

Proposal for Next-Generation Commercial Print Infrastructure: Gutenberg-Landa TCP/IP

I-Jong Lin, Jun Zeng, Eric Hoarau and Gary Dispoto; Hewlett-Packard Laboratories; Palo Alto, CA

Abstract

Using the Transmission Control Protocol/ Internet Protocol (TCP/IP) stack as a model of layered protocols, we propose a next-generation reliable print service called the Gutenberg-Landa TCP/IP to handle the routing of pages through a Print Service Provider (PSP). Gutenberg-Landa TCP/IP will be an expanded TCP/IP protocol: just as data packets flow through the Internet to form web pages, so printed pages flow through a PSP; but unlike data packets, printed pages must be routed through a series of irreversible physical transformations. This wide range of transport mechanisms and transformations (such as raster image processors, high speed digital presses, paper handling, and finishing devices) are all connected by their opportunity to read and add information to the printed page. Given the proper infrastructure, the printed page itself can dynamically route itself through the PSP. Through its design of its protocols, Gutenberg-Landa TCP/IP will unify Reliable Printing, Finishing, and Delivery into a single paradigm that would provide the same quality of service as the Internet: a highly efficient and robust print infrastructure with near 100% uptime and the ability to grow its print capacity without limit.

Introduction: "Printed Pages as Packets"

This paper proposes a next-generation reliable print service called the Gutenberg-Landa TCP/IP that uses the TCP/IP stack [1], the basis of Internet communication, as a model of layered protocols for design of PSPs. By using the same stack architecture, we will recreate three key system characteristics of the Internet for the next-generation print infrastructure: 1) distributed management of heterogeneous devices, 2) robustness and scalability of print capacity, 3) the seamless flow of content from Internet publishing to the printed page, and 4) the seamless flow of the physical manufacturing capabilities of the PSP onto the Internet. Just as the Internet consists of routers, computers, and optical transmission lines of different makes and models that run multiple versions of different software, so Gutenberg-Landa TCP/IP will bind together binders, staplers, RIPs (Raster Image Processors), digital presses and paper path into a cohesive print service. Just as the Internet can bypass failed components and find more bandwidth from connected components, so Gutenberg-Landa TCP/IP will add a level of robustness and scalability to the print capacity of a PSP. Just as Internet has developed services and applications that distribute text, video, and images across multiple computers, so Gutenberg-Landa TCP/IP will seamlessly transform distribution of Internet-created content to the printed page, creating high value pages destined for digital presses.

With Gutenberg-Landa TCP/IP, this paper implicitly proposes a paradigm shift in how PSPs are both internally and externally connected, and demonstrate how reliable print production can occur not only for a single PSP site, but also across a distribution



Figure 1. Two for one: Benny Landa (founder of HP/Indigo, pioneer and champion of Digital Print) as Johannes Gutenberg (Father of Printing). Courtesy Benny Landa.

of PSPs, independent of the underlying print capabilities or finishing hardware. The printing industry will always be an industrial challenge, because the transformation of data into physical print artifacts will always involve heavy machinery and large capital costs. Gutenberg-Landa TCP/IP is the first step in not only integrating industrial machines with IT infrastructure, but also in connecting digital content with its physical manifestations.

Background

Ever since its widespread adoption, analog offset printing has defined commercial print publishing [2]. In spite of Internet, desktop publishing, image processing, and Moore's law, the limitations of analog offset press have defined what it means to be published and what it means to be printed. The analog offset press can deliver a high quality prints for an unbeatably low price per page, but only at page volumes in the hundred thousands, and at a great capital expense. This combination of economics and the mechanics of analog offset press have limited commercial print to a broadcast medium, competing with television, radio and other forms of broadcast. Even with the advent of the digital offset press, the legacy of the analog offset press still defines many different aspects of the print infrastructure beyond the press itself: PSP Resource Management, Customer Access, Computation Architecture, and Color Processing.

PSP Resource Management

Currently, a centralized manager at a PSP provisions the equipment for a run of jobs for a particular class of content. While pre-allocation of equipment and centralized scheduling of jobs is efficient for long run jobs for the analog press, this methodology can only be modified to accept certain classes of short runs jobs, either by explicitly printing only one type of content, e.g. Snapfish with photobooks, or by expressing its content in a variable data

printing (VDP) language, whose language structure automatically filters out content that cannot match the print capabilities of the printing system. Even though the digital press itself can print every page differently, the current commercial print infrastructure does not exploit its flexibility as a general purpose printing component because its resource management has already predetermined what is acceptable content to print.

Customer Access

Currently, PSPs restrict access to their presses, RIP systems and finishing equipment to in-house professionals who have intimate knowledge of each machine. This access to analog offset presses was acceptable for offset presses because only long-run static print jobs were available and economically feasible. The cost of this face-to-face negotiation over each device would be amortized over the number of copies. This lack of automatic access hides the PSP capabilities from the customer and stifles the free flow of internet content into their production lines, especially for the short run internet-based content, *e.g.* VDP documents that are sourced from the web, *e.g.* housing market service mashup via Google maps.

Computational Architecture

The bulk of all computation in a PSP occurs at the RIP in order to create the imaging plate. This RIP time/computation is amortized over the number of copies being made with that plate. Because a typical press run of 10,000 to 100,000 copies take hours to complete, analog offset press have allowed RIP times of minutes per page, requiring lightweight IT infrastructure and precluding other digital content that may demand one RIP per copy.

Color Processing

The transition from analog to digital offset presses has been hardest on color processing. Although digital presses have the capabilities for new color capabilities, the language for commercial print color processing is so geared toward controlling the standard SWOP CMYK inks and spot colors on the analog offset press that it filters out most of new functionalities that digital presses have to offer. Most content providers hope for only consistent CMYK color on their prints, rarely aware of possibilities of anything more.

Job/Page Tracking and Error Recovery

In general, for job/page tracking and print error recover of long run jobs, PSPs use centuries-old solutions to these tasks: manual validation by a human, and overproduction, respectively. Since long runs were merely copies of each other, a PSP cover any copies that were lost with a standard 10%. For short run and VDP jobs, very few copies of a single page are made and overproduction would require 50-100% overhead cost on a job. Furthermore, the tracking of even 200 pages within a PSP would be nightmare if done manually, and, without a means to automatically reprint a job, the PSP floor would come to a standstill. For bill printing and other financial transactions, PSP must dedicate its own proprietary tracking and validation mechanism. Nevertheless, the problem of a multitude of small jobs flowing through a PSP and the long list of VDP need to be printed are the same. We require transparency not only on a job-to-job basis, but also on a page-to-page basis, much like how we do things on the Internet.

To summarize, although we may upgrade the commercial press from an analog to a digital, the lack of connectivity and the centralized management of the PSP's equipment is holding back the conversion from analog to digital commercial print content. We are at the tipping point of a paradigm shift on what material can be published, and how much more immediate and valuable the print media can become. Analog offset presses were broadcast communication devices; digital offset presses are more like the Internet. With the analog offset presses, commercial print publishing is a communication network with one high quality message sent to many people. With the digital offset press, commercial print publishing can print a different page to every person. In this way, the analog press is more like a television or radio tower, broadcasting a signal; while the digital press is structured more like the Internet, treating each page like a datagram packet sent to a particular user. Once again, the communication capabilities of the digital press are hidden behind a wall of protocols that have been defined for the analog offset press. Once the digital presses of PSPs can network properly to handle the streams of digital content, the commercial print industry (like other media industries) will reap the benefits of the Internet revolution.


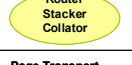
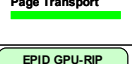
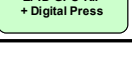
The Challenge: Gutenberg-Landa TCP/IP

The core theoretical challenge is how to remap and expand the TCP/IP standard into a print infrastructure that 1) has the physical routing of information, 2) integrates its own distributed imaging/sensing/marking network and 3) requires a guarantee that a sequential order of physical processing steps, determined by information on the page itself, has been applied to a given page. If we meet this challenge, we would create a print infrastructure that would provide the same quality of service as the Internet: a highly efficient and robust print infrastructure with near 100% uptime and the ability to grow its print capacity without limit.

By directly remapping the parts of the standard TCP/IP protocol, Gutenberg-Landa TCP/IP retains two specific characteristics of the TCP/IP layered system: 1) the ability to accept many different heterogeneous components, and 2) the ability to dynamically provision and route pages for best quality of service, even in the face of component failures. However, protocols for TCP/IP are written for data, not for physical objects; while data is inherently pliable, the naive mapping of TCP/IP would leave thousands of torn envelopes on the PSP's floor. This design would require distributed imaging/sensing network to validate a content, flow and quality on a page-by-page basis and match the extremely low false negative rates of packet error detection on the Internet. Using the same TCP/IP stack as model in both physical and digital domains, Gutenberg-Landa TCP/IP routes information and content seamlessly between the two domains.

Gutenberg-Landa TCP/IP must extend the TCP/IP standard to support the routing and discovery of different physical and the associated infrastructure that is required to bootstrap and implement these protocols. First, we must append the sequence and the parameters of physical transformations (*e.g.*, color, security, collation and folding) to the routing protocol of each page, so that the page can keep track of its own progression. Second, while the packet on the Internet must only make sure that it is approaching to

a)

| Symbol | Who Builds it? | G-L TCP/IP Capabilities | Examples |
|-----------------------------------------------------------------------------------|-------------------------------------------------------------------------|-------------------------|----------------------------------------------------------------------------------------|
|  | 3rd Party Finishing and Processing w/ HP G-L TCP/IP Add-on Capabilities | Read/Write Capable | Folder, Binder, Varnisher, UV Coater, Stapler |
|  | HP or 3rd Party Devices Certified to be HP G-L TCP/IP Compliant | Read Capable | Content Specific stacker, Variable Paper path w/ barcode reader, Page Buffer/collector |
|  | Anybody | None | Straight paper path, forklift, dolly, human being |
|  | 3rd Party RIP Front End, HP Press and GPU Back end | Primary Creator | Adobe + GPU-RIP + Indigo Press, GG + GPU-RIP + IHSP, etc. |

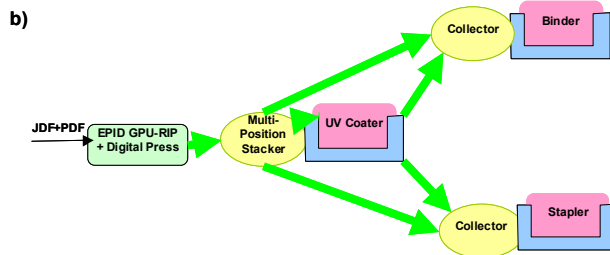


Figure 2. Next-generation PSP as Gutenberg-Landa network. a) illustrates a taxonomy of devices in a Gutenberg-Landa System. b) shows an example of a simple Gutenberg-Landa TCP/IP system that can bind or staple groups of pages with and without UV coating.

its final destination, a page in a PSP must progress through a series of particular physical devices before reaching its final destination.

This proposal envisions the next-generation PSP as a network, consisting of three types of Gutenberg-Landa components (Fig. 2a), connected by point to point paper transfers (Fig. 2b). Within the network, a printed page starts at the EPID (*Every Page Is Different*) RIP and press combinations, and moves through the PSP, requesting color and security services, discovering finishing devices and validating its own content and content transformation after each device. Along with enhancing finish devices with read/write capabilities, Gutenberg-Landa TCP/IP requires the creation of printed page routers, buffers, and switches, *i.e.*, a printed page version of network infrastructure.

Finally, as the Grand Challenge, how general is Gutenberg-Landa TCP/IP for creating a general purpose on-demand content-based manufacturing systems – can it be applied to car, house, or DNA-specific drug manufacturing [3]?

Gutenberg-Landa TCP/IP Stack

As shown in figure 3, we have divided our Gutenberg-Landa TCP/IP stack into a layered system that corresponds closely to the TCP/IP protocol: the Physical Layer, the Datalink Layer, the Network Layer and the Transport Layer. To distinguish from the Internet layers, we denote the Gutenberg-Landa versions of layers with G-L prefix.

Gutenberg-Landa Physical Layer

G-L Physical Layer encodes data onto and scans data off of the physical printed pages. This layer handles the actual physical paper and integrates digital presses and various finishing devices with sensors, imagers and special marking devices. This layer also includes printing specific hardware and algorithms to transform content into printing marks such as RIP, half-toning, color mapping, and security printing. This layer is responsible for

attachable imaging and mechanical components that add these reading and writing capabilities to current paper handling infrastructure. This layer is also responsible for creating the marks on the printed page for content payload, but not format or organization of the marks for routing.

One of the major challenges is the speed of converting data into printed pages: for Gutenberg-Landa TCP/IP, the transitional boundary between the Internet packets and physical pages must not be a bottleneck since all pages are different due to their identity and routing information. The digital press can certainly deliver the ink at speed; however, a great challenge is to design IT infrastructure that can deal with the real-time constraints of digital press' output and the later digitization of its output. Since a digital offset press can print a different page every time, RIP may occur for every page, increasing computational demand by a factor of 100,000 over analog offset press. This massively parallel high-performance RIP computing architecture [4] will integrate multi-core CPUs, GPUs, FPGAs and custom-designed ASICs, as needed.

The other challenge is to provide a distributed imaging/sensing network that can answer four questions with a very, very low error rate (<0.001%): 1) routing verification, *i.e.* whether it is supposed to be here, 2) content validation, *i.e.* whether the current physical page actually matches its corresponding digital page, 3) job validation, *i.e.* whether the page placed correctly in its bound document or print job, and 4) G-L TCP/IP marking, *i.e.* how to append more routing information for the G-L TCP/IP onto the page. The success of the distributed imaging/sensing network depends upon a design that optimizes five different factors simultaneously: 1) the robustness and informational density of the page headers, 2) the complexity of the imaging hardware, 3) the complexity of image enhancement, registration and segmentation, 4) the placement within or modifications of the paper path for these imaging devices, and 5) integration of the marking technologies, if the device is write-capable. Much like EPID RIP, a similar computing infrastructure is required to handle the uptake of this data as well.

Gutenberg-Landa Datalink Layer

G-L Datalink Layer binds information to the printed page, reads information off the page, validates its content payload and amends the routing information, if necessary. G-L Datalink Layer specifies routing information to be created on the page. This layer must balance the robustness of routing indicia with the complexity of its imaging algorithms. In addition to supporting routing information, G-L Datalink Layer also define a language of content transformation to encode finishing, color and security services that the page must reach, and update that information when that processing has been applied.

Gutenberg-Landa Network Layer

Network Layer routes pages within the PSP to the page's desired transformations. This layer is responsible for the protocols to robustly find the required processing resource and move the page toward its processing goal via the most efficient form of physical transport. To match Internet protocols, we would assume some locally optimal greedy selection routing for distributed routing. Unlike the Internet, Gutenberg-Landa TCP/IP has classes of physical device with different functionalities and some pages

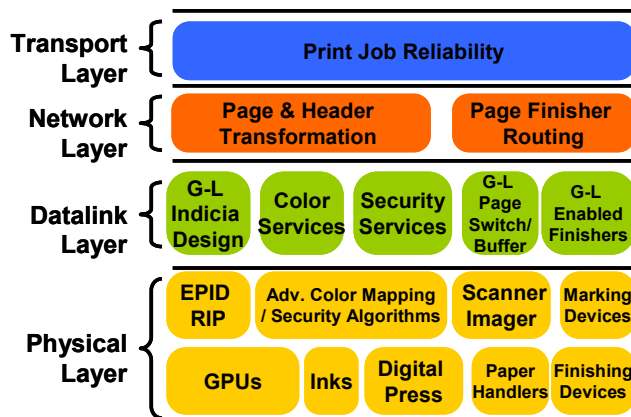


Figure 3. Gutenberg-Landa TCP/IP Stack Implementation.

will require routing to specific classes of physical devices. One approach would be to treat every class of physical devices as its own virtual overlaid network inside of the PSP network. This multiple network view of the system would require substantial analysis and design to make sure that pages moving from one device to another would still have some guarantee of progress.

Gutenberg-Landa Transport Layer

G-L Transport Layer (TCP) maintains the correctness and reliability of the printed job. This layer will rely on page image analogue to data checksumming, but the protocols for job reliability will be determined mostly by the capabilities of the lower layers. This layer's function is to ensure the reliable print of a whole job. The standard TCP protocol is that if packet is missing, all packets are dropped and a request for resend is made. The TCP protocol may have to be significantly modified in its failure modes to work for PSPs. For instance, G-L TCP may have to encode some kind of path robustness for it to work properly. The previous layers of design will also inform how we design the analogue of the TCP protocol in the printed page domain.

The design and implementation of a modified TCP/IP routing protocol must include an infrastructure to discover the correct physical devices and the protocol for marking the progression of page through its physical transformations. For Gutenberg-Landa TCP/IP, every class of physical device defines a different connectivity to printed page network. With these differentiated networks, we need to analyze the stability of these protocols and the performance guarantees. As we better understand the system behavior of the component involved with the design, we can better define the protocol and supporting infrastructure.

Print Services Above Gutenberg-Landa TCP/IP

In consuming jobs for the customer, PSP would control their internal PSP network through the submission of JDF-specified jobs on the network. The PSP would expand (or reduce) production capacity by adding (or removing) new device to the Gutenberg-Landa Network. Jobs flowing through the PSP could be tracked; jobs finished would flow out or fail within the PSP. Ideally, while the physical devices would still be individually installed, maintained and repaired as today's PSP, the management of the PSP could be understood and maintained easily at a system level.

Conclusion

The insights that we gain from this remapping of TCP/IP will inform us about the next untapped market: General Purpose Content-Driven Manufacturing, where design and content of the consumer drives the production process. Commercial print is merely the leading edge example of this new type of manufacturing.

Our first step at HP Labs is to build PSP Simulator [5] for cost-benefit analysis of the current PSP designs vs. Gutenberg-Landa based PSP design. We will optimize our system design and Gutenberg-Landa protocols according to the patterns of communications among system components and the impact on associated design tasks, for instance, imaging and indicia design. As our analysis design progresses, we will work with PSP to test our designs.

References

- [1] Comer, D.E., *Internetworking with TCP/IP*, Vol. 1, 5th Ed., Prentice Hall, 2006.
- [2] Kipphan, H. (Ed.), *Handbook of Print Media: Technologies and Production Methods*, Springer, 2001.
- [3] Shastry, B.S., "Pharmacogenetics and the concept of individualized medicine", *The Pharmacogenomics Journal (Nature)*, 6, 16–21, 2006.
- [4] Lin, I-J., Buzbee, B., Moore, K. and Dispoto, G., "GPU/CPU codesign of an accelerated Raster Image Processor", *HP TechCon 2008*.
- [5] Zeng, J., Lin, I-J., Hoarau, E. and Dispoto, G., "Numerical simulation and analysis of commercial print production systems", *NIP25*, 2009.

Author Biography

I-Jong Lin is the research manager for GPU-RIP Print Services group with Hewlett-Packard Laboratories. He received Ph.D. and M.S. in Electrical Engineering from Princeton University and Stanford University, respectively, and B.S. in Computer Systems Engineering from Stanford University. His research experiences span from VLSI-CAD, digital design, neural networks, video object extraction, to digital print. He is the author of "Video Object Extraction and Representation: Theory and Applications" by Kluwer Academic Publishers. I-Jong holds 13 patents.

Jun Zeng is a senior scientist with Hewlett-Packard Laboratories. PhD (mechanical engineering) and MS (computer science) from Johns Hopkins University, and BS (modern mechanics) from USTC, China. Jun's publication includes 30 peer-reviewed papers and a co-edited book on CAD. He was a guest editor of IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems. Jun also serves as termed faculty and member of PhD dissertation committee with Duke University's Electrical and Computer Engineering Department.

Eric Hoarau is currently a senior researcher at Hewlett-Packard Laboratories in Palo Alto, CA. He received his B.S. from the University of California at Berkeley and his M.S. in Mechanical Engineering from the Massachusetts Institute of Technology. His research interests span several fields: Mechatronic systems, imaging systems, color imaging algorithms, distributed computing and system dynamics.

Gary Dispoto is the director of the Print Production Automation Lab, which seeks to streamline the processes required to produce industrial and commercial digital print, enabling new types of printed products and broader access to commercial print production. He holds several U.S. patents related to color imaging. Gary received B.S. and M.S. degrees in electrical engineering from Stanford University and an MBA degree from the University of Santa Clara.