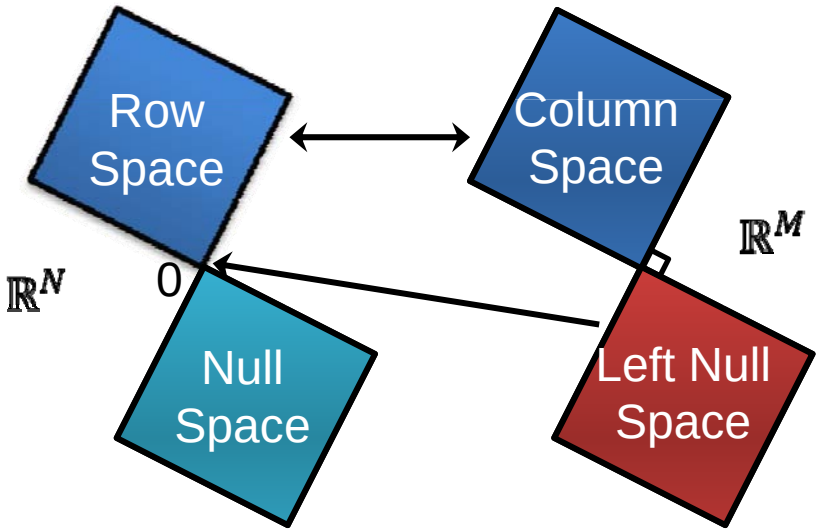


Problems:

- How is the chaff created?
- How is the chaff winnowed out?
- How do we ensure the data is a protected from a nosey provider?

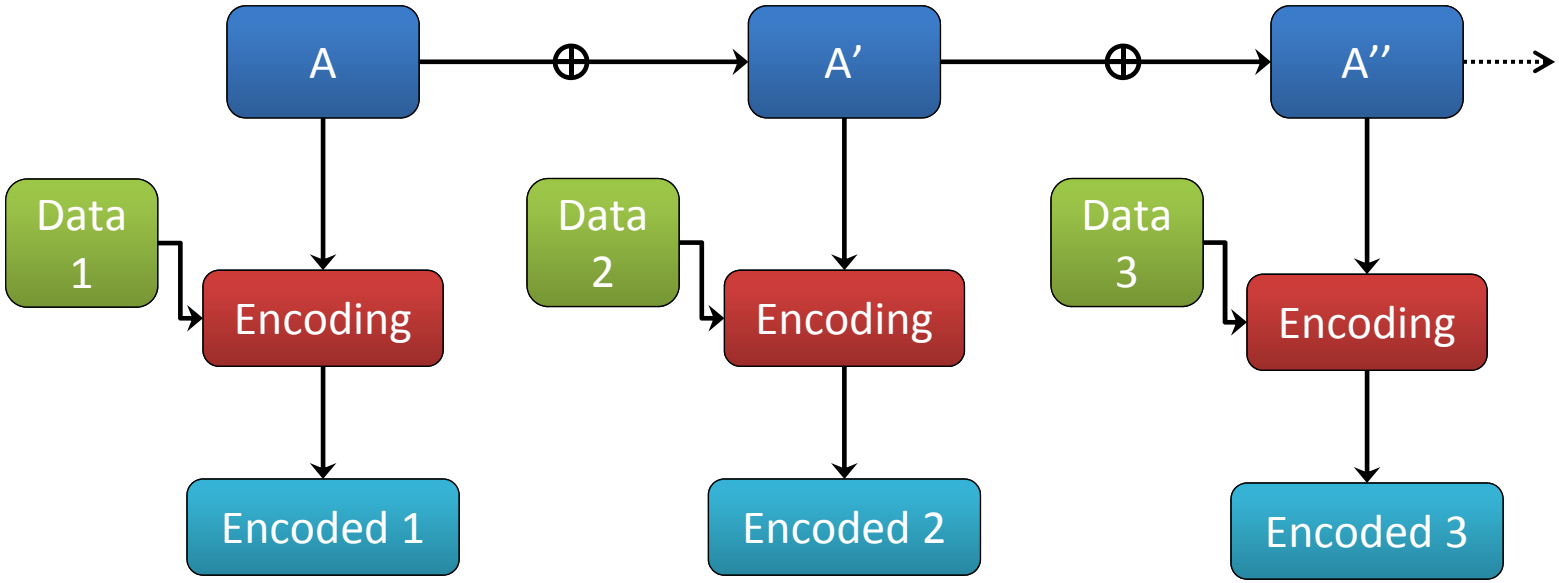
W&C Algorithm Sketch

- Treat data as Matrix *D*
- Generate a random coding matrix *A*
  - *Encoded = A * D*
- Chaff is derived from the left null space of A
- Incorporate bits of chaff into *Encoded* data



Improving Encoding Performance

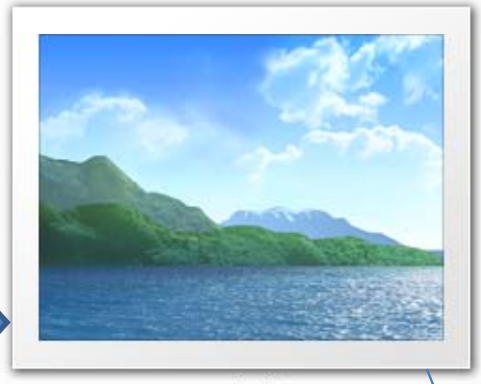
- Large amounts of data can be costly to manipulate (time, memory, processing)
- *Matrix Block Chaining:* Partition the data into smaller chunks and encode
  - Avoids creating multiple encoding matrices



Visual Results

- Before

Encoded



encoded.jpg

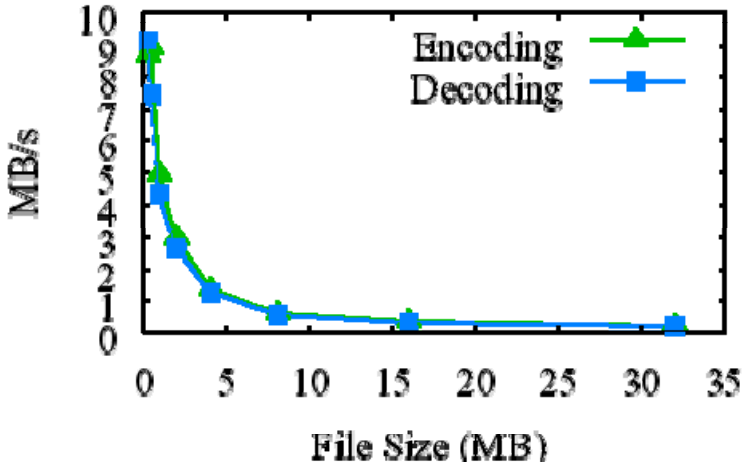
crypto.jpg

- Decoded

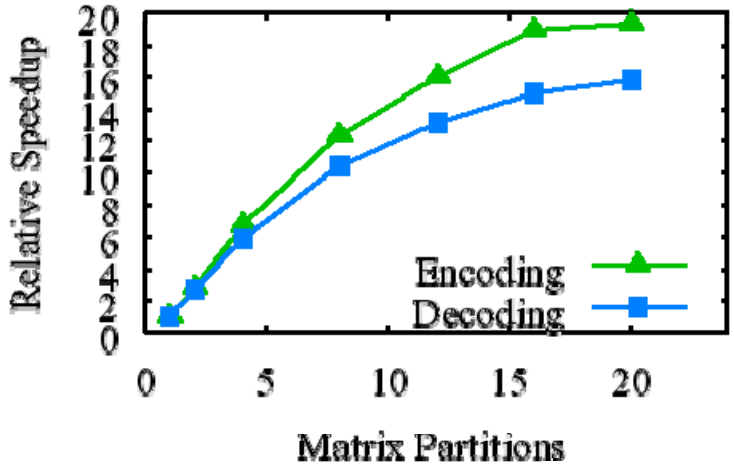
The image is not viewable!

Numerical Results

- W&C Encoding Performance

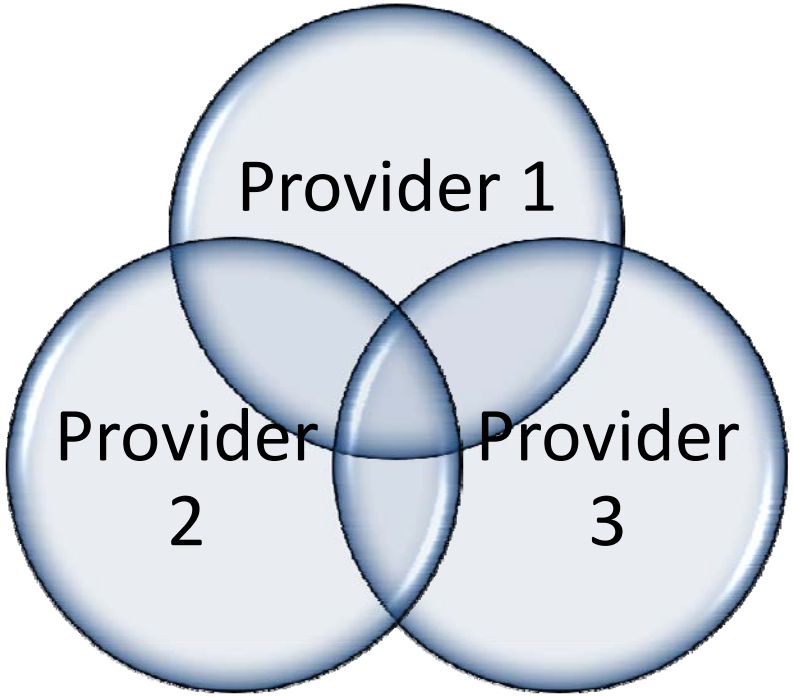


- Matrix Block Chaining Speedup



What's next in W&C?

- Integration of efficient integrity mechanisms
- Extend storage to multiple providers
  - Included redundant information for fault-tolerance
- Examine multiple providers and multiple participants to provide heightened privacy for users



Overlap of data stored at three service provider

Closing Thoughts

- Current threats and challenges to security professionals
  - Can't ignore the complexity
- Research areas
  - Enhanced deployment strategies
  - Cloud full-stack analysis
  - Improved cloud storage

May All Your Clouds Have Silver Linings

Vince Urias
veuria@sandia.gov

David Zage
djzage@sandia.gov

Questions?

Interesting Links

- Fate of data held by Megaupload up in the air
- Megaupload host wants out
  - Carpathia says it is paying $9000/day to keep data
- VMware causes second outage while recovering from first
  - Cloud Foundry (PaaS)
- Windows Azure Leap-Year Glitch Takes Down G-Cloud
- Amazon says it's getting a handle on EC2 outage
- The Hidden Risk of a Meltdown in the Cloud
- Is there Too Much Complexity in the Cloud?
  - 'Black Swans' Are Sure to Fly in the Public Cloud