Information Protection in Content-centric Networks

Christopher C. Lamb

Department of Electrical and Computer Engineering
University of New Mexico

November 6, 2012

THE UNIVERSITY of NEW MEXICO

Outline

Original Goals

Contribution of Work

The contribution of this work is a quantitative analysis of policy-centric overlay network options, associated taxonomies of use, and prototypical technology proofs-of-concept.

- *Network Control Options* — This includes various types networks and associated strengths and weaknesses addressing centralized and decentralized models.
- *Taxonomies of Use* — Depending on the specific usage management requirements and context, different overlays have different applicability; this work will provide guidance on suitability; it will eventually lead to how to manage data flow within SDN-capable infrastructure.
- *Prototypical Technologies* — Examples and proofs-of-concept will be required to appropriately analyze various architectural alternatives.

Meeting the Goals

Network Control Options

I have developed and analysed multiple types of overlay systems, both centralized (hierarchical) and non-centralized (non-hierarchical), with differing topologies and integrated content-centric control.

Taxonomies of Use

I have established an verified a taxonomy of usage management and applied that within the network providing mechanisms extendable to SDN use.

Prototypical Technologies

Prototype information-centric networks are running between the Rackspace and Amazon clouds.

Impact and Originality

- Information-centric architectures common in future internet designs
- Significant work with respect to name/object binding, overall topologies, approaches
- No significant work yet on exploiting information-centricity for enhanced security
- They have significant new capabilities inherent in approach that allow for better information security

Additional Contributions

This work, as well as providing alternatives analysis with respect to security in information-centric architectures and approaches, also demonstrates the first implementation of granular context-sensitive security functionality embedded in an information-centric network.

Publications

Conference Papers:

C.C. Lamb and G.L. Heileman. *Overlay architectures enabling cloud computing for multi-level security environments*. In Services (SERVICES), 2012 IEEE Eighth World Congress on, pages 116-124, june 2012.

Christopher Charles Lamb, Pramod A. Jamkhedkar, Mathew P. Bohnsack, Viswanath Nandina, and Gregory L. Heileman. *A domain specific language for usage management*. In Proceedings of the 11th annual ACM workshop on Digital rights management, DRM '11, pages 51-62, New York, NY, USA, 2011. ACM.

Christopher C. Lamb, Pramod A. Jamkhedkar, Gregory L. Heileman, and Chaouki T. Abdallah. *Managed control of composite cloud systems*. In System of Systems Engineering (SoSE), 2011 6th International Conference on, pages 167-172, june 2011.

P.A. Jamkhedkar, C.C. Lamb, and G.L. Heileman. *Usage management in cloud computing*. In Cloud Computing (CLOUD), 2011 IEEE International Conference on, pages 525-532, july 2011.

Pramod A. Jamkhedkar, Gregory L. Heileman, and Chris C. Lamb. *An interoperable usage management framework*. In Proceedings of the tenth annual ACM workshop on Digital rights management, DRM '10, pages 73-88, New York, NY, USA, 2010. ACM.

Publications

Journal Articles and Book Chapters:

Christopher C. Lamb and Gregory L. Heileman, "Content-centric Information Protection in Cloud Computing", *International Journal of Cloud Computing and Services Science* vol. 1, no. 5, December 2012.

Pramod A. Jamkhedkar, Christopher C. Lamb, and Gregory L. Heileman, *Digital Rights Management: Technology, Standards and Applications*, Auerbach Publications, 2013.

Christopher C. Lamb and Gregory L. Heileman, "Dynamic Context-sensitive Information Protection", *IEEE Internet Computing - Dynamic Collective Work*. *(pending)*

Motivation for Work

Unified Cross-Domain Management Office (UCDMO)

The UCDMO is responsible for cross-domain classified information flow. They are focused on protecting information while providing it to whomever needs it.

Defence Advanced Research Projects Agency (DARPA)

DARPA has been trolling for solutions in this space under the Clean-slate design of Resilient, Adaptive, and Secure Hosts (CRASH) program.

Medical

Medical systems handle highly sensitive information, and the distribution of that information must be controlled and traced.

This work examines the impacts of specific **strategies** on the **security properties** of information-centric networks of different **topologies**.

Strategies

What are these strategies?

Strategies are methods designers can use to protect information. More than one can be used at a time.

- **Redaction** is removing sensitive content from the body of a message, but leaving non-sensitive content.
- **Rerouting** is removing sensitive content and retransmitting that content over a more secure channel or route.
- **Encryption** involves enciphering sensitive content in a message while leaving non-sensitive content in the clear.

Security Properties

What are these security properties?

In this work, the primary properties of concern are confidentiality, integrity, and availability. These are the system properties most traditionally associated with cyber-security evaluation.

- **Confidentiality** is a measure of the overall confidentiality of sensitive information. Information that can be accessed by unauthorized users has low confidentiality.
- **Integrity** describes whether the information has been altered in any way. If information has changed, its integrity decreases.
- **Availability** is related to the ability of a user to acquire information to which he or she is authorized.

Topologies

What do we mean by hierarchical, non-hierarchical?

These are two different network types that represent two extremes of network organization.

- **Hierarchical** networks are ordered in a strict top-down structure. Here, they are organized in a tree in which only the leaves of the tree content content, and all internal nodes only route requests and reponses.
- **Non-Hierarchical** networks are more loosely organized. In non-hierarchical networks, all nodes route information and contain content. They are akin to peer-to-peer networks.

Motivation for Selection

Why Confidentiality, Integrity, and Availability?

These are system properties wide associated with cyber-security. They provide insight into the ability of a system to protect and deliver information.

Why Redaction, Rerouting, and Encryption?

These are three distinct possible strategies that system designers can use to protect sensitive information. Encryption is commonly used to protect sensitive information. Redaction is commonly used with documents, and rerouting has been commonly used for sending classified data to remote locations.

Why Hierarchical and Non-Hierarchical networks?

These are two distinct types of network approaches to storing information commonly used in today's computer systems.

Results Overview (1 of 2)

Confidentiality, Integrity characteristics based on approach.

- **Redaction**, by removing information, by definition destroys integrity while guaranteeing confidentiality; unavailable information that is cannot be leaked
- **Rerouting** removes information from a context damaging integrity that can possibly be repaired later, potentially increasing confidentiality by rendering that information unavailable
- **Encryption** minimizes integrity impacts be keeping ciphered data with original context at the expense of possible interception and cryptanalysis exposure

Availability is based on performance.

- **Performance** is measured via end-to-end time of transmittal

Results Overview (2 of 2)

Overall evaluation of impact against strategy:

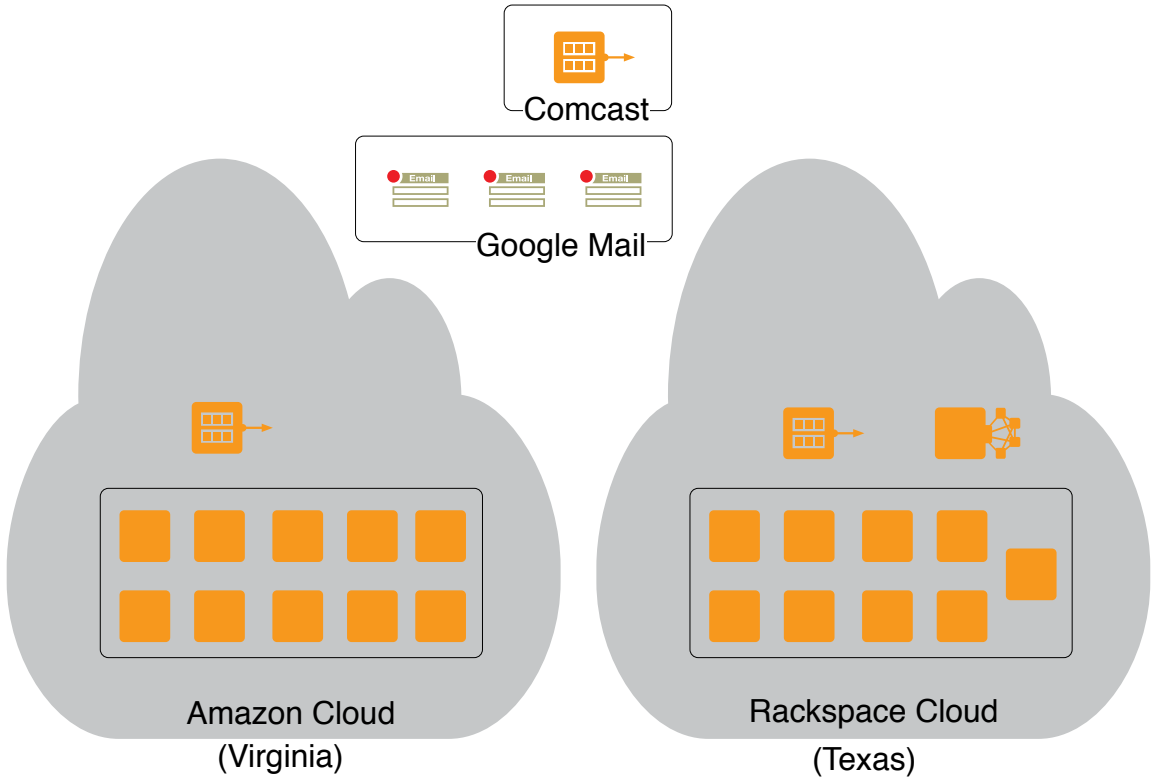
- Encryption most likely to be used...
- ...Rerouting likely the best compromise (but expensive)
- Hierarchical and non-hierarchical networks had similar performance
- No clear leading strategy under all conditions

Property	Redaction	Rerouting	Encryption
Confidentiality	3	2	1
Integrity	0	1	3
Availability	3	1	2

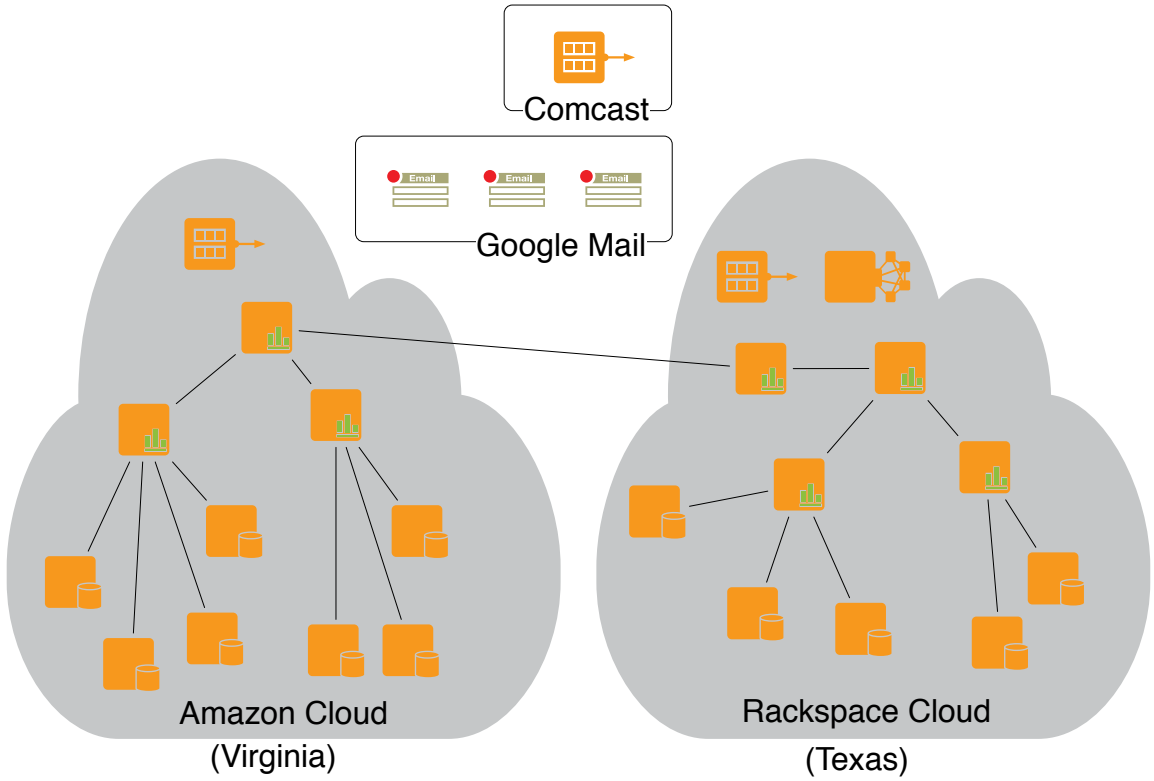
Strategy Impact by Attribute (3 is best, 0 is worst)

What does this mean? How did we get it?

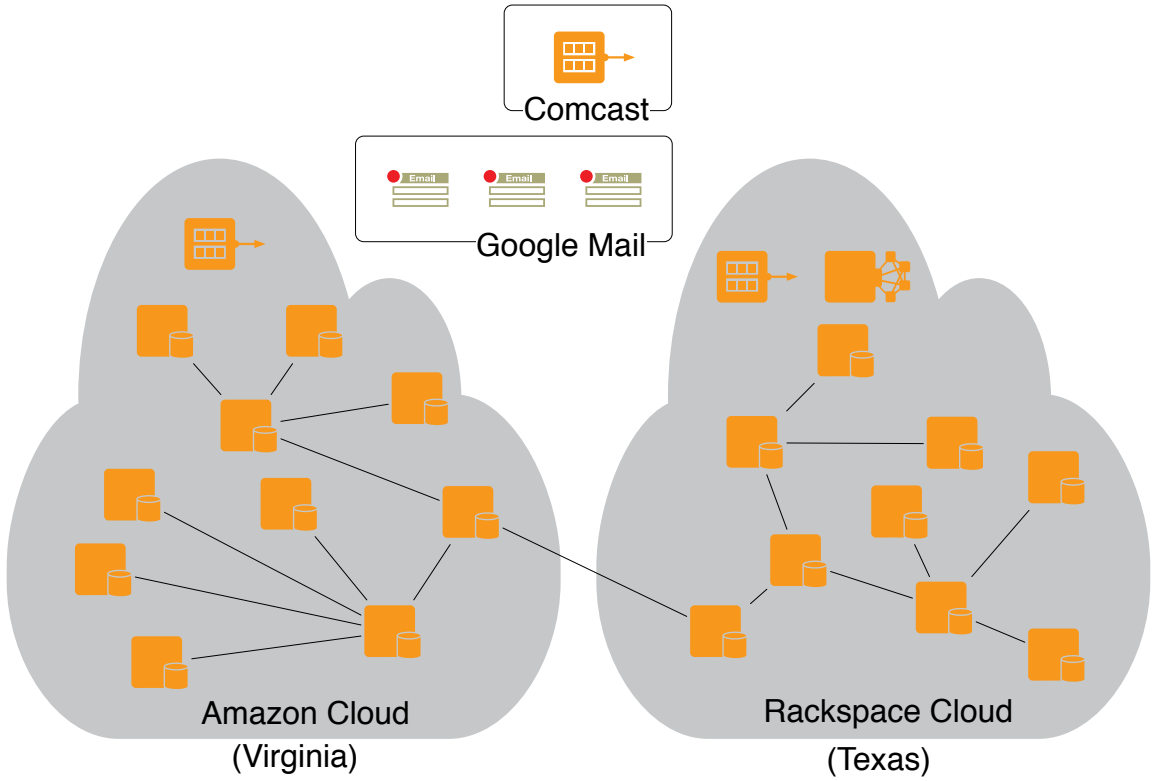
Physical Topology



Hierarchical Topology



Non-Hierarchical Topology



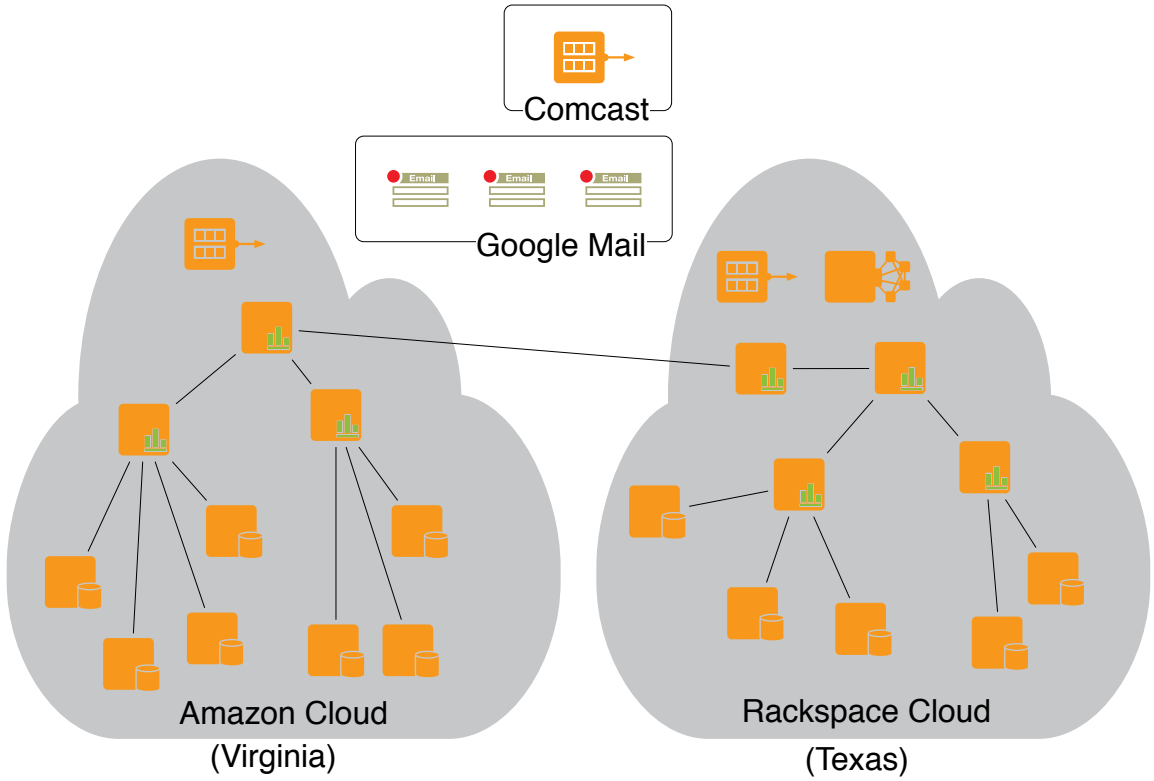
Hierarchical Experimentation

The nodes fielding content requests are scattered throughout the network, though the content is fixed.

- The **home** node contains the requested content and is hosed at Amazon.
- The **peer** node is under the same immediate router as the home node, and is also hosted at Amazon.
- The **neighbor** node is under the neighboring router, again at Amazon.
- The **distant 1** and **2** nodes are under routers in Rackspace infrastructure on the other end of the network.

Control testing does not apply usage management to content. All other cases apply usage management and the specific outlined strategy.

Hierarchical Topology Revisited



Hierarchical Effects

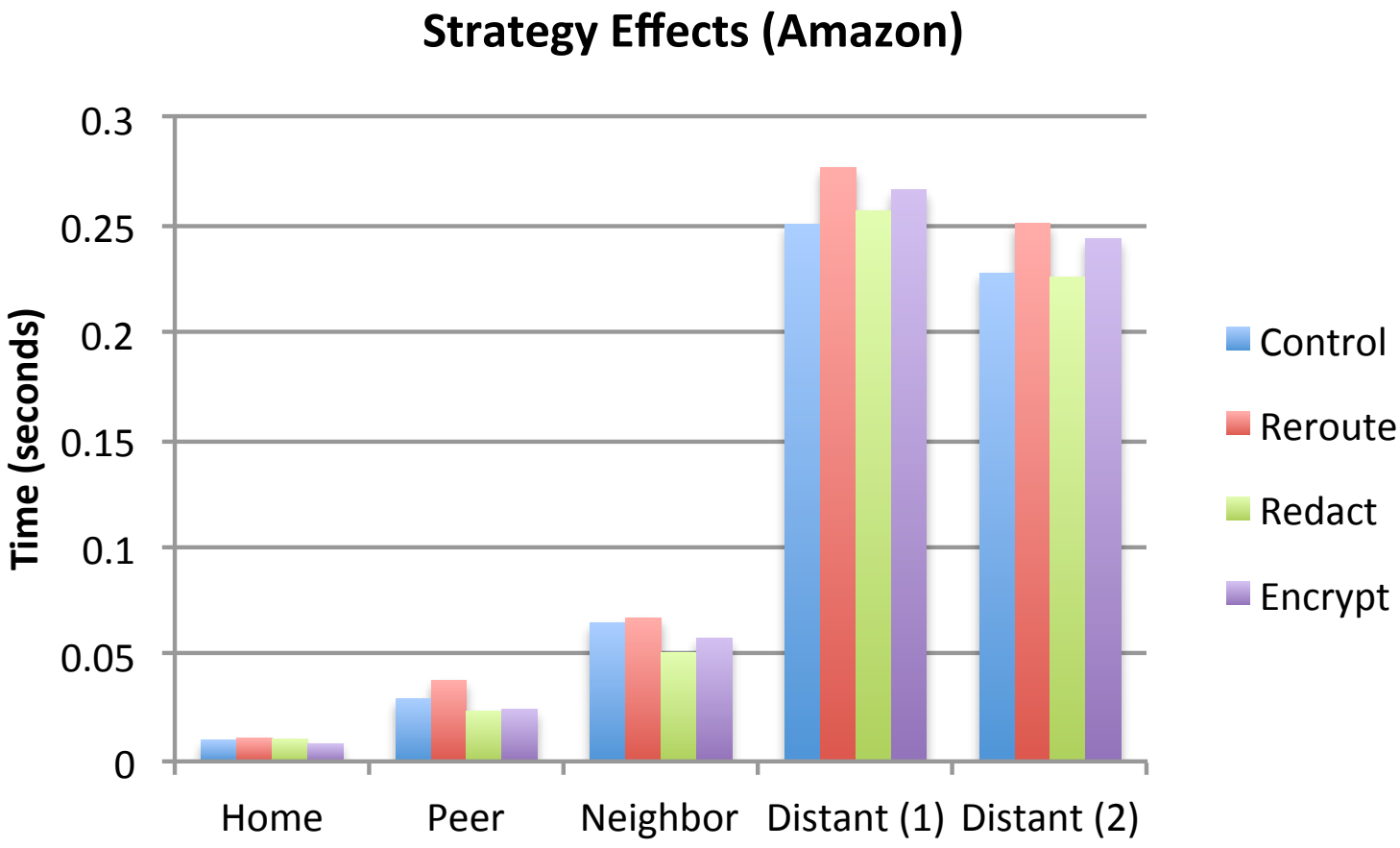


Figure: Hierarchical Results from Amazon

Hierarchical Effects

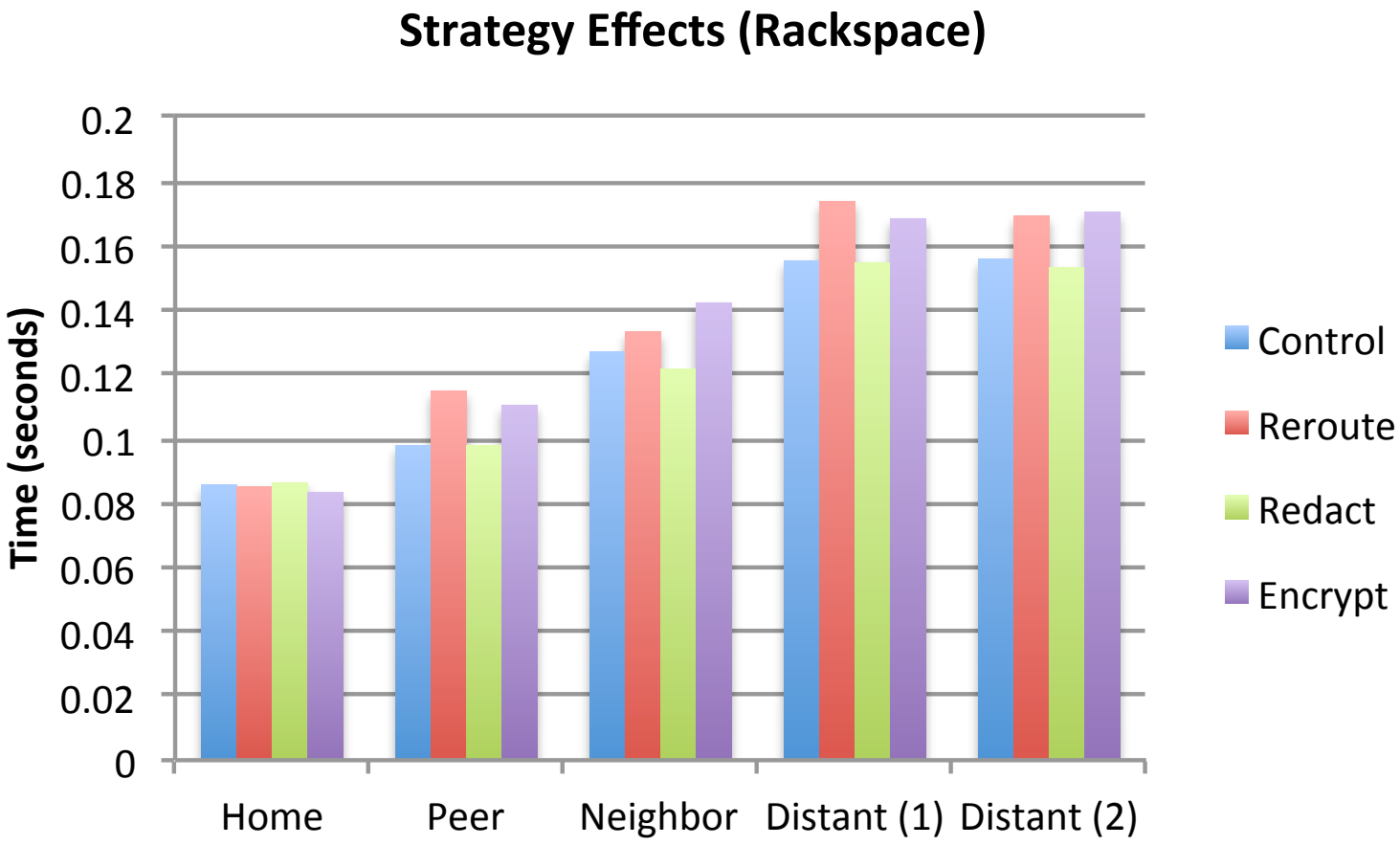


Figure: Hierarchical Results from Rackspace

Hierarchical Effects

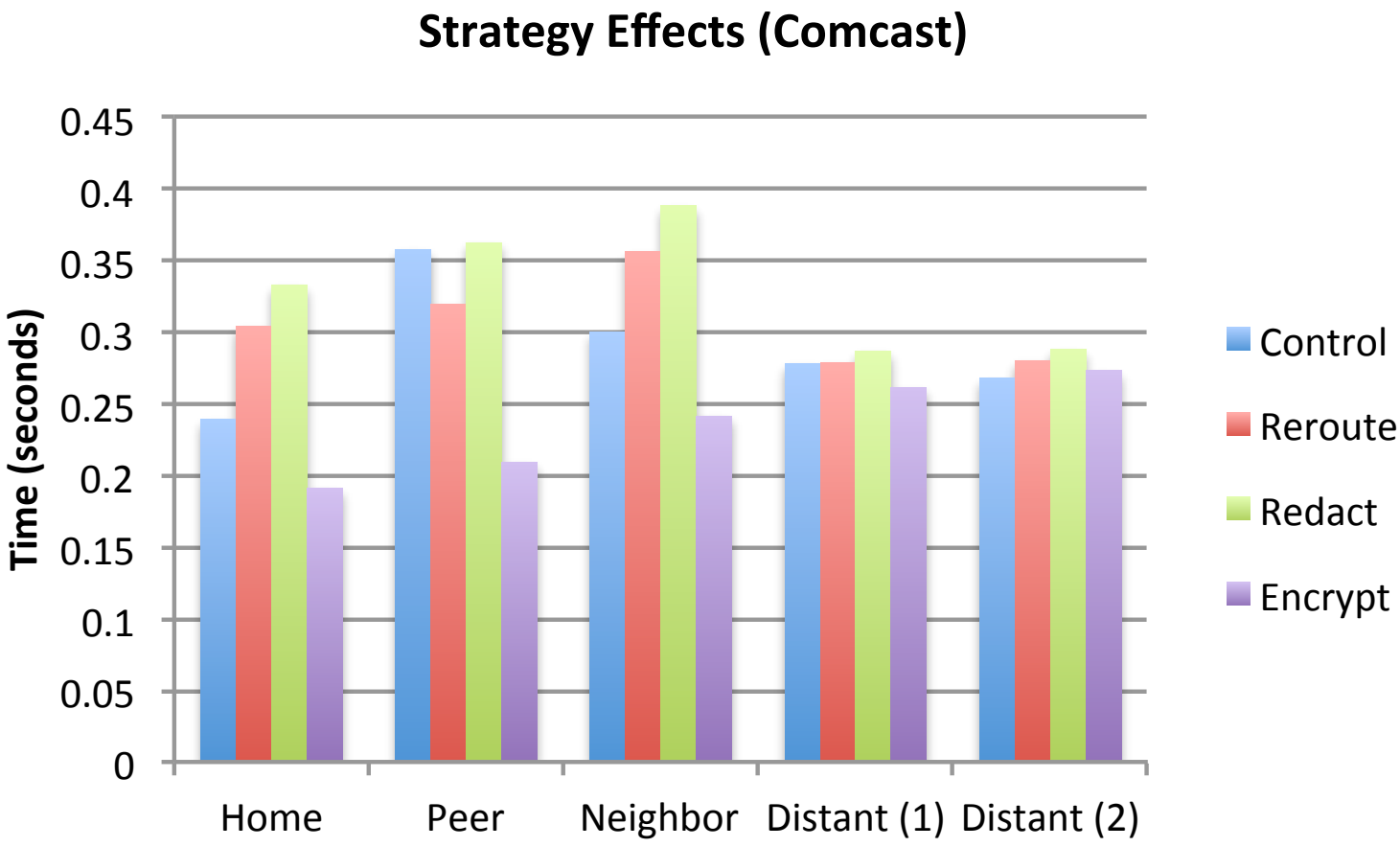


Figure: Hierarchical Results from Comcast

Hierarchical Analysis

Caching is important.

Network effects have a tremendous impact on performance. Using in-node content caching helps take advantage of query locality.

Less naive routing wouldn't hurt either.

A query from the amazon test node to rackspace nodes travels to rackspace, to amazon, to rackspace, and then back to amazon. Better node location services or caching could eliminate this back-and-forth.

Infrastructure costs are high.

Rerouting content has some significant conceptual advantages, but establishing reliable secondary infrastructure can be expensive and difficult.

Encryption is a realistic compromise.

Encryption seemed performant and didn't sacrifice integrity and availability for confidentiality gains.

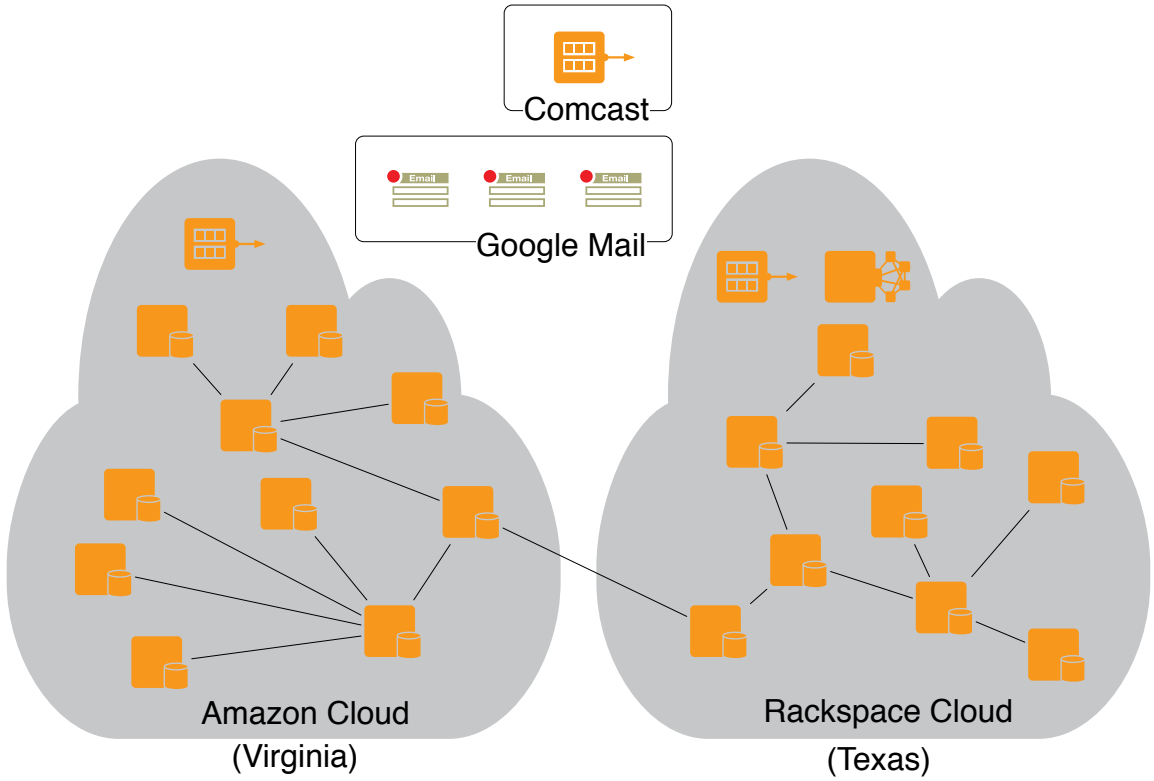
Non-Hierarchical Experimentation

Nodes are in both Rackspace and Amazon. The home node and neighbor nodes (1) and (2) are in amazon, while the remaining nodes are in Rackspace. Content is again fixed.

- The **home** node contains the requested content and is hosed at Amazon.
- The **neighbor (1)** node is next to the home node.
- The **neighbor (2)** is next to the neighbor (1) node.
- The **neighbor (3)** is next to the neighbor (2) node.
- The **neighbor (4)** and **(5)** nodes adhere to the same pattern.

Control and strategy application is the same as in hierarchical testing.

Non-Hierarchical Topology Revisited



Non-Hierarchical Effects

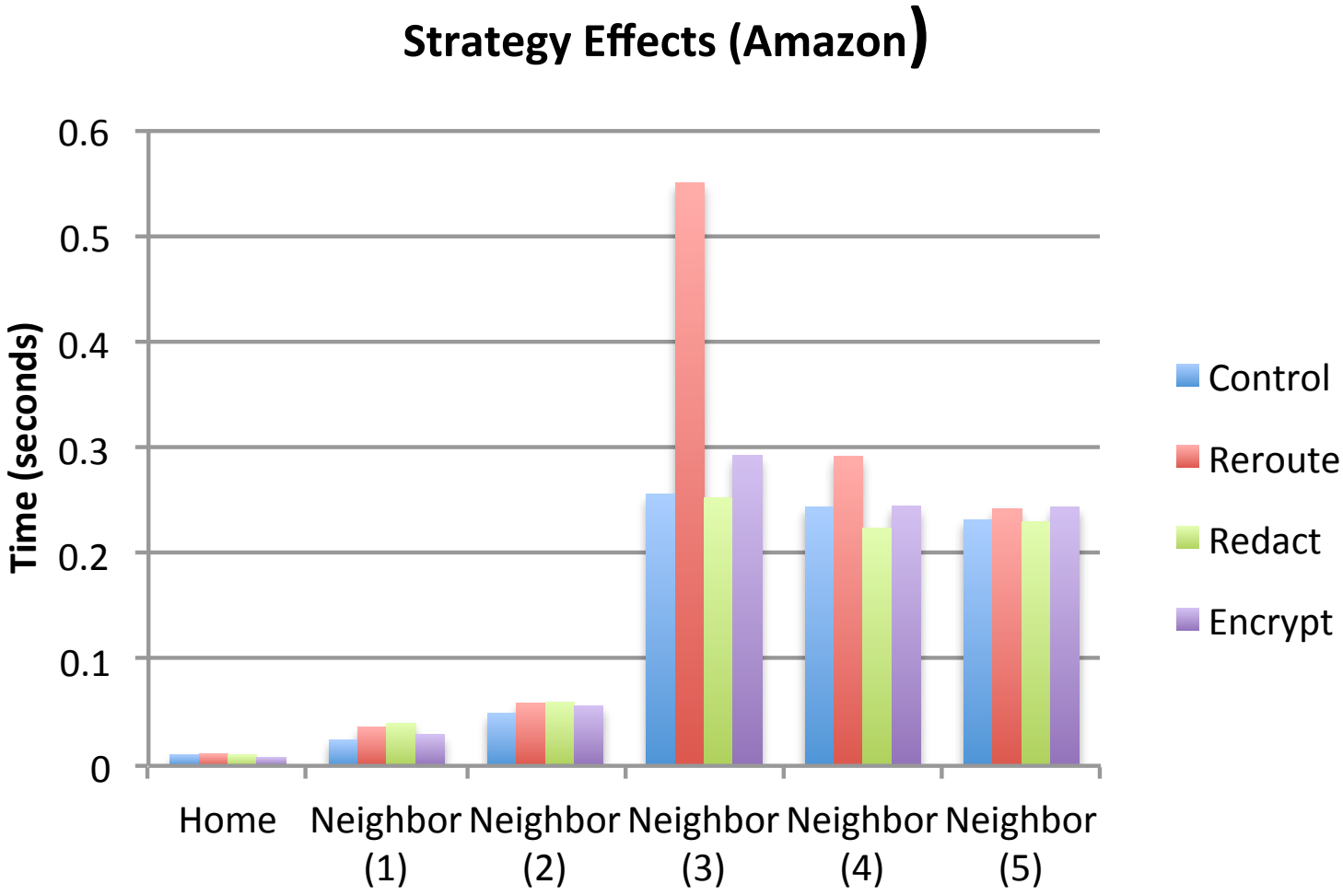


Figure: Non-Hierarchical Results from Amazon

Non-Hierarchical Effects

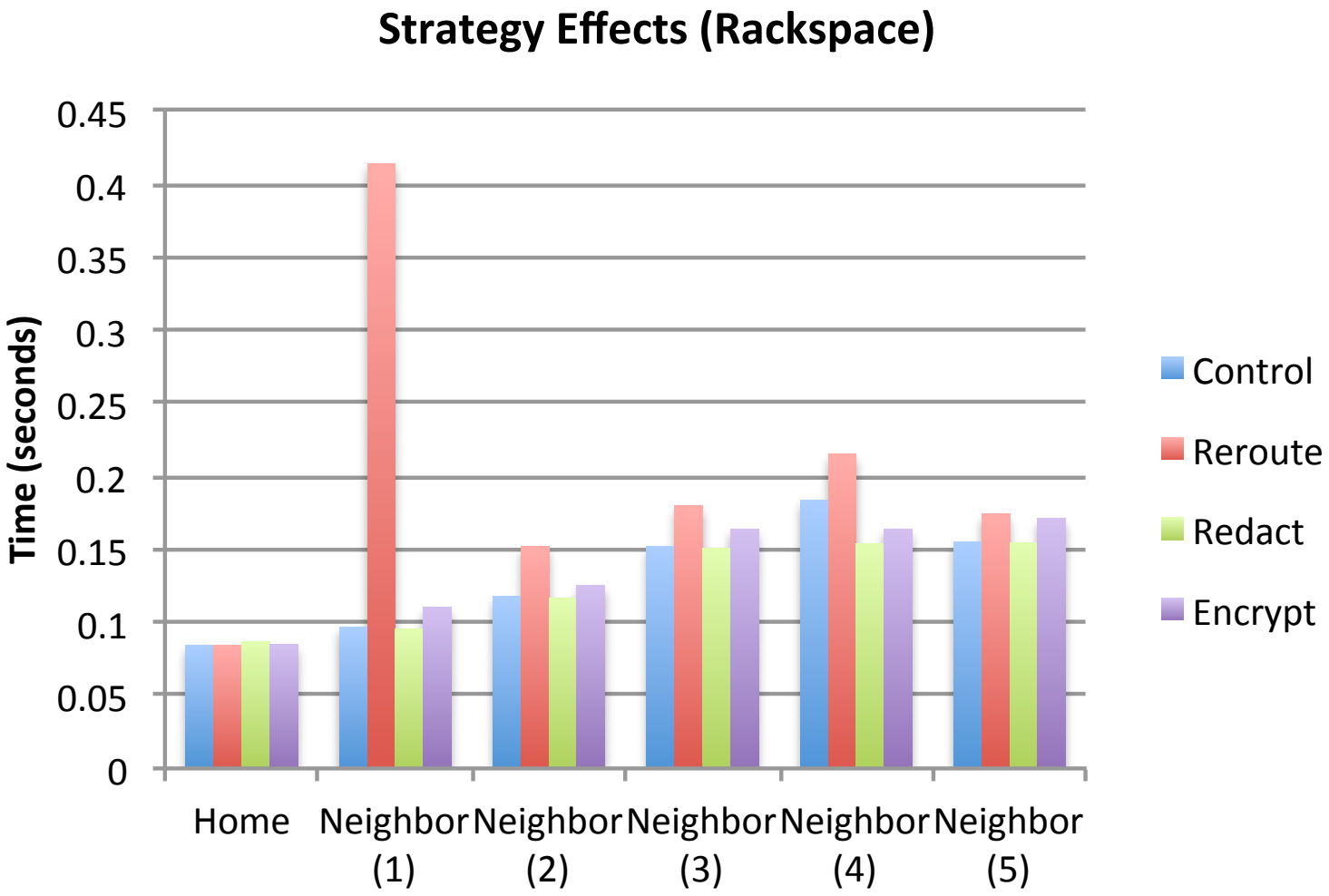


Figure: Non-Hierarchical Results from Rackspace

Non-Hierarchical Effects

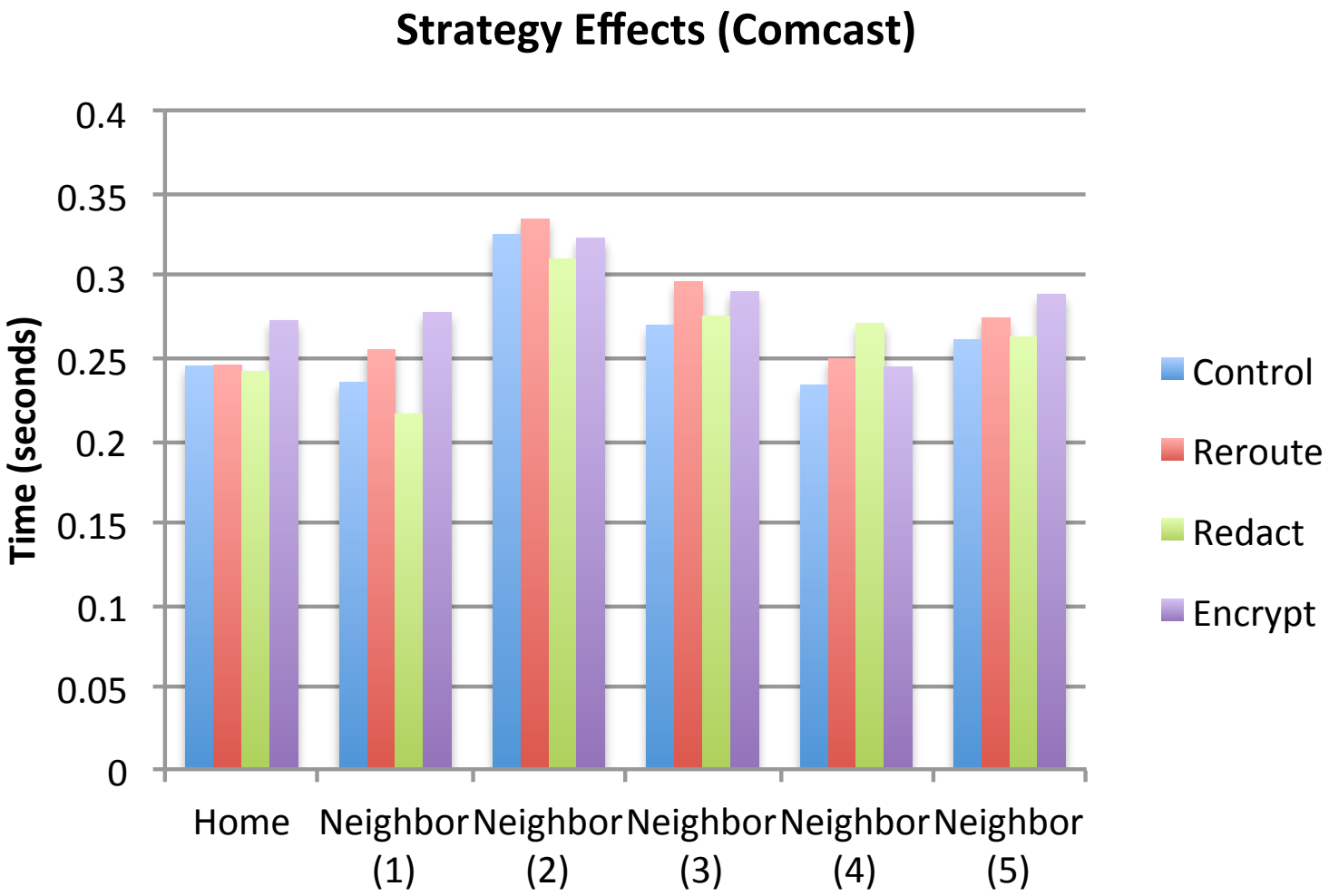


Figure: Non-Hierarchical Results from Comcast

Non-Hierarchical Analysis

Similar performance to hierarchical topologies.

We see similar performance profiles for hierarchical and non-hierarchical networks. Conclusions from hierarchical networks with respect to routing and caching hold here as well. Non-hierarchical networks do have additional content processing demands.

Secondary infrastructure effects more prominent.

In these tests, we had significant infrastructural issues with respect to rerouting with our provider. Again, secondary infrastructure can be problematic.

Non-hierarchical and hierarchical have different strengths.

Hierarchical networks may more effectively resist random failures due to functional centralization. Non-hierarchical networks may better maintain functionality in the face of directed attacks. This will likely depend on topology as well as hierarchical characteristics. *More research is needed to support these conclusions.*

Network-Free Experimentation

Why network free?

This highlights the basic characteristics of a given strategy without any network effects clouding results. End-to-end results are important, but so are basic strategy effects.

How implemented?

- Single node in Amazon
- Requests and responses over loopback
- Strategies and control tests are the same
- Used the home node containing content of interest

Network-Free Evaluation

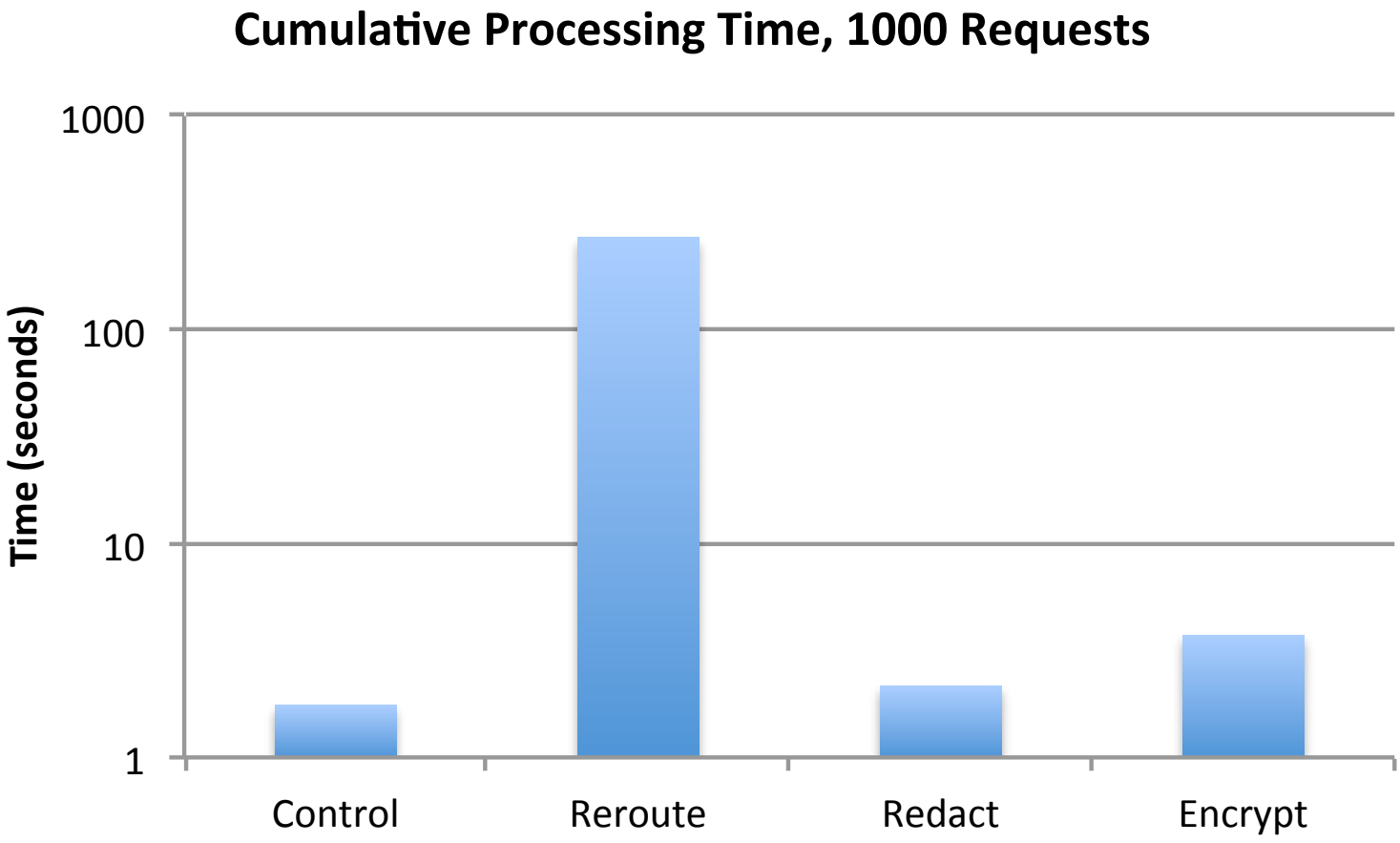


Figure: Results from Requests to a Single Node

Network-Free Analysis

Doing nothing is cheapest.

As expected, the control group has the best performance. We are not processing content in the control example, rather we just pass content directly through the node.

Redaction is next.

Computationally, redaction is the simplest strategy, and is just slightly more expensive than the control set, but it destroys integrity.

Encryption is just slightly more expensive than redaction.

We do use symmetric ciphers here, so this is very fast. Public key encryption would be more expensive.

Rerouting is by far the most expensive.

Note that this is a logarithmic scale graph! Setting up secondary channel communication is expensive and time consuming.

Conclusions

What did I learn?

- Redaction destroys integrity
- Rerouting can be expensive
- Encryption can be broken

Surprisingly, rerouting is not the most realistic option!

- Rerouting may provide the best compromise between confidentiality and integrity, but at a potentially high availability, and likely financial, cost. Maintaining dual infrastructures, especially if the second is highly secure, will more than double infrastructure spend. Results will vary based on implementation more than just strategy.
- Encryption may be eventually broken, but is difficult to decipher short-tem, is easier to implement, and is likely much less expensive than rerouting.

Encryption is the most *realistic* solution.

What's next?

This work is funded for an additional six months, and likely for an additional three years of continuing funding. Goals currently include:

- **Productionization** of the current system for eventual deployment in operational systems. This will involve extending the current infrastructure to handle additional types of data and potentially different and lower-level protocols.
- **Public Key Infrastructure** to provide non-repudiation as well as confidentiality to protected information. This will likely involve some combination of symmetric and asymmetric cryptography (e.g. TLS, IPsec).
- **Software Defined Networking** to enable real-time network configuration in response to changing contexts.
- **Publish/Subscribe** approaches to registering for content of interest, and associated decoupling of request and response network paths.

Questions? Comments?