Could we fit the Internet in a Box?

Gareth Tyson$^\diamond$, Arjuna Sathiaseelan$^\bullet$ and Jörg Ott†

$^\diamond$Queen Mary University of London, $^\bullet$University of Cambridge, †Aalto University

Abstract. It is estimated that only 40% of the world's households are connected to the Internet. Half of them are in less developed countries, where household Internet penetration has only reached 28%. This is in stark contrast to the 78% of households in more developed countries. A key challenge facing the next generation is therefore enabling wider participation in the Internet, as well as the benefits it brings. This paper explores the feasibility of capturing network applications and services in a single locally usable "Internet Box". The Box will operate independently from the rest of the Internet, allowing those without traditional connectivity to use the "Internet" in an simulated and disconnected manner. We conclude that the concepts have great potential, and explore some of the remaining challenges, as well as milestones achieved in the literature so far.

1 Introduction

It is estimated that only 37.9% of the world's population have Internet access [1]. In developing countries, this value is only 29.9%, in stark contrast to 75.7% in the developed world. The economic benefits of Internet access are also disproportionately biased, with developed economies ahead by a factor of 25% [2]. Despite this, remarkably, even some wealthy countries suffer from similar problems; in New Orleans, for example, the poorer wards have broadband subscription rates between only 0 and 40%. A key challenge facing the next generation is therefore enabling wider participation in the Internet, as well as the benefits it brings. Before facing this challenge, it is first necessary to ask one question: What is the Internet? If you ask network engineers, they would explain the many details of TCP/IP. However, the everyday person might likely respond with services available via the Internet such as Google, Facebook, Twitter or Netflix. Hence, we argue that users want access to services — they are not concerned about how they are delivered.

The above observation is a powerful one because it relaxes some of the constraints on deploying the "Internet" to new regions. Specifically, we explore the feasibility of capturing network applications and services in a single locally usable box which we call the "Internet Box". The Box will operate independently from the rest of the Internet, allowing those without traditional connectivity to use the "Internet" in an simulated and even (partly) disconnected manner. For example, locally caching a copy of a map on the Box would allow a user to access it regardless of the backhaul availability. This is shown in Figure 1, where a user connects to the local Box via WiFi to download a video, which is locally stored.

Beyond this, there might be periodic backhaul connectivity to cloud service(s) that can provide fresh content when available. Advanced models could be built to make this process automatic. For instance, predictive algorithms could be developed to predict the content that will be requested from the Box in the future. Such algorithms could be executed whenever the Box achieves connectivity with the wider Internet. Through this, all future required resources could be pushed or pre-fetched and stored in the Box in anticipation of their usage.[1] If these principles are proven feasible, many localities that possess no connectivity could hopefully start to use Internet services immediately. The paper explores why this is a positive first step towards global access for all.

Recent efforts towards an Internet Box have already resulted in deployments. Most prominently, a project actually named "Internet in a Box" has had multiple successful deployments [4]. Such projects, however, face a number of challenges. An obvious problem that emerges is the inevitable lack of resources available in such environments. Storage and local network limitations, alongside poor backhaul connectivity, can result in many users being frustrated with the service they receive. Traditionally, such problems would be dealt with via typical "fair share" algorithms that operate agnostic to the higher level applications (e.g. TCP congestion control). However, with extremely limited resources, these fair share algorithms can result in everybody receiving a unusable quality of service (much like experienced during past congestion collapse events [5]). Similarly, existing Box implementations operate in very constrained ways, generally just providing static content. The "Web 2.0" revolution has proven that this limitation is not long-term viable and, therefore, we also argue that the Box should be equipped with local services (e.g. social networks, picture sharing) that can be accessed by citizens.

This paper explores the potential of deploying a Box possessing local storage and computation into areas that have very limited Internet access (e.g. rural villages). Section 2 surveys related work in the area and highlights milestone implementations and deployments of these principles so far. Section 3 details the key system components that must exist within the box to correctly operate. We explore how the Box might be implemented and used in Section 4. Finally, in Section 5, we conclude the paper, highlighting further interesting avenues of work.

2 Background and Related Work

Recent years have witnessed several projects that propose concepts related to an Internet Box. Most prominent is the aptly named Internet in a Box initiative [4]. This is a small networked device that is pre-loaded with curated content, e.g. books, Wikipedia. The box is enabled with local WiFi connectivity, allowing

---

[1] Flavours of this idea were suggested for connecting users accessing the Internet via satellite, e.g. to save capacity of satellite channels when serving the same content to many users [3] and for intermittently connected users (c.f. http://www.gedanken.org.uk/software/wwwoffle/).

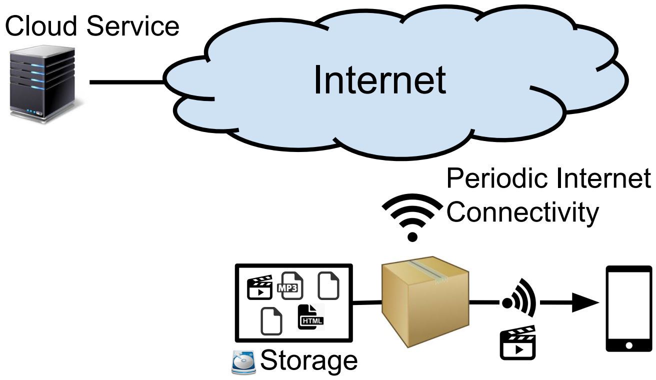

Fig. 1. Overview of the Box. A smartphone connects the the Box and requests a video that is transferred via WiFi. The film is locally stored on the Box and is therefore provided regardless of backhaul Internet connectivity. The Box may also have periodic Internet connectivity, allowing it to communicate with remote cloud services that can provide updated content.

users to connect and access the locally stored content. A similar approach is taken by Liberouter [6], which offers access to generic content that might be of interest to a given neighbourhood. A common use case is exploiting this type of equipment for learning purposes. The Open Learning Toolkit [7] is one such example; it exposes an HTML5 interface to allow learners (generally school children) to access various resources such as textbooks. It can operate in both connected and disconnected modes, the former allowing new content to be fetched. Other closely related systems include The Library For All [8], and the Library Box [9].

An alternative to disconnected devices is Outernet [10], which takes a unique approach to content delivery. Rather than following the Internet's usual request-response model, where a client (e.g. the Box) requests content, Outernet *broadcasts* content via Satellite. Any Boxes within range, will therefore passively receive new bundles of content (much like a TV receives broadcast signals). Recent work has also explored the possibility of using 4G broadcast signals for similar aims [11]. Preliminary evaluations of this concept have been very positive; in [11], the authors found that a 100 MB broadcast of content could preempt upto 40% of requests. Work has also been performed into pre-fetching particular types of content; for example, SCORE [12] predicts users' catch-up video consumption patterns and automatically records videos from TV broadcast signals, rather than leaving the user to fetch content over IP (thereby saving energy and network overheads). A key enabler to these technologies is past measurement studies that have explored how users access the Internet. There have been a handful of studies into how people in developing regions use the Internet (e.g. which services they access). Most noteworthy is Johnson *et al.*; the authors inspected a two month network trace in Zambia; this, amongst other things, showed the impact of traffic from content-heavy peer-to-peer systems. They used this insight to design VillageShare, which operates as a time-delayed proxy for use in rural villages.

There are a number of remaining technical challenges that needs to be addressed in this domain. There are obviously key hardware issues that must be faced, particularly relating to energy management and cost. Building reliable devices at affordable prices (that can be deployed in developing regions) is not trivial. This is exacerbated by the fact that environments are often quite severe, where devices may be exposed to extreme weather (e.g. very high or low temperatures). Another huge challenge is the loading of appropriate content onto the Box. In existing projects, this is generally done in a relatively manual way, whereby content is curated and selected by an appropriate party. There are techniques for citizens to request specific content, but this is often cumbersome. For example, the Outernet allows users to request content be added to the storage by SMS. As such, it would be much better if this could be an automated process, where the Box learnt and automatically loaded "optimal' content for its locale. In either case, it is likely that some users would be disappointed by the choices made, leading to the need for interaction between users and the Box to find a compromise (c.f. interactive web caching [13]).

Another key limitation of existing solutions is their lack of support for dynamic services. Whereas it is relatively easy to load a Box with static content, it is much more difficult to host and execute services. The benefits of achieving this goal are significant. It would allow developers (both local and global) to instantiate beneficial services that could perform tasks far beyond simple content provision. A classic example would be to introduce a local social networking services that allows people to share photos etc. There are many conceivable services that could be used for more practical purposes too. For instance, a crop disease diagnosis service could be provided that allow farmers to upload pictures of diseased crops; these could then be automatically analysed to return treatment advice. Although deploying services like this into edge networks has yet to become mainstream, new technologies (e.g. unikernels [14] or service-centric networks [15]) makes this increasingly possible.

Of course, all the above concepts further depend on one thing: Being able to transport content and/or services from the source to the Box post-deployment. Without such a capability, the Box would forever remain at its factory default. A key challenge is therefore finding and exploiting (intermittent) low cost backhaul connectivity. This would most likely be occasional satellite communications, broadcast-style delivery and/or the ability to periodiccally move the Box to an area that has connectivity. A closely related topic to this is that of delay-tolerant networks (DTNs). A DTN is a type of network that supports the existence of significant delays or disruptions between sending and receiving parties [16]. Specifically, DTNs propose a store-and-forward architecture in which data units, termed *bundles*, can be temporarily stored at nodes (during network disruptions) until an appropriate next hop can be found [17]. Traditionally, these disruptions and delays could be caused by long distances (e.g. interstellar communications

[18]) or, alternatively, by network partitions. The latter is the case for the Box as, in essence, the Box operates as a disconnected component of the network. These principles were made available, for example, in the KioskNet project [19], which introduced boxes into known small kiosks. Passing vehicles (e.g. buses) would then carry content and load it into the kiosks as they drove past.

Another closely related concept is that of information-centric networks (ICNs). An ICN is a network with the sole purpose of delivering content [20][21][22]. As such, an ICN exposes a publish/subscribe style abstraction unlike the existing Socket API [23]. This is because a host-centric network (HCN) is designed to scalably route packets from a source to a destination, whilst an ICN is designed to scalably deliver content from providers to consumers. This is clearly a technology that maps closely to the Box, which primarily is oriented towards the delivery of content. Prominent examples of ICNs include DONA [24], PURSUIT [21], Named Data Networking (NDN) [20] and Juno [25]. We believe that these principles combined (ICN and DTN) could offer a strong foundation to build Box technology. This could then be combined with emerging service-oriented technologies (e.g. Jitsu [14]) to allow a Box to run local services. Our past work has taken the first steps towards this by integrating ICN and DTN functionality [26]. A further overview of key scenarios and uses for ICN can be found here [27].

3 Strawman Requirements and Design

We next detail a strawman design of the Box, exploring key components that must be built. There are four architectural components; each constitute both hardware and software

3.1 Storage

To enable totally disconnected operation, it is necessary to maintain a significant amount of local storage in the Box. This allows popular objects (both content and services) to be served locally without needing to use any egress network connectivity. In the simplest case, this storage device could be pre-loaded with important content (e.g. Wikipedia, medical information) at the factory. Storage is a cheap commodity, with multi-TB drives available for ≈£50 (and below). This might constitute significant capital expenditure in a developing region, but it could be treated as a long-term investment for a community. Perhaps the most notable bottleneck is energy consumption. Mechanical disks, although cheap, consume large energy quantities due to the need for moving parts (although low energy mechanical storage disks are available). In contrast, Solid State Drives are much more energy efficient, although financial cost is greater.

Beyond these hardware considerations, it is clearly also necessary to build intelligent algorithms that can appropriately select what is stored. In situations where content and services are statically loaded at the factory, this is less important. However, if there is any level of content/service churn over time, this must be managed in an appropriate manner (more information is provided in Section 4). This is closely related to recent work on ICN caching [28], as well as work on pre-fetching, which aims to predict the content requests before they happen [12]. In these cases, the challenge would be to perfectly predict all requests by users so that they can be pre-emptively stored on the Box in advance.

3.2 Local Connectivity

Clearly, it is necessary to provide some mechanism by which citizens can gain access to the Box. The simplest approach would be to introduce a WiFi network interface. Of course, this assumes that users nearby own a device capable of connecting. This would be the case for all network alternatives, e.g. Bluetooth or wired Ethernet. A different approach would be to introduce some sort of user interface (touch screen) to the Box itself. On the one hand, this would make the box more self contained, however, on the other hand, it would limit the number of people who could simultaneously use it. Further, it would dramatically increase complexity, cost and energy consumption. As such, providing simple WiFi communications, that can be accessed using a low-cost tablet, would be preferential.

A further use of local connectivity would be to update the storage with fresh content. One possibility would be to use vehicles, which can carry content and transmit it to the box when they come into range (like KioskNet [19]). Alternatively, somebody could physically move the Box to an area that has fresh content to download. Of course, this would all require seamless mechanisms that automatically update the Box without human intervention (i.e. as soon a the Box comes into range of another content source).

3.3 Backhaul Connectivity

An optional addition to the Box would be a network interface that provides backhaul connectivity (i.e. wide area Internet communications). This would be highly desirable, as it would allow the box to refresh the content and services it stores without the need to move the box or to use a KioskNet-like mechanism. Unfortunately, the regions where the Box might be deployed are unlikely to easily support widescale communications at a low cost. Consequently, the box must be able to operate with and without Internet connectivity. Particularly, the Box should support on-off backhaul support, where connectivity is intermittent. By enabling all Box deployments with both capabilities, we also argue that this can more seamlessly allow citizens to transition to “real” Internet access when wide area connectivity is made available in a region.

3.4 Service Management

As previously stated, it would be beneficial if the Box could support functionality beyond static content storage. Running a small service platform within the

Box would allow things like local social networks and even games to be hosted. Obviously the capabilities of these would be limited by the resources of the Box and, therefore, service mangement in this context would be extremely important. Specifically, it would be necessary to build a platform that can securely host *very* lightweight services with the ability to individually control and limit their resource consumption (based on a range of requirements). These constraints mean that running large numbers of complex services would be difficult. Instead, specific techniques would be needed to strictly limit the operations of each service in a fair manner.

3.5 Remote Management

As stated above, the Box would be ill-suited to running and managing complex services and tasks locally. For example, deciding what to store on the Box could potentially involve complicated operations (compiling user histories, modelling their interests, predicting future request patterns, selecting a subset of content to request). Due to cost constraints, for some deployments, it would be undesirable to build the Box with sufficient memory and processing capacity to execute all these algorithms. It would therefore be desirable to be able to offload some of these tasks to remote execution (i.e. in a cloud). This could operate in a similar manner to CDroid [29] or ThinkAir [30] for mobile phones, where tasks are split between local and remote execution. Through this, the computational load on the Box could be reduced. Complexity in a disconnected scenario, however, is higher. For instance, the Box might not be able to communicate with the remote cloud service for an extended period. This would be particularly problematic if data must be kept up-to-date on both the Box and the cloud. As such, techniques must be developed to balance the need to offload computation with the limitations of the backahul communications availability.

4 Implementation Levels

There are several potential stages to the implementation. Currently, a small set of hardware systems have been built (e.g. [4]). The key challenge remaining is therefore devising means by which the resources of this hardware can be appropriately managed and extended to support more sophisticated behaviour (e.g. services). Here, we devise several key implementation steps for our future work:

1. The simplest implementation would be to provide a single Box that possesses static storage and local network connectivity. The Box would contain a large block of storage that is filled with curated static content (decided at the factory). Local WiFi would allow users to connect to the Box. Upon connection, users would be able to access a locally available portal that gives access to all content on the Box (e.g. via a web browser). In line with common usage, the most sensible portal might be a search engine interface with more structured access to content categories (similar to interactive web caching [13]).

2. The next implementation stage would extend the Box to support user uploaded content. This would allow users to create and share their own static media (e.g. webpages, videos). The uploaded content would then be integrated into the portal to make it locally available to all other users. Of course, this is accompanied by various audit, privacy and security challenges that would need addressing.
3. The next implementation stage would be to introduce active services onto the Box. This would move beyond static content provision. Potential services could be online social networks and voice communications. This would be limited to local interactions; for example, a social network service would only provide accounts for local users. Such services could be built over a unikernel platform like Jiitsu that allow extremely lightweight virtual machines to operate in resources constrained environments [14]; each service would exist as its own micro virtual machine that could be executed.
4. The next step would be to add one-way inbound communications to the Box. This is likely to be periodic and, potentially, unpredictable. This would allow the Box to receive one-way information from external parties. For example, this could involve satellite or radio broadcasts of fresh content/services (similar to Outernet [10]).
5. The next tier would introduce two-way communications. However, this will not necessarily be synchronous. Request/response intervals could be in the order of hours or days. Further, there may be extended periods without any connectivity whatsoever. This would allow information to be uploaded from the Box to third parties (e.g. offloaded cloud services).
6. Finally, the most advanced stage of implementation would be to add two-way synchronous connectivity. This would, in essence, provide full Internet connectivity to citizens, which is the final goal of our work.

5 Conclusion

This paper has explored a potential approach to deploying Internet-like services in areas that currently struggle to gain wide area connectivity. There are a number of challenges that remain. We therefore conclude with a brief summary of future issues to address. A particular problem is how such a Box could be filled with content and services. Currently, the predominant approach is through curated streams and content packages chosen by third parties. This raises a number of issues. It poses technical challenges, such as how dynamic shifts in content or service demand could be reacted to. Beyond this, it also raises ethical questions: Who is qualified to have control over what content should be accessible by a group of people? Making such as process transparent would not be trivial. Transporting content and services from the Internet onto the Box poses another key challenge. Clearly, this is straightforward if the Box is solely preloaded with content/services at the factory (i.e. there is no need to transport anything). However, if the Box is to be updated over time, it is necessary to execute algorithms that can decide what should be stored. A number of possibilities have been discussed in this paper, however, the most appropriate is not clear.

An open issue that we have not touched upon is that of commercialisation. It is likely many stakeholders (e.g. Box manufacturers, content providers) would need to monetise their "products" somehow. For example, how would content producers make money from uploading their content onto the Box? This is an issue that we leave for future exploration. However, to produce sustainable deployments, it is likely that this challenge must be faced.

It is important to finalise by saying that the purpose behind our work is *not* to create a two-tier Internet, in which certain people are limited to locally available content/services provided by the Box. Rather, it is intended as an intermediate step to better enable future deployments. Consequently, the nature of these future deployments must be considered. An obvious possibility is to introduce more sophisticated features to the Box, e.g. various novel services, that offer near identical support to the wider Internet. Of course, the next step beyond this would likely be to offer "real" Internet access, where the Box operates as a simple gateway (with a cache). An attractive property of the Box would be that by this point in time, people would have been given the opportunity to become familiar with the technology. Thus, the transition for local communications to global would be more seamless.

Acknowledgment

Arjuna Sathiaseelan and Jörg Ott were funded by the EC H2020 RIFE project, grant reference 644663

References

1. U. N. International Telecommunication Union (ITU), "World telecommunication/ict indicators database." http://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx, 2013.
2. Strategy Dynamics Global SA, "The emerging digital economy." https://www.globaltrends.com/monthly-briefings/60-monthly-briefings/192-gt-briefing-june-2013-the-digital-economy, 2013.
3. M. Rabinovich and O. Spatscheck, "Web caching and replication," *Sigmod Record*, vol. 32, no. 4, p. 107, 2003.
4. "Internet in a box." http://internet-in-a-box.org/.
5. B. Braden, D. Clark, J. Crowcroft, B. Davie, S. Deering, D. Estrin, S. Floyd, V. Jacobson, G. Minshall, C. Partridge, *et al.*, "Recommendations on queue management and congestion avoidance in the internet," 1998.
6. T. Karkkainen and J. Ott, "Liberouter: Towards autonomous neighborhood networking," in *Wireless On-demand Network Systems and Services (WONS), 2014 11th Annual Conference on*, pp. 162–169, IEEE, 2014.
7. "Open learning toolkit." http://www.col.org/progServ/programmes/KM/Pages/Aptus.aspx.
8. "Library for all." https://www.libraryforall.org/.
9. "Library box." http://librarybox.us/.
10. "Outernet." https://www.outernet.is/en/.
11. A. Finamore, M. Mellia, Z. Gilani, K. Papagiannaki, V. Erramilli, and Y. Grunenberger, "Is there a case for mobile phone content pre-staging?," in *Proceedings of the ninth ACM conference on Emerging networking experiments and technologies*, pp. 321–326, ACM, 2013.

12. G. Nencioni, N. Sastry, J. Chandaria, and J. Crowcroft, “Understanding and decreasing the network footprint of catch-up tv,” in *Proceedings of the 22nd international conference on WWW*, 2013.
13. J. Chen and L. Subramanian, “Interactive web caching for slow or intermittent networks,” in *Proceedings of the 4th Annual Symposium on Computing for Development*, p. 5, ACM, 2013.
14. A. Madhavapeddy, T. Leonard, M. Skjegstad, *et al.*, “Jitsu: Just-in-time summoning of unikernels,” in *12th USENIX Symposium on Networked System Design and Implementation*, 2015.
15. T. Braun, V. Hilt, M. Hofmann, I. Rimac, M. Steiner, and M. Varvello, “Service-centric networking,” in *Communications Workshops (ICC), 2011 IEEE International Conference on*, pp. 1–6, IEEE, 2011.
16. K. Fall, “A delay-tolerant network architecture for challenged internets,” in *Proc. SIGCOMM*, 2003.
17. K. Fall and S. Burleigh, “Bundle Protocol Specification.” RFC 5206 (Experimental), Nov. 2007.
18. S. Burleigh, A. Hooke, L. Torgerson, K. Fall, V. Cerf, B. Durst, K. Scott, and H. Weiss, “Delay-tolerant networking: an approach to interplanetary internet,” *Communications Magazine, IEEE*, vol. 41, pp. 128 – 136, june 2003.
19. S. Guo, M. Derakhshani, M. H. Falaki, U. Ismail, R. Luk, E. A. Oliver, S. U. Rahman, A. Seth, M. A. Zaharia, and S. Keshav, “Design and implementation of the kiosknet system,” *Computer Networks*, vol. 55, no. 1, pp. 264–281, 2011.
20. V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, “Networking named content,” in *In Proc. 5th ACM CoNEXT*, 2009.
21. D. Trossen and G. Parisis, “Designing and realizing an information-centric internet,” *IEEE Communications Magazine*, vol. 50, july 2012.
22. D. Kutscher, S. Farrell, and E. Davies, “The NetInf Protocol.” Internet Draft, 2012.
23. G. Tyson, A. Mauthe, S. Kaune, P. Grace, and T. Plagemann, “Juno: An adptive delivery-centric middleware,” in *Proc. 4th Intl. Workshop on Future Media Networking (FMN)*, 2012.
24. T. Koponen, M. Chawla, B.-G. Chun, A. Ermolinskiy, K. H. Kim, S. Shenker, and I. Stoica, “A data-oriented (and beyond) network architecture,” *SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 181–192, 2007.
25. G. Tyson, A. Mauthe, S. Kaune, P. Grace, A. Taweel, and T. Plagemann, “Juno: A middleware platform for supporting delivery-centric applications,” *ACM Transactions on Internet Technology*, 2012.
26. G. Tyson, J. Bigham, and E. Bodanese, “Towards an information-centric delay-tolerant network,” in *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*, pp. 387–392, IEEE, 2013.
27. K. Pentikousis, B. Ohlman, D. Corujo, *et al.*, “Information centric networking: baseline scenarios,” tech. rep., ICNRG Internet-Draft., 2014.
28. G. Tyson, S. Kaune, S. Miles, Y. El-Khatib, A. Mauthe, and A. Taweel, “A trace-driven analysis of caching in content-centric networks,” in *Proc. Intl. Conference on Computer Communication Networks (ICCCN)*, 2012.
29. M. Barbera, S. Kosta, A. Mei, V. Perta, and J. Stefa, “Cdroid: Towards a cloud-integrated mobile operating system,” in *Computer Communications Workshops (INFOCOM WKSHPS), 2013 IEEE Conference on*, pp. 47–48, IEEE, 2013.
30. S. Kosta, A. Aucinas, P. Hui, R. Mortier, and X. Zhang, “Thinkair: Dynamic resource allocation and parallel execution in the cloud for mobile code offloading,” in *INFOCOM, 2012 Proceedings IEEE*, pp. 945–953, IEEE, 2012.