On the Feasibility of *Data Loss Insurance* for Personal Cloud Storage

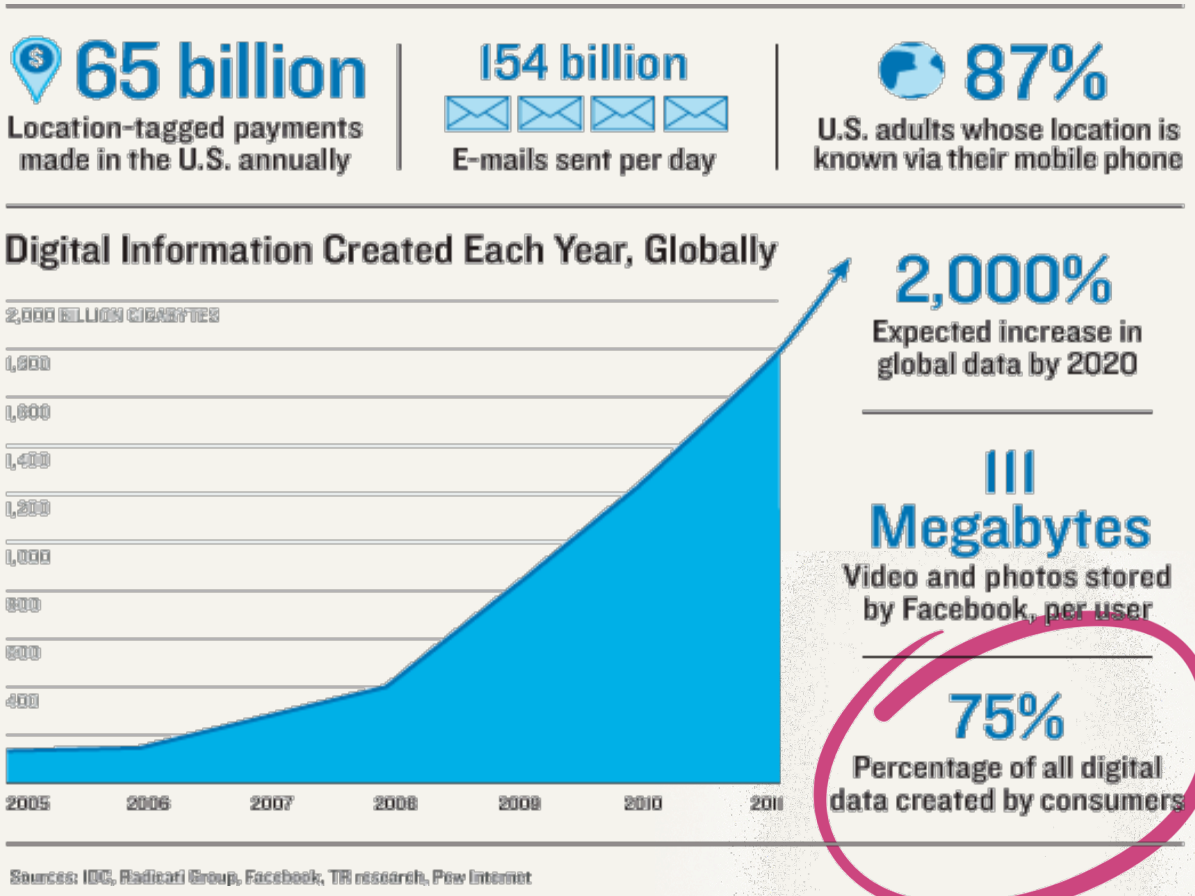
Xiaosong Ma

Qatar Computing Research Institute
North Carolina State University





Personal Data Growth



Source: http://www.technologyreview.com/news/514351/has-big-data-made-anonymity-impossible/

- More data accessed on the go [1]
  - Mobile traffic grew 81% 2012-2013
  - Projected to grow 11-fold by 2018

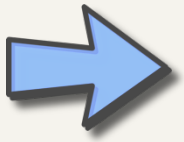
[1] Cisco VNI Global Mobile Data Traffic Forecast, 2013-2018

Traditional Personal Data Storage



From Physical Devices to Cloud

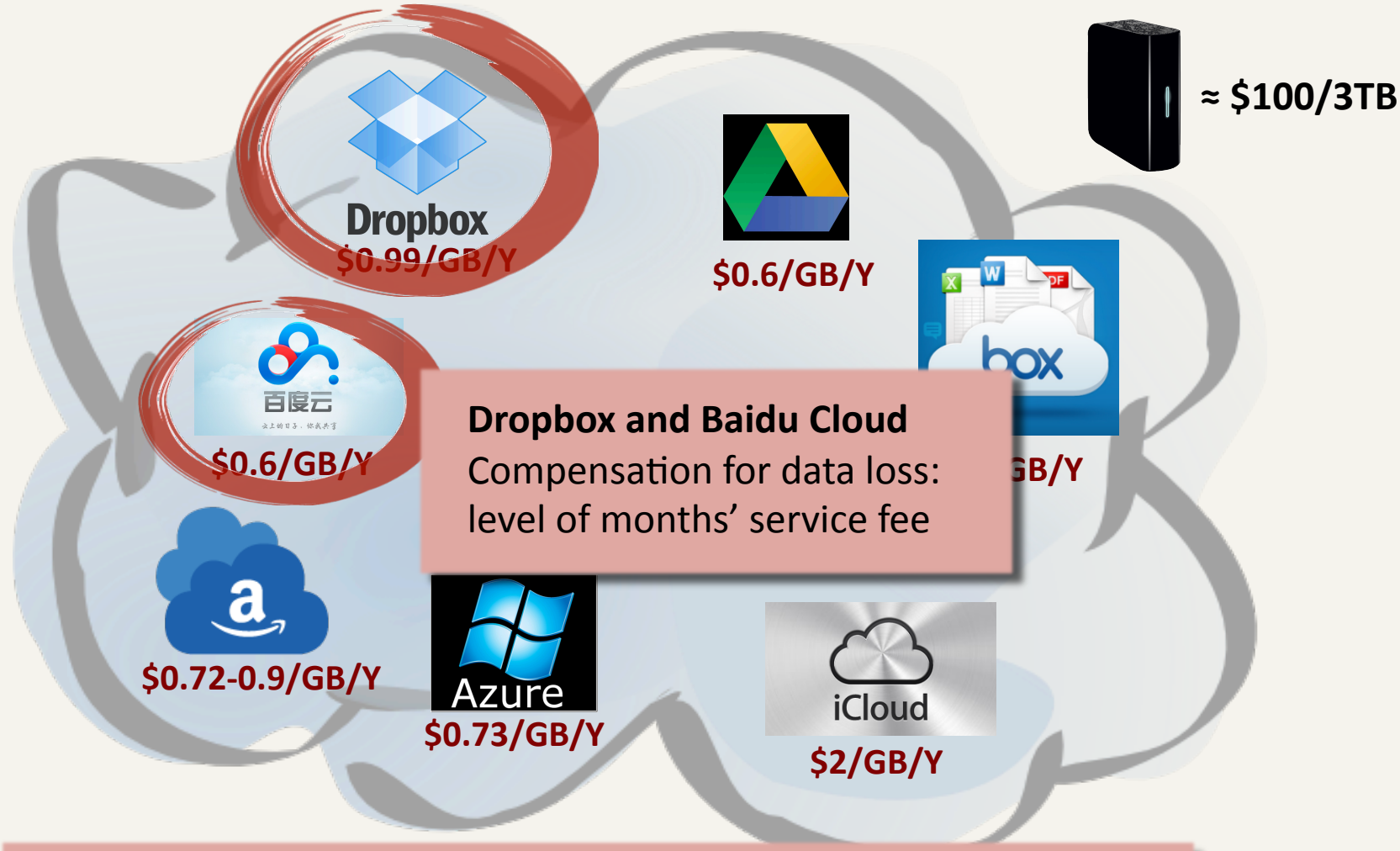






- Widely used
  - Accessible everywhere
  - Trouble-free storage
- Not yet fully adopted
  - Cost
  - No reliability guarantee

Current Services: Cost and Liability



"As-is" or "best-effort" services

- Provider makes reasonable effort to avoid data loss
- But has no liability "to the fullest extent permitted by law"

Conservative Both Ways

- Case study: Amazon S3 storage
  - Explicit availability guarantee
    - Standard: “11 nines”, $0.9/GB/Y
      - 0.000000001% annual object loss rate
    - RR (Reduced Redundancy): “4 nines”, $0.72/GB/Y



- Lots of copies exist
  - Lose data, lose customer
  - Cost included in service
- Still no promise against loss

Traditional Property Protection



Source: http://danapoint-ca-3694.theupsstorelocal.com

Insurance for data stored in cloud?



Common Hazards for Cloud Storage

- Typically include
  - Hazards causing data loss/corruption
    - Storage hardware failures
    - Security attacks
    - Incorrect handling: human errors, software bugs
    - Environmental damage: building corruption/fire
    - Natural disasters: earthquake, flood, storm, wild fire …
  - Fraudulent claims from customers
    - Moral hazard
  - Termination of business

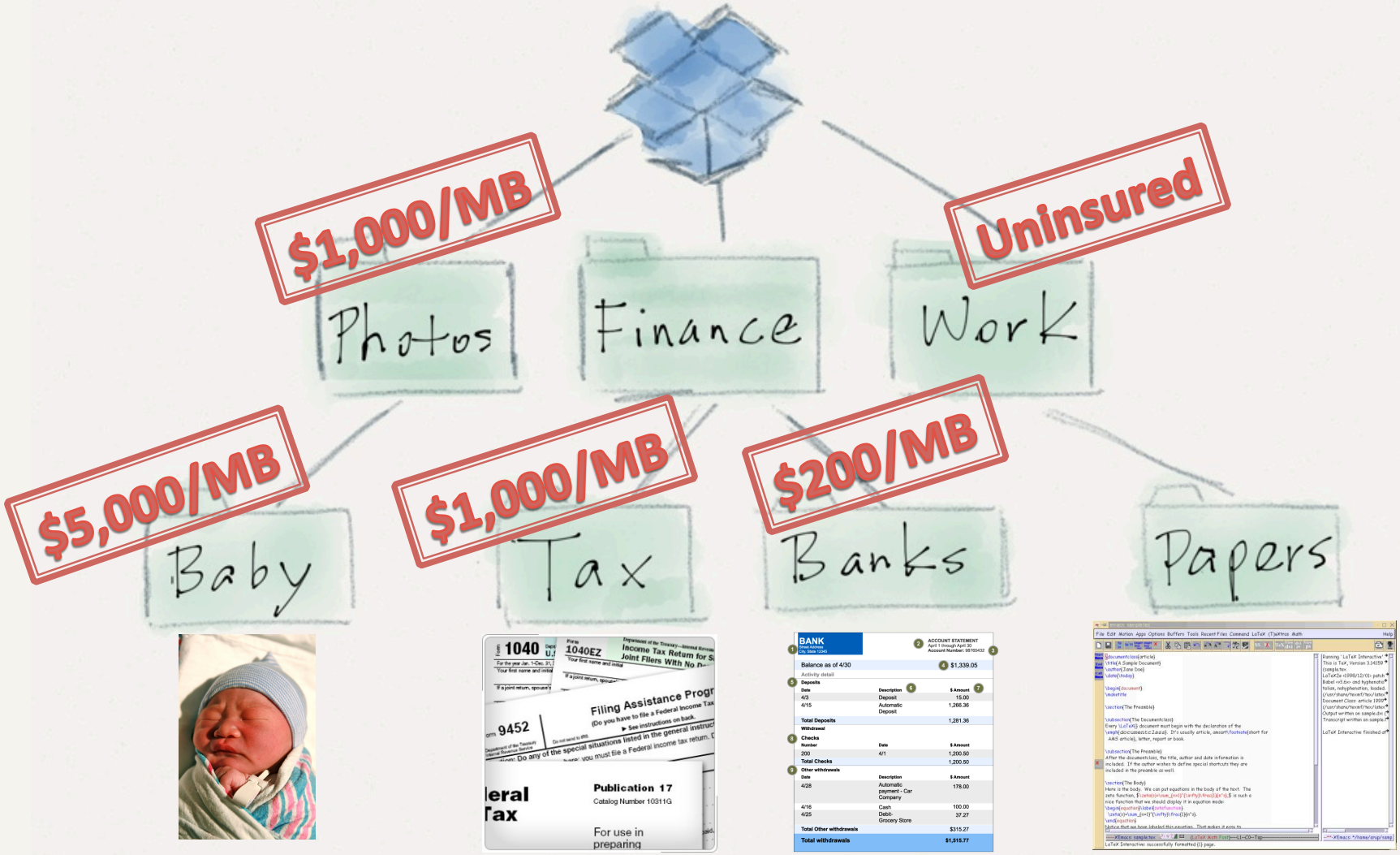


Data Loss Insurance: Rationale

- Unique advantages of data insurance
  - Service providers can actively **reduce** risks
    - By making more copies
    - Any copy of data as good as original
    - Cost of extra copy much lower than value
  - May reduce baseline replication aggressiveness
    - From “11 nines” to “4 nines” or even “2 nines”

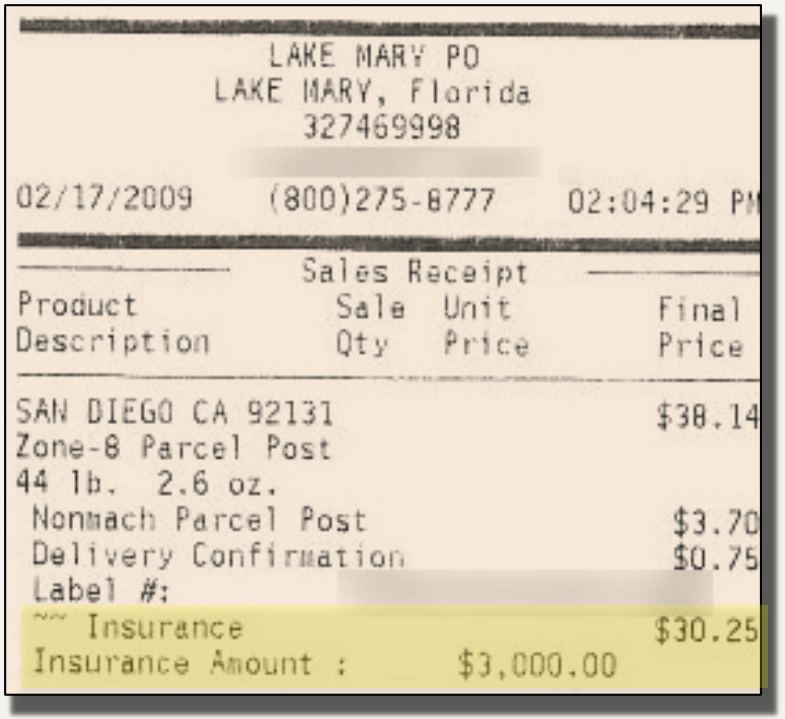
- Optional to user
  - Baseline service charge reduced for all
    - Attract more paying customers

Proposal: Optional Data Insurance



Proof of Loss

- Data receipt containing
  - Traditional metadata
    - Date
    - Dataset size
  - Hash value
    - **Pre-image attack** impractical
    - Challenge compounded by using two hash algorithms



LAKE MARY PO
LAKE MARY, Florida
327469998

02/17/2009 (800)275-8777 02:04:29 PM

Sales Receipt

Product Description	Sale Qty	Unit Price	Final Price
SAN DIEGO CA 92131			$38.14
Zone-8 Parcel Post			
44 lb. 2.6 oz.			
Nonmach Parcel Post			$3.70
Delivery Confirmation			$0.75
Label #:			
~~ Insurance			$30.25
Insurance Amount :		$3,000.00	

```
02/17/2009
-----------------------

Data size: 64MB
Insurance amount: $3,000.00
Hash value list 1:
Hash value list 2:
```

Premium Estimate

- Traditional formula for premium calucation
  - $P = L + U + E$
    - P: premium
    - L: expected loss
    - U: underwriting expense
    - E: profit

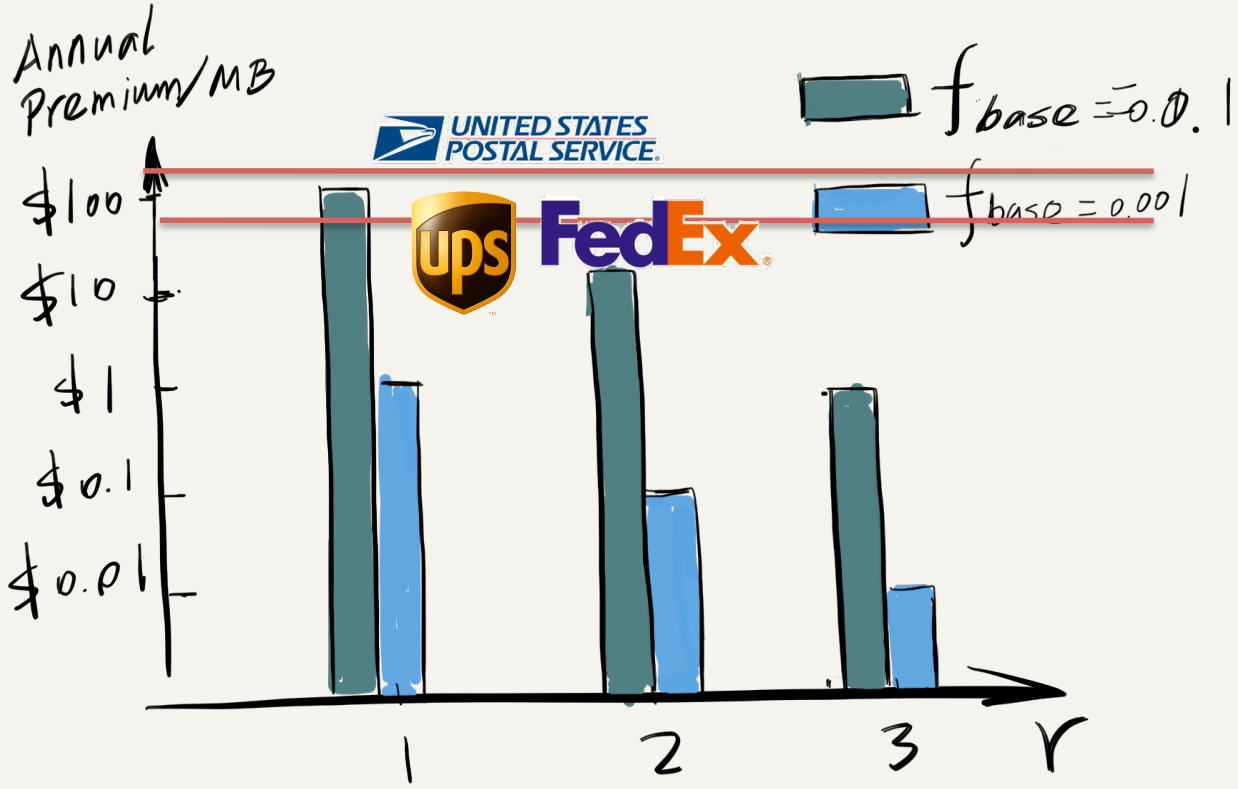
Indemnification

Cost of extra copies

- Annual data loss insurance premium
  - $p = f_{base} f^r v + c_{disk} r$
    - f_{base} : baseline availability (annual object loss rate)
    - f : annual disk failure ratio, aggressively estimated as 0.1 [2]
    - r : replication degree
    - v : declared data value ($/MB)
    - c_{disk} : disk price ($/MB), set to $100/2TB

[1] E. Pinheiro et al., "Failure Trends in a Large Disk Drive Population", *FAST 2007*

Sample Premiums: v = $10,000/MB



- Observations on p-v (premium/insured-value) ratio
  - Highest number matching shipping industry
    - With baseline availability of 0.1 and one extra copy
  - Declines when growing r or f_{base}
  - Indemnification dominates
    - With claimed value significantly larger than storage cost
    - Till replication degree and/or baseline availability large enough

Miscellaneous Issues

- Discontinued storage service
  - Long-term trust required
  - Possible solution: third-party insurers (reinsurance)
- Access pattern aware optimizations
  - Precious objects tend to be read-only
    - May help lower risk and cost
- Going beyond personal data
  - Risks and liabilities more complicated for business data
  - Techniques in business risk management applicable

Closing Remarks

- ***Data insurance might make cloud storage (appear) safer and cheaper***
  - Psychological security for customers
  - Reduced baseline redundancy for providers
- Users explicitly marking important data
  - Implicitly, **unimportant data**
- Requires cross-discipline investigation
  - Computer systems
  - Cnsumer behavior study
  - Actuarial science
  - Law

Thank You!

- Questions?