Hearing before the New York City Council
Committee on Technology

Cloud Computing and Storage

Steven M. Bellovin*
Department of Computer Science
Columbia University
https://www.cs.columbia.edu/~smb

December 15, 2020

*Affiliation listed for identification purposes only.

Introduction

Thank you for inviting me to speak here today on the proposed cloud storage bill.[1] As I noted the other time I was invited to speak before a Council committtee, I'm a native New Yorker. I grew up in Brooklyn and attended the city's public schools, my first paid summer job was in the Municipal Building, and I like to spend my free time bicycling or photographing birds in city parks. In other words, I really live here, and I'm delighted to have a chance to give back to the city.

By way of introduction, I'm a professor of computer science at Columbia University Engineering and affiliate faculty at Columbia Law. Security and privacy have been my main focus for more than 30 years; I caught my first hackers in 1971 while working at the City College Computer Center. In 1994, I co-authored the first book on Internet security.[2]

The issue of cloud computing is a complex one, with many aspects; I discuss some of them below. Furthermore, it is difficult to discuss cloud storage without discussing cloud computing; the two are intimately linked. I do note that I know little of the operations of the Department of Information Technology and Telecommunications (DoITT); my experience as a computer programmer for the city's Comptroller's Office was more than 50 years ago! My conclusions, in brief, are:

- New York City may be large enough that, with one exception (the prices of real estate and electricity) it may not need to use cloud providers—but that is true if and only there is stable, predictable funding for information technology (IT) operations. Apart from that, there are likely cost, security, and stability advantages to using commercial cloud providers. Furthermore, there are some cloud services that are very hard for any single entity to replicate on its own.

- Cloud-based storage is maximally effective if coupled with cloud computing and cloud applications. Furthermore, there may be significant advantages in reliability and functionality from using such services, if done correctly.

- If the city does decide to use commercial cloud providers, it will necessitate some change in how DoITT is organized.

- Because of that, external as well as internal advice should be sought on the proposed conversion. It is desirable to do a trial deployment; this will be most effective on a service that would otherwise require a significant in-house upgrade.

All of this is explained in more detail below.

1. A Local Law in relation to an assessment of the feasibility of storing city agencies' electronic data on cloud computing systems, T2020-6906, New York City Council (2020), https://legistar.council.nyc.gov/LegislationDetail.aspx?ID=4700300&GUID=EBB5B86F-C182-4A17-A4B2-FEE9E21043B4&Options=&Search=.

2. William R. Cheswick and Steven M. Bellovin, *Firewalls and Internet Security: Repelling the Wily Hacker*, 1st edition (Reading, MA: Addison-Wesley, 1994), ISBN: 0201633574, http://www.wilyhacker.com/1e/.

What is "The" Cloud?

People speak of "the cloud", "cloud computing", and "cloud storage", but those simple phrases hide a world of complexity. In reality, there is not one "cloud"; rather, there are many different services offered by the various providers. Even something as simple as "storage" can be provided in many different ways. It is not possible to do any single analysis; each cloud service has to be analyzed separately. Furthermore, there are interactions; the desirability of, say, a cloud-resident disk depends heavily on how that disk is used and whether the computation is local or remote.

The proposed bill speaks of "data classification categories"[3] without explaining what those are. It appears, though, that it refers to the contents of the data—payroll records, tenant complaints, taxi trip data, etc.—rather than how the data is organized and how it is used. While this is an appropriate way to reason about, say, security concerns, how the data is used in a technical sense can be more important for assessing the larger costs and benefits. A simple file with a list of names and salaries has different characteristics than a database; both can differ from data that is to be fed into a machine learning model.

There are three overriding issues that affect all analyses, though. Two are obvious: the cost in New York City of both real estate and electricity. Data centers can take a lot of space, and while they need not be located in Midtown, nowhere in the city is particularly cheap. Electricity is expensive everywhere in New York, and data centers use an amazing amount of power, both for the computers and for the air conditioning units they need.

The third issue, though, is little known outside of the computing field: latency, how long it takes for a message to reach its destination. Latency can have a significant effect on the performance and responsiveness of a system; the effective bandwidth of an Internet connection is intimately related to the latency between the two points. Furthermore, the ultimate limit is the speed of light, the absolute speed limit of the universe. The laws of physics dictate that a response to a message from New York City to an Amazon Web Services (AWS) data center in Northern Virginia *cannot* take less than 3.6 milliseconds; it will always take longer. While this seems fast to us, it is a long time to a computer. In many situations, then, it is wise to put computation near the storage. That is, cloud storage can often imply the need for cloud computing.[4]

In fact, there are more categories than those two. Roughly speaking, cloud services can be classified into three different categories: storage, computing, and applications. Each has its own variations and nuances.

3. T2020-6906 at §1(c)(1).

4. Latency can be so critical for applications like high-frequency Wall Street trading that some firms are using novel fiber optic cables to save just a few *billionths* of a second; see, e.g., Alexander Osipovich, "High-Frequency Traders Push Closer to Light Speed With Cutting-Edge Cables," *Wall Street Journal*, December 15, 2020, https://www.wsj.com/articles/high-frequency-traders-push-closer-to-light-speed-with-cutting-edge-cables-11608028200.

Cloud Storage

Basic cloud storage is the simplest possibility. In effect, one has an Internet connection to a set of remote disk drives. Used this way, it is almost certainly not worthwhile for general-purpose use—as noted above, it would be far slower than local disks for complex uses—unless cloud computing is used as well.

The analysis is somewhat different for more sophisticated uses, notably services optimized for sharing files. Dropbox is the best-known example, though there are others. The question here, and it is a question best answered by the DoITT, is to what extent such sharing is necessary or desirable. Some organizations block Dropbox et al. because they do not want their employees working from home on sensitive enterprise files. On the other hand, inter-organizational sharing within the city government is simplified if such services are used. The security of cloud storage solutions is likely to be better than trying to do it in-house with standard mechanisms. Not only are the remote access protocols better designed for hostile networks compared with local network-oriented protocols such as Oracle's NFS and Microsoft's CIFS, managing user authentication is easier: the services do that. Managing access control rules—who is allowed to do what with the data—is still a challenge but is likely easier for cloud storage services.

A third use for cloud storage is off-site backup of crucial files. This is almost certainly an excellent idea; it's faster, easier, and more automatic than manual solutions such as shipping backup media to specialist companies. Preference should be given to services that permit client-side encryption of the data, since that protects the information from any possible security failures by the cloud provider. Of course, tracking, managing, and protecting the encryption keys will require careful and reliable management by the DoITT.

Many of the cloud storage providers offer file versioning—having rapid, automatic access to older versions of files—and even their own long-term archival storage.

There is an important caveat: a sometimes-ignored part of the price structure of many cloud services is the cost of exporting data from them. That is, there can be a significant charge for the network traffic necessary to retrieve bulk data from the providers. Their goal, of course, is to encourage you to do more of your computing on their platforms.

Cloud Computing

Cloud computing has several interesting use cases. The first, of course, is to move computation near the cloud storage disks, to avoid latency-related problems. There is little more to say about that aspect: the cost-effectiveness of the two must be evaluated together.

A second common use case is for surge computing capacity: if an organization's need for CPU cycles varies dramatically over time, it can be considerably more cost-effective to rent it when needed, rather than purchased. The classic example is a movie studio that suddenly needs a vast amount of computing power for digital special effects, but may not need anything near that much until they produce another science fiction movie—but most organizations' needs computing needs are far less bursty. DoITT is

best equipped to ascertain if any city applications require surge capacity. (If there are such, cloud storage, either permanent or temporary, may be needed, too.)

There's an important special case of this: the ability to rapidly create new, high-quality, interactive services. They could start small on virtual machines; more capacity could be added as needed, using the built-in load-balancing features of cloud platforms.

The most interesting issue, though, is whether cloud computers are more or less secure than locally managed ones. The answer is very dependent on the precise scenarios being considered. In particular, since the quality of system administration is vital to security—an issue I raised in Congressional testimony some years ago,[5] my testimony here in February,[6] and which I discussed more fully in a recent book[7]—the relative security of a cloud-resident server compared with a locally-run one turns heavily on how well the systems are administered. That in turn depends on both the nature of the cloud computing service and on city policies.

System administration is easy to scrimp on. Many of its products—keeping things running smoothly and installing patches—are often invisible; indeed, many people wonder about system administration only when something breaks. If budgets are tight, it is all too easy to postpone hardware and software upgrades, and to hire too few system administrators or give them the budgets they need to do tool development.

This risks disaster—you find yourself in a trap where you *must* upgrade, for security reasons, but you're so far behind that the upgrade has become a major project. It's easier to stay current in the private sector, where computers can be depreciated; this aids in long-term planning. In government, at all levels, political necessities can result in budget cuts. If these cuts are to system administration, security can suffer. Obviously, there can be budget cuts in the private sector as well, but the tax benefits[8] make equipment upgrades less painful.

The tradeoffs are different for the large cloud providers. They have learned to do system administration at scale; the marginal human cost of adding another computer or another thousand computers is close to zero. They are thus much better equipped than almost any other organization to stay current on patches, which in turn improves security.

There is, however, a crucial caveat. "Cloud computing" can mean a virtual machine (VM) managed and administered entirely by the customer, in which case it is no better than a local machine administered by the same organization. Alternatively, cloud computing can be access to a system administered by the cloud provider, in which case there can be significant security benefits.

5. Cybersecurity Research Needs: Testimony before the House Select Committee on Homeland Security, Subcommittee on Cybersecurity, Science, Research, & Development, hearing on "Cybersecurity—Getting it Right", Transcript at https://archive.org/details/gov.gpo.fdsys.CHRG-108hhrg98150, 108th Cong. (July 22, 2003) (statement of Steven M. Bellovin), https://www.cs.columbia.edu/~smb/papers/Statement.pdf.

6. Steven M. Bellovin, "Cybersecurity for Small Businesses": Testimony for the New York City Council Committee on Technology and Committee on Small Business, New York City Council (Feb. 25, 2020) (statement of Steven M. Bellovin), https://www.cs.columbia.edu/~smb/papers/nyc-council-testimony.pdf.

7. Steven M. Bellovin, *Thinking Security: Stopping Next Year's Hackers* (Boston: Addison-Wesley, 2016), Chapter 15, ISBN: 978-0-13-427754-7, http://www.informit.com/store/thinking-security-stopping-next-years-hackers-9780134277547.

8. See IRS, *Frequently Asked Questions: Depreciation & Recapture*, October 14, 2020, https://www.irs.gov/faqs/sale-or-trade-of-business-depreciation-rentals/depreciation-recapture for details.

Even here, though, cloud providers have an edge: they have sufficient capacity that customers can avoid the need to patch a running computer. Instead, one can simply bring up new, already-patched virtual machines and switch the application to them, knowing that it's easy to revert to the older version if necessary. To be sure, this takes customer effort: it requires planning, automated build and configuration management tools, and more. But it is easier to implement on a cloud platform.

The major cloud providers also provide good support for better forms of authentication than passwords. It is hard to overstate how important this is; many devastating attacks, such as Russia's 2016 hack of John Podesta's emails, were due to reliance on passwords.[9] Ubiquitous support for, e.g., FIDO2 tokens, an issue I discussed here in February,[10] is crucial, and it is much more easily done via cloud providers—they've already implemented it.

There is a more subtle benefit to using someone else's authentication service. If it is ever necessary to audit use or abuse of credentials, a service operated by an outside party can maintain logs not accessible to a rogue system administrator—and rogue system administrators can happen.[11] (I should note that cloud authentication services are generally designed so that not even the provider can steal someone's credentials.)

In other words, there is no one answer to the question, "Is cloud computing more secure?" Apart from the risk of mischief by the provider's employees or of security problems in their own infrastructure—and those risks, though low, certainly exist—there are enough different models for cloud computing and for that matter local administration that an accurate assessment depends entirely on what service is being proposed and what it would replace.

Cloud Applications

What I call "cloud applications" are the least understood, most important, and most fraught aspect of cloud computing. Cloud applications are just that: ordinary applications, but run by the cloud provider. The application could be something as commonplace as email—we're all familiar with Google's `gmail.com` and Microsoft's `outlook.com`—or quite complex. AWS offers things like databases, machine learning systems, facial recognition, video analysis, and more.[12]

Cloud services offer the greatest potential for both cost savings and for security improvements. They are administered by the cloud providers; as noted, they have learned to do this well, and at scale. Furthermore, the applications are often far more sophisticated than what an individual organization is capable of developing or managing on its own. Spam filtering accuracy is entirely dependent on scale; a company that sees

9. Nicole Perlroth and Michael D. Shear, "Private Security Group Says Russia Was Behind John Podesta's Email Hack," October 20, 2016, https://www.nytimes.com/2016/10/21/us/private-security-group-says-russia-was-behind-john-podestas-email-hack.html.

10. Bellovin, "Cybersecurity for Small Businesses".

11. Richard Esposito, Matthew Cole, and Robert Windrem, "Snowden Impersonated NSA Officials, Sources Say," *NBC News*, August 29, 2013, http://investigations.nbcnews.com/_news/2013/08/29/20234171-snowden-impersonated-nsa-officials-sources-say.

12. See, e.g., *Find the hands-on tutorials for your AWS needs*, November 22, 2020, https://aws.amazon.com/getting-started/hands-on/?awsm.page-getting-started-all=1. Every cloud provider will of course have its own service offereings.

billions of messages a day—Gmail has about 1.8 billion users[13]—has a far better grasp of what spam is than any single organization's email can show.

Cloud-based web hosting is almost the norm. The web site from which I read this bill, https://legistar.council.nyc.gov/, appears to be run by a private company, Granicus, which specializes in digital government applications; it in turn is owned by a venture capital firm.[14] There is not only nothing wrong with using this service, it is an excellent practice. It saved the city the money that would have been needed to develop and run a custom application.

Cloud applications can, of course, pose security risks. Foremost among them is the necessity to administer them correctly. A Google search for "open s3 bucket"—sites that have not properly protected their AWS cloud storage—will return many hits.[15] Overall, though, with a modicum of care they are likely to be far more secure than in-house alternatives.

There is, however, a subtle danger: lock-in. Program access to, say, an AWS facial recognition service is likely to be different than the equivalent Microsoft Azure service. In other words, migrating from AWS to Azure or to an in-house offering is likely to require significant reprogramming and hence significant expense. This does not mean that it is a bad idea to use these cloud services—as noted, they can be more secure and more functional than locally developed versions—but the cost of migrating away from them should be taken into account when making a decision.

Cloud Availability

One major benefit of cloud services is high availability: applications run on whichever virtual machines are available. If some hardware fails, the application is migrated to another VM. That can be done locally, too, of course, but only if there is sufficient extra hardware. With a cloud provider, the cost of the extra hardware can be amortized over many more physical machines.

The large cloud providers also have multiple data centers, both in the US and around the world. There can thus be redundancy even against the failure of a single data center. Add-on services include applications that automatically synchronize databases between different disks and different sites.

To be sure, even the large cloud providers can experience outages. Google had two significant outages this week,[16] AWS experienced a major hiccup within the last

13. Christo Petrov, "50 Gmail Statistics To Show How Big It Is In 2020," *TechJury*, June 30, 2020, https://techjury.net/blog/gmail-statistics/.

14. Tamara Chuang, "Vista Equity buys another Denver tech firm, Granicus," *Denver Post*, August 18, 2016, https://www.denverpost.com/2016/08/18/vista-equity-buys-denver-tech-firm-granicus/.

15. Yifat Perry, "Amazon S3 Bucket Security: How to Find Open Buckets and Keep Them Safe," *NetApp.com Blog*, September 14, 2020, https://cloud.netapp.com/blog/aws-cvo-blg-amazon-s3-buckets-finding-open-buckets-with-grayhat-warfare.

16. Adam Satariano, "Google's apps crash in a worldwide outage.," *New York Times*, December 14, 2020, https://www.nytimes.com/2020/12/11/business/google-down-worldwide.html; Ron Amadeo, "Google sees major services outages two days in a row," *Ars Technica*, December 16, 2020, https://arstechnica.com/gadgets/2020/12/google-sees-major-services-outages-two-days-in-a-row/.

month,[17] and Microsoft has also had problems.[18] Such outages are very visible because their scale: many customers are affected. It is likely, though, that if their customers all ran their own servers, their aggregate outage time would be far greater than they have experienced via their cloud providers. Critical services can fail, even if—or because–they are locally run. New York City has experienced at least two significant failures of its 911 system,[19] though of course more centralized systems can experience similar issues.[20]

Recommendations

As I noted at the start, New York City is large enough that it could have its own first-class computing infrastructure. It is not clear, though, that this would be the most cost-effective path. Making the right decision will take considerable effort.

New York City might be able to do it itself: It might be feasible for the city to run its own information technology infrastructure at standards of reliability approaching that of cloud providers. Doing this will require a long-term commitment of stable funding and an ongoing program of equipment and software modernization.

I cannot speak to New York City's experiences; in the Federal government, IT modernization efforts have sometimes been delayed by budget sequestrations due to political deadlocks. For example, the GAO reported that the US Patent and Trademark Office "reduced spending on IT modernization contracts by $80 million,"[21] the Alcohol and Tobacco Tax and Trade Bureau dealt with its cuts by delaying IT updates,[22] the Nuclear Regulatory Commission delayed or rescoped contracts for its modernization efforts,[23] and so on.

For reliability, the city would need to create a backup data center To guard against problems that affect the entire city—think of Hurricane Sandy—and to reduce costs for electricity and perhaps real estate, it may be wise to locate this center

17. Jay Greene, "Amazon Web Services outage hobbles businesses," *Washington Post*, November 25, 2020, https://www.washingtonpost.com/business/economy/amazon-web-services-outage-stymies-businesses/2020/11/25/b54a6106-2f4f-11eb-860d-f7999599cbc2_story.html; Jay Greene, "Amazon's cloud-computing outage on Wednesday was triggered by effort to boost system's capacity," *Washington Post*, November 28, 2020, https://www.washingtonpost.com/technology/2020/11/28/amazon-outage-explained/.

18. Ed Targett, "Microsoft Azure Throttles Cloud Access, Blames Capacity Crunch," *Computer Business Review*, April 2, 2020, https://www.cbronline.com/news/microsoft-azure-capacity-crunch.

19. Michael Cooper, "Reviews of 911 System Are Promised After Hourlong Failure," *New York Times*, February 2, 1999, https://www.nytimes.com/1999/02/02/nyregion/reviews-of-911-system-are-promised-after-hourlong-failure.html; Michael Schwirtz, "Tracking the Storm: New York's 911 System Overloaded," *New York Times*, October 28, 2012, https://cityroom.blogs.nytimes.com/2012/10/28/hurricane-sandy-live-updates/#new-yorks-911-system-overloaded.

20. Brian Krebs, "Who's Behind Monday's 14-State 911 Outage?," *Krebs on Security*, September 29, 2020, https://krebsonsecurity.com/2020/09/whos-behind-mondays-14-state-911-outage/.

21. GAO, *Report to the Chairman, Committee on the Budget, House of Representatives: 2013 Sequestration*, GAO- 14-244 (Government Accountability Office, March 2014), p. 73, https://www.gao.gov/assets/670/661444.pdf.

22. Ibid., p. 151.

23. Ibid., p. 179.

elsewhere. (Relatively speaking, the impact on jobs will not be large; data centers do not have very many employees per square foot of floor area except during construction.) Using a commercial cloud service, with live storage but standby computing, is likely the most economical way to do create backup capacity. Doing this could lead to significant savings in real estate costs.

But it isn't clear that this would save money compared with commercial cloud services. They scale-dependent advantages to cloud services cannot be replicated by any single entity, not even New York City.

Using cloud computing and services As described earlier, cloud storage can be hard to use effectively unless cloud computing or cloud applications are used. If the decision is made to move some storage to a cloud provider, serious consideration should be given to moving the associated processing as well.

The need for very high availability is another possible driver for cloud computing. Although the principles underlying the mechanisms are well known, implementation is not trivial, and data replication may have to be added to many programs. However, the software underlying many cloud storage and cloud application services was designed with replication in mind. This pre-existing infrastructure can make it much easier to implement extremely reliable services.

The most intriguing advantages come from the ability to use high-level cloud services. On the one hand, there is the chance to build creative new applications, and to do so at a far lower cost than starting from scratch. On the other hand, this needs to be weighed against the lock-in problem. There is no "one size fits all" answer to these questions; they must be answered separately for each IT subsystem.

I stress that simply moving the existing service architectures to the cloud, without modification—a strategy known in the business as "lift and shift"—is unlikely to produce savings. The real benefit from the cloud is the ability to work in a completely different way.

Changes to DoITT The more functionality that DoITT moves to cloud applications, the less need there will be for in-house system administration—cloud providers do that for their own application platforms. Furthermore, such a move will almost certainly improve the security of the city's applications.

This does not, however, mean that the city does not need IT expertise. Indeed, in many ways it needs more expertise just to manage outside contractors. A National Academies report (disclaimer: I was on the study committee) noted that about a Federal Aviation Administration effort:[24]

> Even if the FAA were not acting as systems integrator, it would still need to be a "smart customer,"—meaning that it needs expertise

24. David E. Liddle and Lynette I. Millett, eds., *A Review of the Next Generation Air Transportation System: Implications and Importance of System Architecture* (Washington, DC: National Academies Press, 2015), p. 77, https://www.nap.edu/catalog/21721/a-review-of-the-next-generation-air-transportation-system-implications.

that will enable it to effectively structure and manage its supplier relationships.

Developing and retaining this expertise will be a challenge. However, proceeding with inexperienced or less than the best personnel in key leadership positions is a significant risk.

The need for in-house expertise does not stop at the project management level. Administering cloud-resident resources is still necessary, to avoid errors such as the open bucket problem. Programmers will still be needed to effectively utilize the cloud applications.

Finally, great care in system design is needed to minimize the lock-in problem: while there is likely no way to avoid having platform-specific modules, good software engineering can provide the software abstractions to isolate such dependencies, making them much easier to replace if necessary.

Making the Decision As I've outlined here, there are many complex, interacting factors that affect the city's possible decision to use assorted cloud services. Given the complexity of the city's existing IT operations, it is clear that DoITT involvement in the process is utterly vital—no one outside would have nearly enough knowledge of what is done today.

However, given that some scenarios would imply significant changes to DoITT, it is important to that the decision process include outsiders as well. These people should have significant operational experience with cloud services and data center operations, and they should work closely with the experts in the DoITT.

Overall, I recommend an initial trial deployment of some services on a cloud platform. Moving an existing service, in any form other than lift-and-shift, is a considerable effort; as noted, though, lift-and-shift will not result in nearly as many benefits. I recommend, then, that the next few systems that need a significant hardware or software upgrade be rebuilt as "cloud-native", and (if necessary) that the city procure or provide proper training to developers on how to do this well. This will provide better baseline data on the costs and benefits of such moves.

To sum up: the issues in use of cloud platforms are complex and there is no simple answer. Furthermore, it is not an all-or-nothing decision; some services may be best kept in-house while others are moved to the cloud. But the advantages can be compelling; the city should at least try.

Biography

Steven M. Bellovin is the Percy K. and Vida L. W. Hudson Professor of Computer Science at Columbia University, a member of the Cybersecurity and Privacy Center of the university's Data Science Institute, and an affiliate faculty member at Columbia Law School. Bellovin does research on security and privacy and on related public policy issues. In his copious spare professional time, he does some work on the history of cryptography. He joined the faculty in 2005 after many years at Bell Labs and AT&T Labs Research, where he was an AT&T Fellow. He received a BA degree from Columbia University, and an MS and PhD in Computer Science from the University of North Carolina at Chapel Hill. While a graduate student, he helped create Netnews; for this, he and the other perpetrators were given the 1995 Usenix Lifetime Achievement Award (The Flame). He has also received the 2007 NIST/NSA National Computer Systems Security Award and has been elected to the Cybersecurity Hall of Fame. Bellovin has served as Chief Technologist of the Federal Trade Commission and as the Technology Scholar at the Privacy and Civil Liberties Oversight Board. He is a member of the National Academy of Engineering and has served on the Computer Science and Telecommunications Board of the National Academies of Sciences, Engineering, and Medicine. In the past, he has been a member of the Department of Homeland Security's Science and Technology Advisory Committee, and the Technical Guidelines Development Committee of the Election Assistance Commission.

Bellovin is the author of *Thinking Security* and the co-author of *Firewalls and Internet Security: Repelling the Wily Hacker*, and holds a number of patents on cryptographic and network protocols. He has served on many National Academies study committees, including those on information systems trustworthiness, the privacy implications of authentication technologies, and cybersecurity research needs; he was also on science versus terrorism. He was a member of the Internet Architecture Board from 1996-2002; he was co-director of the Security Area of the Internet Engineering Task Force from 2002 through 2004.

More details may be found at http://www.cs.columbia.edu/~smb/.