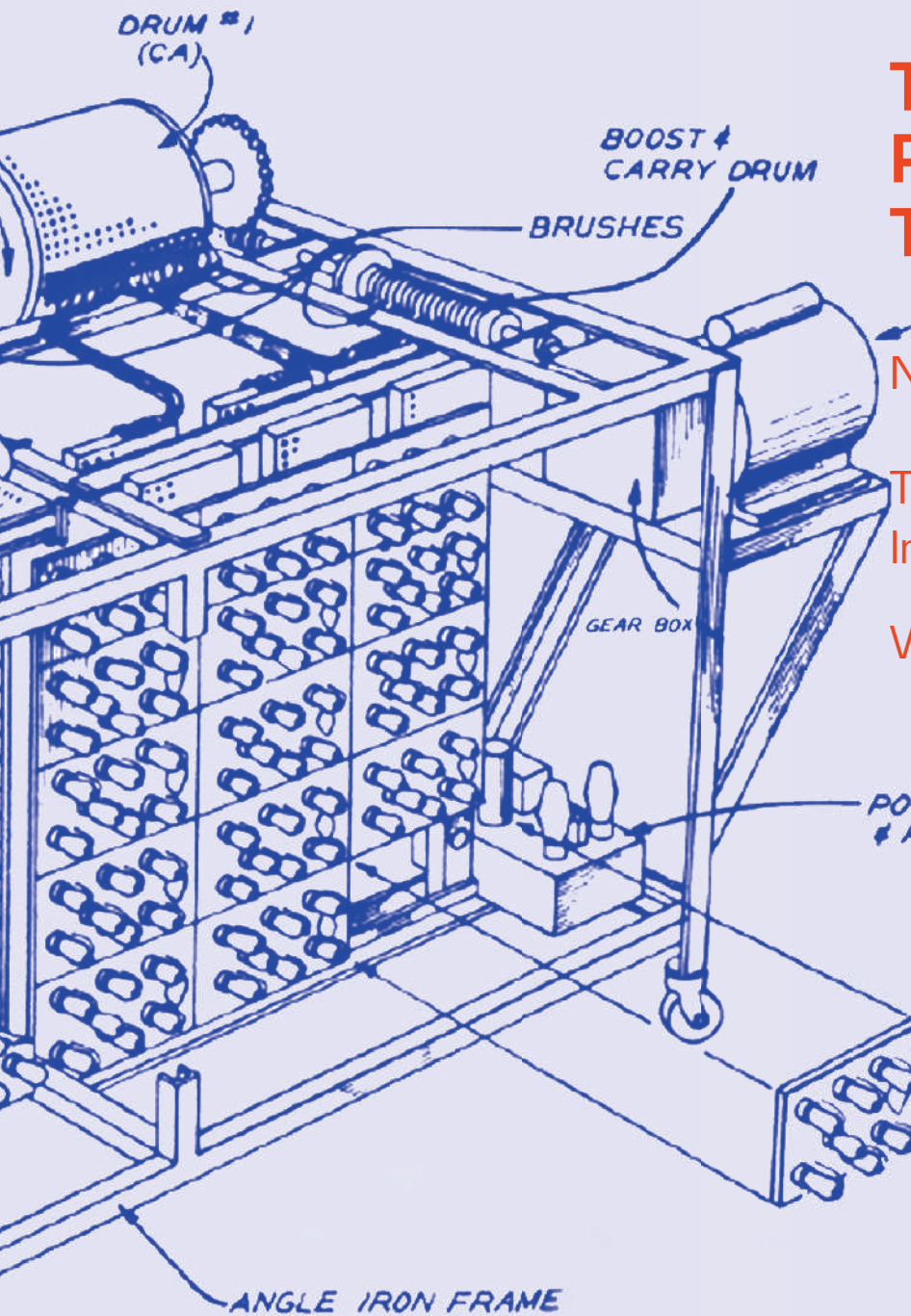


The Next Wave

The National Security Agency's Review of Emerging Technologies
Vol 17 No 3 • 2008



The Myth of Revolutionary Technologies

NetTop Eight Years Later

Text Extraction from Color Images

Web 3.0 — “Are we there yet?”



READER

Letter from the Editor

How does a new technology—hardware or software—make it onto the shelves at your local WalMart or Best Buy? Where does it start? Who first thought of it? Who first created it? Some commercial technology starts right here at the National Security Agency (NSA).

When I began working at NSA two years ago, I was thrilled to learn about some of the amazing technology that happens here. As I settle into my job as Managing Editor of TNW, the thrills remain as I continue to learn about new technology, new software, and new research. Much of the amazing technology is classified, of course. But a surprising amount is not.

Of the unclassified technology, a small percentage is patented and can then be licensed for use by the commercial sector. This issue contains articles about several patented technologies now commercially available from NSA. Additionally, we have a thought-provoking commentary on the origins of technology from a senior scientist at the Defense Intelligence Agency (DIA).

I hope you enjoy your peek behind the curtain and find learning about NSA's technologies just as thrilling as I have.

This cover of TNW shows a drawing of the Atanasoff-Berry Computer. The image, courtesy of Iowa State University, can be found at <http://www.cs.iastate.edu/jva/images/abc-artists-concept.gif>

About the Atanasoff-Berry Computer: John Vincent Atanasoff, an electrical engineering professor at Iowa State College (now Iowa State University), along with graduate assistant Clifford E. Berry built the Atanasoff-Berry Computer (ABC) in 1937-1942. It wasn't until 1973 that the ABC was officially recognized as the first electronic digital computer, when Federal Judge Earl R. Larson ruled that patents for the better-known ENIAC were invalid. A full-scale replica of the ABC is on display in Durham Hall on the Iowa State University campus.

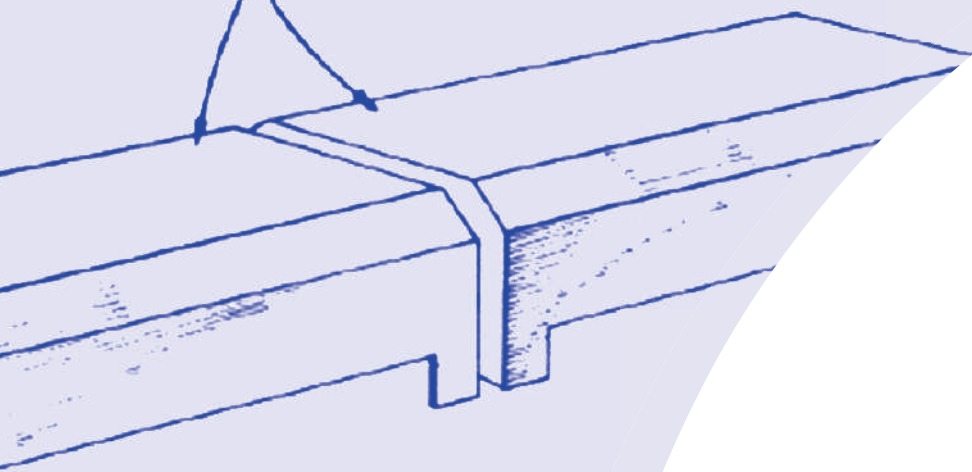
For more information about the Atanasoff-Berry Computer, read *Atanasoff, Forgotten Father of the Computer*, by Clark R. Mollenhoff.

The Next Wave is published to disseminate significant technical advancements and research activities in telecommunications and information technologies. Mentions of company names or commercial products do not imply endorsement by the U.S. Government. Articles present views of the authors and not necessarily those of NSA or TNW staff.



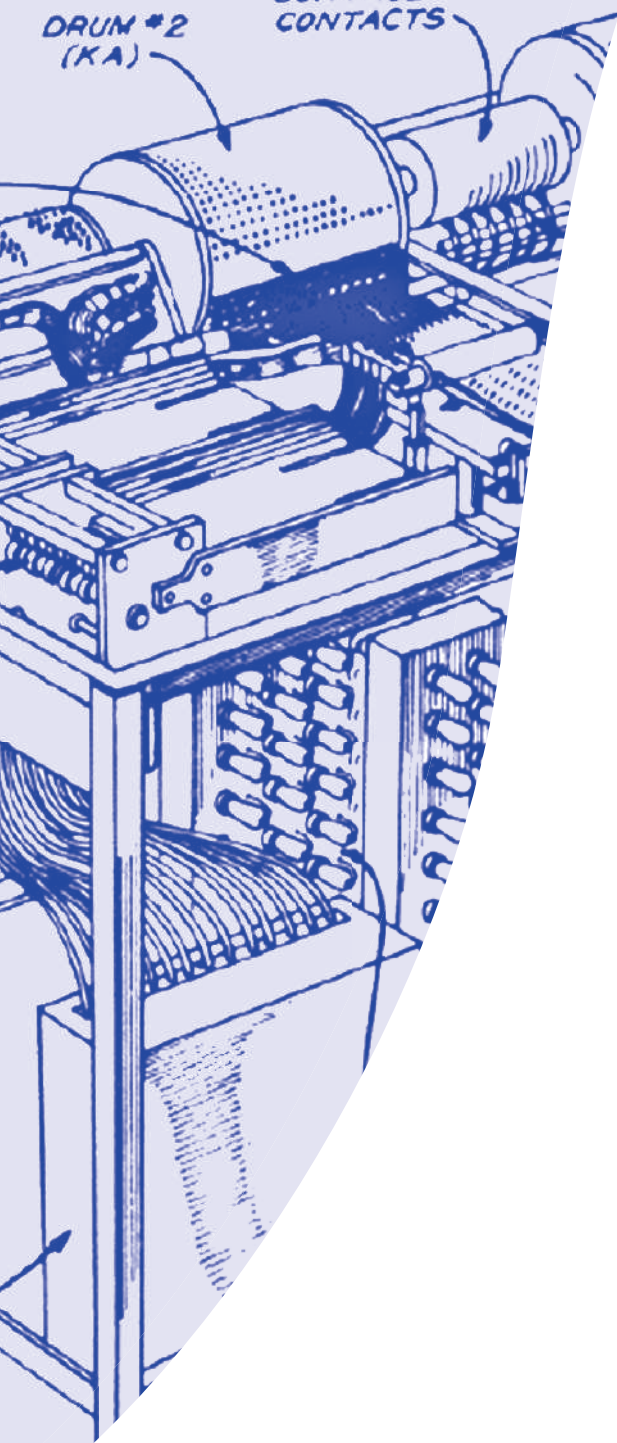
For more information, please contact us at
TNW@tycho.ncsc.mil

DRUM COVERS



TIMING
CONTROL
CONTACTS

DRUM #2
(KA)



CONTENTS

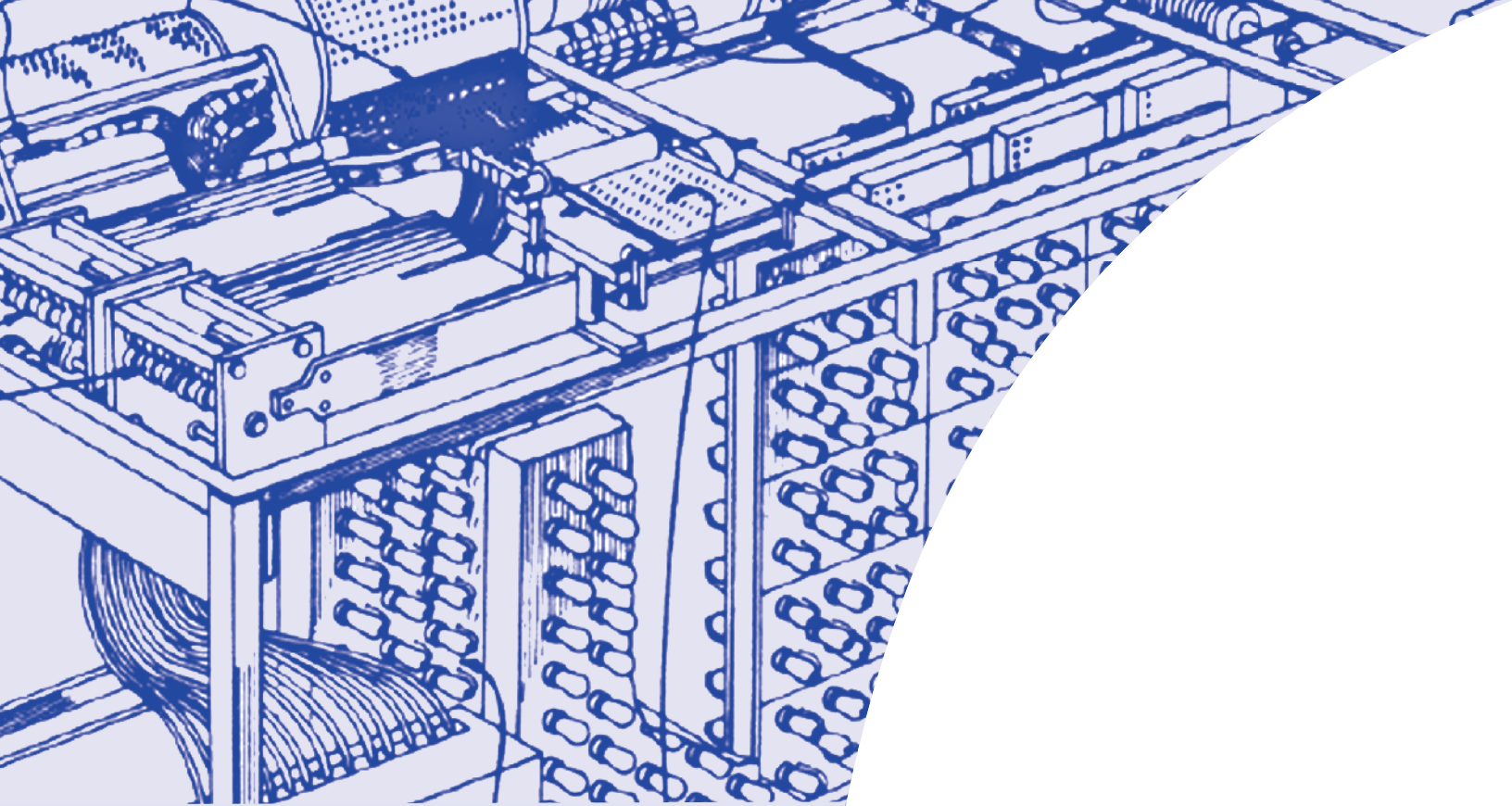
FEATURES

- 4 The Myth of Revolutionary Technologies
- 10 NetTop Eight Years Later
- 22 Text Extraction from Color Images

FOCUS

- 26 Web 3.0 – “Are we there yet?”





The Myth of Revolutionary Technologies

If you build it they will come, eventually!

Compared with past populations, we can consider ourselves the most fortunate humans to have ever lived. People in developed nations today experience the inverse of Malthusian subsistence. With nearly all physical needs met relatively easily, we have devoted enormous amounts of time and effort to intellectual pursuits and leisure. We can attribute this existence for the most part to the amazing evolution of science and technology and its ready application to human needs. Society today is the beneficiary of an evolution of science and technology stretching as far back into history as records of discovery exist. From fire to file servers, surfing on an ever advancing wave of technology, we have ridden the shoulders and often the backs of many generations of inventive giants and fertile

minds that have come before us.

Advancing at an exponential pace, we have come to expect revolutionary technology to appear at any moment out of intellectual ether. In the span of fifty years we have leapt from the Atomic Age, to the Space Age, to the Information Age, and now verge upon the Biological Age, an Age of willful genetic manipulation and affordable designer genes. To most it might seem that the technological advances which brought about these revolutionary Ages immediately resulted from revolutionary discovery.

This misperception persists even in the face of vast amounts of evidence to the contrary.

Revolutionary scientific discoveries and technical advances foreshadowed the Ages of the 20th century. But, without

exception, each discovery and advance was nurtured by numerous parallel and often unrelated societal and technological developments. Ultimately, limitations of existing technologies and their inability to meet a pressing need drove these advances. At the same time, new technologies had to fulfill these societal and economic imperatives. Essentially, to become revolutionary, technologies had to emerge from inventive minds, and then lay in wait for the proper niche, fertile environment, and an imperative for change. Today's technologies developed by what paleobiologists have come to call punctuated evolution.

Punctuated evolution, or "the theory of punctuated equilibrium in evolutionary processes," explains the mechanism by which genetic mutations within a species



Photo of Trimetric view of modern jet engine (GE90-115B)
Image courtesy of General Electric Company, ©2009

come to rapidly dominate the entire genus. It also explains how one species becomes dominant, replacing the older and previously more prevalent form of life. Punctuated evolution holds that the biosphere consists of a pool of mutations that constantly emerge to test the environmental waters. If a mutation loses viability, it dies out. If it remains viable, it may survive for a time. But if a confluence of mutations, environmental change, the appearance of a unique niche, or the loss of an old niche occurs, the new mutations will rapidly replace the current dominant species or genus and expand to dominate the biosphere until another punctuation event occurs. In other words, had earth's environment stayed hospitable for dinosaurs or even had it changed at a geologically slower pace, they'd still be here and we would not. Though, most likely, a polyglot of differentiated mice and mole species would exist as dino-food. Similarly, for a technology to become a revolution, a confluence of technological concepts, environments, and events must exist that bring it to the forefront, rendering the older capability obsolete.

To demonstrate the confluence of conditions that must occur to result in technological revolution, let us trace the development of five technologies that came of age in the 20th century. Each of these technological revolutions emerged from an evolutionary process that in one

case stretches back to the last ice age. By any measure, the gas turbine engine, nuclear energy sources, the laser, digital electronic computing, and genetic design and manipulation have caused a revolution in the way we travel, conduct war, communicate, and ultimately view and manipulate ourselves and our environment.

Jet Engines

Sir Frank Whittle in England and Hans von Ohain in Germany independently developed and introduced the gas turbine engine to aviation in 1939. Ohain developed the world's first jet powered airplane when he fit his engine to an airplane, the Heinkel He 178, in August 1939. The British Gloster E28/39 followed the Heinkel in 1941. The world's first commercial jet airliner, the De Havilland Comet, flew ten years later in July 1949. But nearly another two decades passed before intercontinental propeller-driven air service departed the commercial aviation scene.

We can trace the entire history of turbine powered aviation to well before Frank Whittle, attributing the first working gas turbine to Hero of Alexandria in the 1st century AD. Little more than a steam-driven sphere, it remained unready for adaptation to supersonic flight. Issued in 1791, the first patent for a turbine engine emerged too early for the nascent aviation industry. Montgolfier balloons and jet engines were not a proper mix! And, while

the first operating gas turbine engine was demonstrated in 1903, Orville and Wilbur opted for the more conventional and reliable piston-driven internal combustion engine attached to a propeller to provide thrust at Kill Devil Hills.

But in 1939 an evolutionary punctuation occurred with the confluence of aluminum airframes; precision engineering technologies; and inherent speed, altitude, thrust, and complexity limitations of the piston engine for aviation. These factors combined with the imperative of global warfare to drive the turbine engine to the aviation forefront. In the postwar world, the market for high-altitude, high-speed commercial flight above the troposphere's weather, made commercial jet flight the dominant means of global transportation and gave the general population revolutionary access to the world.

Nuclear Power

On December 2, 1942, in Chicago, Enrico Fermi and his team of scientists and engineers achieved the first self-sustaining nuclear chain reaction. Two-and-a-half years and several billions of dollars later, on July 16, 1945, the world's first intentional runaway chain reaction took place on the desert of New Mexico. In several microseconds, that chain reaction yielded a detonation energy equivalent to approximately 38 million pounds of high explosive. The concept of global war had changed forever. Nuclear power, the E in $E=mc^2$, had in a matter of three years come under human control, ushering in a revolutionary era in the concept of warfare and power generation.

But in 1945, what appeared to take place almost overnight had in reality taken 50 years of research and development. In 1896, the theory of the atom, which dates back to ancient Greece, still remained in question. That year, Henri Becquerel, while studying phosphorescence, discovered spontaneous radiation in uranium salts. Spontaneous emission of particles from certain minerals indicated that a new source of energy was somehow entrapped within these radioactive materials. Thirteen years later, Earnest Rutherford and Hans Geiger proved that atoms consisted

of a positively charged nucleus, which in the case of a radioactive material spontaneously ejected positively charged alpha particles, negatively charged beta particles, and neutral gamma rays. The emission of alpha particles transmuted the original element into an entirely different element. This transmutation resulted in a net gain of energy and loss of mass. Thus, in terms of mass the sum of the parts equaled less than the whole.

In 1905, a young examiner in the Zurich patent office proposed two revolutionary scientific theories. One theory suggested that space and time were not absolute quantities to any observer, but that the quadratic sum of these four dimensions (three space dimensions plus time) remained constant to all observers. Matter (m) and energy (E) became interchangeable and the constant of proportionality between these two quantities evolved as c^2 , the square of the speed of light. This proportionality is mathematically represented by the familiar $E=mc^2$. When applied to radioactivity, Einstein's theory uncovered that the huge energy gained in radiation equaled the nearly imperceptible missing mass. Enormous amounts of energy lay available and untapped in the atomic nucleus. Einstein's second revolutionary theory that year held that matter interacted with radiant energy (light) as if the light came in discreet bundles (quanta),

and that matter would only absorb or emit light of specific energies. This "photoelectric" theory would win Einstein the 1922 Nobel and foreshadow the development of the laser forty years later.

One key particle remained missing from the nuclear energy equation. In 1932, James Chadwick identified the stealthy charge-free neutron as a component of the atomic nucleus. With nearly identical mass to the proton, the neutron became identified as the missing key to nuclear fission. Electrically neutral and yet massive, it served as an energetic probe that could penetrate deeply into the atomic nucleus to investigate its properties and ultimately split the atom.

Nuclear fission of a heavy element, with a resulting emission of lighter nuclei and slow neutrons, became the key to harnessing atomic energy. In 1934, Fermi experimented on the neutron bombardment of uranium but either failed to detect or failed to report on fission. So in 1939, Otto Hahn and Fritz Strassmann, working in Nazi Germany, reported to Lise Meitner, a German émigré physicist living in Sweden, on the production of a barium isotope from the neutron bombardment of uranium. Meitner recognized this as the fissioning of the uranium nucleus. Based on this work, Fermi predicted that he could create and sustain a chain reaction in uranium enriched in the isotope U^{235} . With

the onset of European war in 1939, Fermi moved from Italy to the United States and continued his work at Columbia University in New York and at the University of Chicago.

On Saturday December 6, 1941, Vannevar Bush, science advisor to Franklin Delano Roosevelt, convened a meeting of preeminent US nuclear scientists at the Cosmos Club in Washington, DC. He aimed to focus the structure and direction of nuclear research in the United States on development of an atomic fission bomb. Forty-five years of basic research and development had established the scientific basis for such a device. A letter from Albert Einstein to President Roosevelt primed the fiscal environment. The punctuation to this scientific evolution came the next day at 07:48 Pacific time in Hawaii.

The Laser

In 1959, Gordon Gould filed a patent application for a device he called the laser. Named for its mechanism of operation, light amplification by the stimulated emission of radiation now commonly seen as "laser"—his patent suggested he could use the laser for spectrometry, interferometry, radar, and nuclear fusion. Gould failed to predict the two most prolific applications of lasers today. Since fiber optic communication and high density data storage and retrieval had not yet come onstage, we can excuse his oversight. Yet, today, due to serendipitous and parallel technological developments, the speed of fiber optical cable laying approaches the speed of light and the Blu-ray DVD has grown into a multi-billion dollar industry.

Theodore Miaman, working at Hughes Labs, produced the world's first laser in 1960. An awkward device, it consisted of a ruby crystal pumped by white-light flash tubes. The lineage from that first device to blue laser diodes has taken many paths simultaneously. Government and industry have made enormous investments in laser development, and the paths of development have diverged wildly. Imagine a modern world without laser communications, manufacturing, scanning, measuring, and computing. Current military navi-

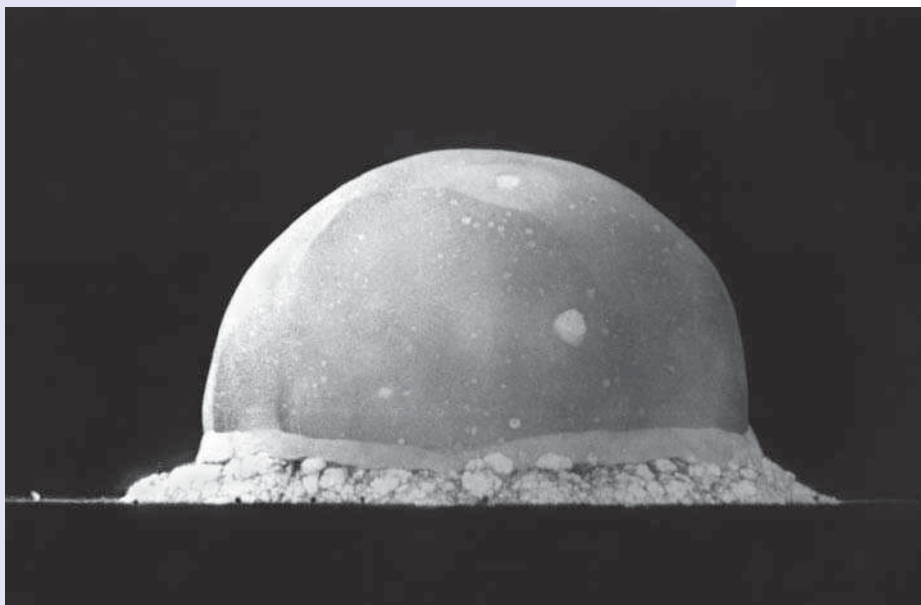


Photo of Trinity site explosion, .016 seconds after explosion, July 16, 1945

Photo credit: Los Alamos National Laboratory

gation, guidance, and tracking and targeting technologies largely depend on lasers. High-power chemical lasers have emerged to shoot down missiles in flight. And at the Department of Energy's National Ignition Facility (NIF) the hope of inertial confined nuclear fusion may eventually be realized. The investments in laser technology have paid off smartly and caused a revolution in business, industry, and warfare, though it first entered the lexicon of modern technology and language 47 years ago.

Einstein's 1905 and 1917 papers on the photoelectric effect and the quantum theory of radiation laid the theoretical foundation for the laser. But 43 years of research, development, and investment had to take place before a practical laser came to fruition.

In the 40 years following that first laser, parallel technologies were developed that could employ the new technology, making way for the current revolution in its use.

The Digital Electronic Computer

The use of a counting machine to aid tedious numerical calculation has roots to the abacus, which existed over a thousand years ago. Even the term "computer" originated in the 17th century when individuals named for their profession, *computed* logarithmic, ballistic, and astronomical tables for ships and guns of Britannia's Royal Navy. By the 18th century, analog mechanical devices for numerical calculations were in development to aid these computers. (Recent archeological evidence suggests that the ancient Greeks used such devices for astronomical calcu-

lations.) These calculating devices generally consisted of wheels and slides operated by cranks, thumb wheels, or push rods. Until the 1970s, one of these devices, the slide rule, dominated the scene, dangling from the belt of almost every engineering student worldwide.

In addition to one-time calculations, repetitive sub-routines, introduced in the early 19th century in the form of coded, wooden, card-like devices or paper rolls, automatically controlled repetitive tasks performed by manufacturing machinery and beer-hall musical instruments. By the early 20th century, electric motors joined the ever more complex wheels, slides, and keys to turn cranks faster and more continuously. They enabled the mechanical calculator to address difficult problems created by numerically solving complex differential equations. Specialized problems such as flight simulation prompted the genesis of electrical analog models.

In the later part of the 19th century, two technologies critical to the operation of a digital electronic computer dawned: the vacuum tube and Boolean algebra. But not until 1943 did Colossus, the first fully electronic, partially programmable, digital computer, begin operation. It was sited at Bletchley Park, England, and was designed to address the complex task of breaking German military codes. Electronic computers evolved rapidly after the war. By the 1950s, a census-taking company converted itself into International Business Machines, IBM, and came to dominate the computer field. But the vacuum tube kept the cost, size, power consump-

tion, and reliability of the computer out of reach for most users. The computing market thus stayed relatively small, with these factors combining to limit the use of electronic computers to governments and large corporations.

Electrical semi-conductive properties of solid-state materials became the key to making the computer available to a new generation of engineers, scientists, and businessmen. For decades, the unique properties of silicon and germanium crystals had been exploited to demodulate AM radio frequency signals into acoustic electric signals. Crystal sets, which preceded the vacuum radio, employed these crystals to drive an earphone and create sound. Though a patent for the point-contact crystal diode was filed in 1906, it wasn't until 1947 that Bardeen, Brattain, and Shockley at Bell Laboratories developed the first practical point-contact triode employable as a switch or amplifier—the *transistor*. Eight years later, the Tokyo Tsushin Kogyo Corporation paid a \$50,000 license fee to use the Bell Lab transistor design to build radios. In the same year, the Tokyo Tsushin Kogyo Corporation decided to use the name "Sony" to better market the new transistor radio to the West.

Besides amplification and rectification for radios, the new transistor could act as a switch, the fundamental device inherent in all computer components from memory to processors. This new switch could convert the computer from a room-sized, unreliable, power-consuming behemoth to a comfortable, wall-cabinet sized device

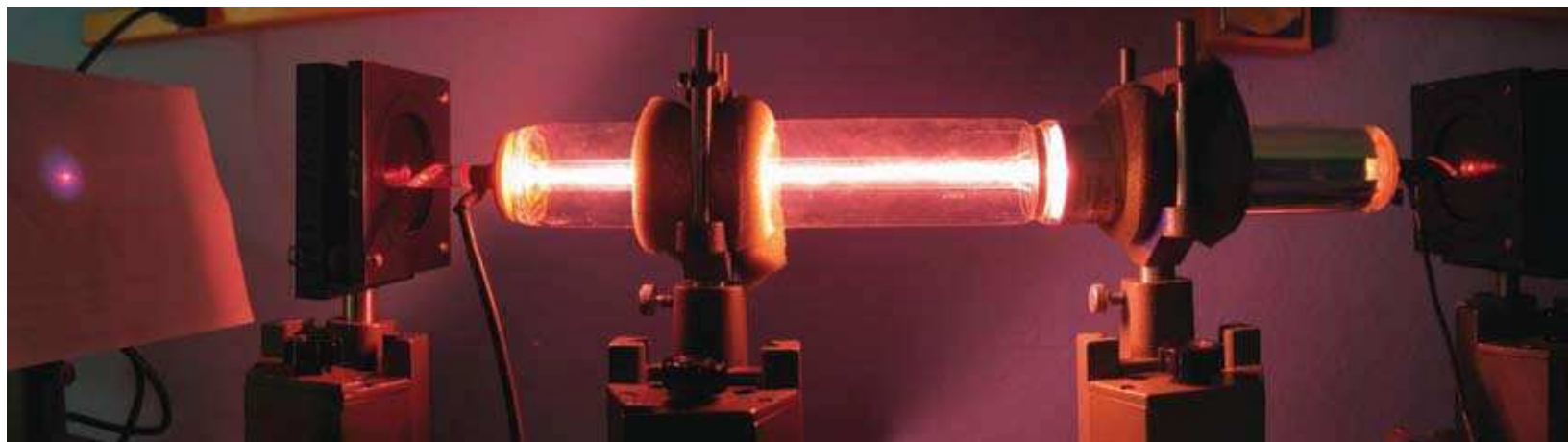


Photo of helium neon laser.

Photo courtesy of NSF - Center for Biophotonics

that both universities and small businesses could afford. In 1957, Olsen and Anderson formed the Digital Equipment Corporation (DEC) and by 1964 had produced the first practical solid-state “minicomputer,” the programmable digital processor (PDP-8). The PDP-8 sold for \$16,000, equaling \$200,000 today.

Despite the revolution in computers instigated by the transistor, the combination of two old and one new technology into an application actually started the personal computer revolution. Photolithography, acidic etching, and vapor deposition allowed for the design of integrated electronic circuits with hundreds to millions of transistors. These circuits have the capability of performing billions of operations per second on silicon wafers no larger than a postage stamp. These integrated circuits constitute the guts of all digital electronic systems, replacing vacuum tubes, discrete transistor components, and magnetic core memory.

While minicomputers proliferated, the means by which humans could interact with them and they, in turn, could interact with one another paralleled the pace of hardware production. The development of one of the first interactive and inter-networked computer systems began at MIT’s Lincoln Laboratory in 1958. The Air Force’s Semi Automatic Ground Environment (SAGE) system consisted of redundant AN/FSQ-7 computers interconnected with distributed radars and operated from radar consoles with light pen designators. Operators guided fighter-interceptor aircraft to targets by designating the target with the light pen. The information up-linked to the airborne aircraft directed the intercept. Data communicated between remote radars and their local computers traveled digitally to the main SAGE system over standard telephone lines.

As SAGE operations began, an even more ambitious concept unfolded at MIT and subsequently at the Advanced Research Projects Agency (ARPA). This idea of a “galactic network” incorporated hundreds of interconnected computers around the nation. First conceived in 1962, a network of military computers became the topic of

the science fiction nuclear holocaust genre in 1966 in the novel *Colossus*, the first of this ilk. On January 14, 1969, researchers at Stanford Research Institute and UCLA established the first ARPANET link between computers. By 1983, newly created computer networking protocols allowed a wide variety of computers to interconnect and ARPANET became the initial backbone of the Internet. The ubiquitous use of the Internet and its components continued to develop in parallel with market driven priorities and an exponentially expanding set of global users.

The revolution in computation and communications, from cell phone to wireless internet, has reached its current state through a process of continuous evolution from basic technologies to advanced capabilities. None of today’s revolutionary computer/communications products emerged overnight. Slow progression of research enabled the rapid proliferation of technologies. In communications and computers, the ever increasing consumer demand and associated enormous profit potential serves as the most recent catalyst. These punctuations in a technological evolution began with the code-breaking applications of a world at war.

Genetic Manipulation

Genetic manipulation, which is at the core of the emerging Biological Age, is arguably also the oldest of the new technologies. Attempts to alter genetic traits may have started after the last ice age. The symbiosis that developed between human and canine hunters led to the selective breeding of wolves for hundreds of purposes. Over thousands of years of genetic modification, our carnivorous competitors came to be called “man’s best friend.”

For millennia, humans have known that animal temperament and the physical characteristics of plants and animals transfer by some means from one generation to another. By selective breeding for desired characteristics, we have empirically created all kinds of domestic animals and agriculture that would not have existed without human intervention.

Although it had little basis in

biochemistry, Gregor Mendel’s empirically derived seminal paper on plant hybridization in 1865 showed that inheritance properties had a scientific basis and could be described and manipulated with mathematical precision. This work was the foundation of modern genetic theory.

Born in 1877, Oswald Avery, a physician and molecular biologist, spent most of his career at the Rockefeller Institute in New York City. In 1918, Avery became a key researcher in uncovering the nature of the Great Influenza Pandemic that decimated populations globally. Avery was constantly on the verge of a breakthrough in the field of highly infectious bacterial and viral disease, but it was not until 1944 that he discovered DNA as the stuff of genes and chromosomes. Since genes and chromosomes serve as critical protagonists of heredity, as described by Mendel, it followed that the DNA molecule would somehow determine inherited characteristics.

In 1953, Francis Crick and James Watson identified the now familiar double helix and paired base structure of the DNA molecule. The unique structure and chemical makeup of the DNA molecule provided them with the clues to its function as life’s information storage device.

DNA is a long polymer comprised of simple units called nucleotides. Its backbone consists of sugars and phosphate groups joined by ester bonds. Each sugar in the backbone has one of four unique base molecules attached. The directions for constructing an organism are encoded in the sequence of these four bases arranged along the backbone. DNA is organized into structures called chromosomes, and the complete set of chromosomes within a cell make up a genome. DNA replication duplicates chromosomes before a cell divides. The genetic information from one generation of cells is thereby transferred to the next.

With this knowledge and the associated biochemical tools at hand, it became possible to map and to modify genomes. Thus genetically modifying an existing organism at the molecular level became a reality. This new capability revolutionized

genetics and enabled geneticists to change the characteristics of an organism in one generation rather than in hundreds.


By 1990, sufficient understanding of the critical role of DNA in human development, disease, and characteristics warranted an attempt by the government to begin the gargantuan task of mapping the entire human genome in a project named the Human Genome Project (HGP). The Department of Energy, which led the project and invested three billion dollars over 13 years, released its report on the human genome in 2003—50 years after the Watson and Crick discovery of the structure of DNA.

As we can see from these few simplified examples, revolutionary discoveries and technologies need time and trial before society embraces the revolution. I do not need or intend to detract from the tremendous scientific and technological contribution of 21st century researchers to appreciate that underlying evolution enables revolutionary technologies. Rather, I hope these examples convey a better understanding of the processes inherent in research. A scientific discovery or a technical breakthrough does not make a revolution. Chasms in time and technology generally exist between discovery and application. As evident from these five examples, fundamentally new discoveries disrupt the status quo. This is seldom popular. Therefore, revolutionary shifts will occur only when the technical and societal environment can accept the new paradigms. We cannot bypass the slow process of transitioning discovery to application. Final application so often depends on key interrelated factors that remain irresolvable at the time of discovery.

I hope these examples clearly illustrate that the original intent of research very often does not relate to its final application. Because of parallel yet interrelated technical advancements in diverse fields, or because of unforeseen societal developments, a technology created to effect change in one application will cross from one field of endeavor to revolutionize yet another.

We can seldom predict serendipitous results from research but we should always

expect them. Even in directed research, serendipity will often occur if the environment supports diverse investigations and interactions. Future technological advances that change our ways of doing business will come about only if we plant the seeds of innovation and nurture the crop of ideas. With this consideration, we must remember that failure to seed the future ensures that we will harvest the past. 🌱



NetTop Eight Years Later

In the summer of 1999, NSA's most senior Advisory Board issued a report warning of a serious and growing problem in the protection of government's most sensitive information systems. The Board's concerns acknowledged the dramatic decline in information assurance that many professionals had observed over the previous decade. Surprisingly, this decline occurred in spite of major advances in computer security spurred by the establishment of the National Computer Security Center (NCSC) in 1981 expressly for the purpose of securing critical information systems. Numerous high-assurance computing platforms were produced as a result of the NCSC's efforts, but none seemed capable of coping with the impact of the decade's technological phenomenon—the Internet. The 90s produced an explosion of new networking systems and services, and the widespread availability of powerful and inexpensive commodity workstations powered by Microsoft's ubiquitous Windows operating system. Not even the national security community could resist the functionality and cost savings this technology delivered, despite numerous assessments of its negative impact on security. Commercial-off-the-shelf (COTS) information technology had established a permanent foothold within government.

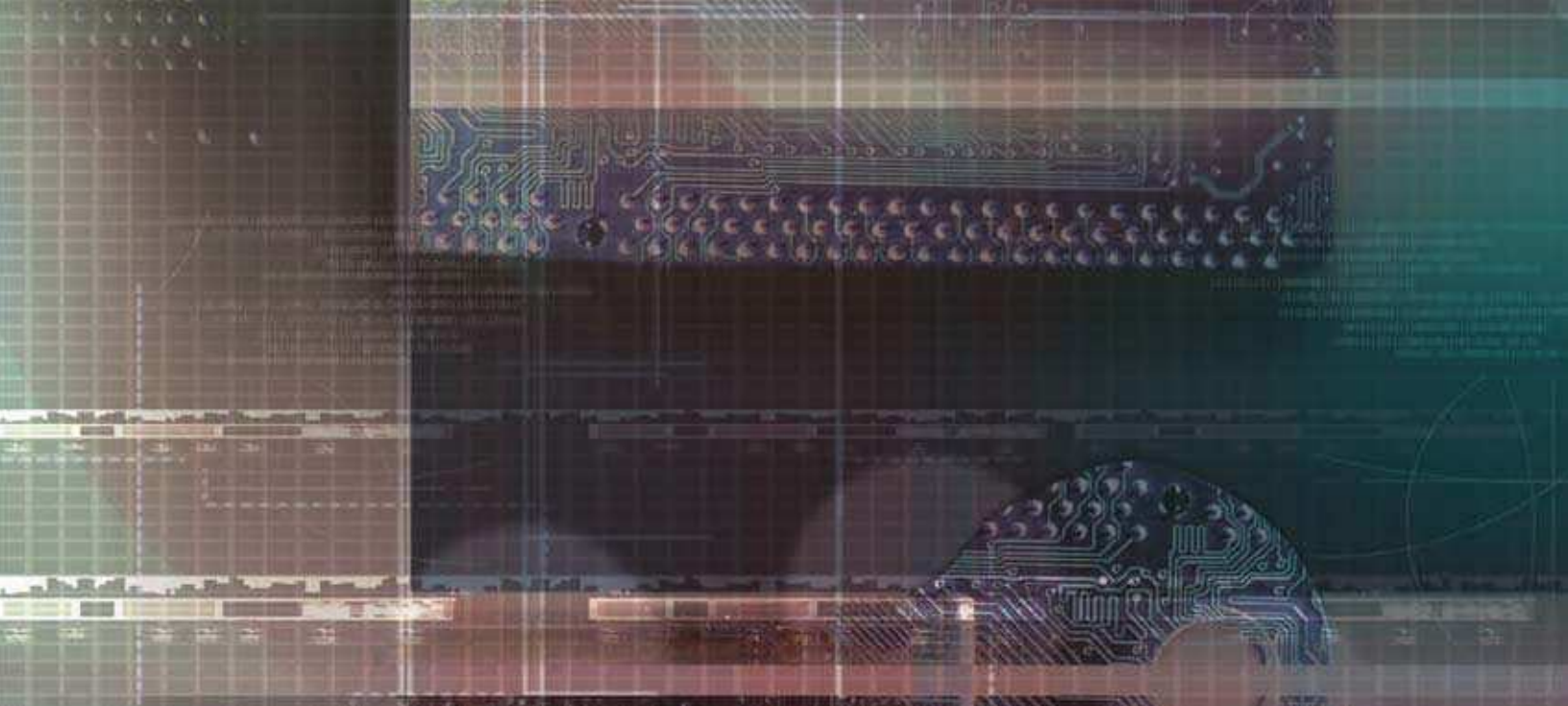
NSA's Information Assurance Directorate (IAD) responded to the growing use of commercial technology with a number of new initiatives. Some attempted to raise the level of security provided by commercial products, but the government market was far too small to have any real influence with successful commercial vendors. The much-publicized and valiant Multilevel Information System Security Initiative (MISSI) with its flagship Fortezza encryption card attempted to provide a high assurance overlay to bolster the security of commercial products, but it too lost the battle of cost, convenience, and interoperability in the desktop space. As the 90s drew to a close, NSA's approach to information assurance shifted to emphasize perimeter defense, intrusion monitoring, and risk management. This situation prompted NSA's Advisory Board to sound an alarm. The Board saw the government's most sensitive information systems being dominated by COTS technology incapable of providing the necessary levels of security, and a government market insufficient to influence vendors to provide the requisite protection. The Advisory Board called for a new strategy to be developed that could leverage the commercial technology that users wanted but still provide the higher levels of assurance they needed. The Board's report included a specific

challenge to NSA's Information Assurance Research Group to launch a new effort to deal with this problem. To ensure that their challenge was handled with appropriate urgency the Board insisted that a solution be developed within one year! The challenge was accepted, and the result was NetTop.

NetTop was originally described in the Fall 2000 issue of *Tech Trend Notes* (predecessor to *The Next Wave*). At that time the project had just started and we were developing many new ideas for potential applications of the technology. We optimistically thought that within three to five years we could get the technology into our customer's hands. Today, eight years have passed and NetTop has yet to achieve widespread use. So what happened? The editors of *The Next Wave* thought that a retrospective look at NetTop's history might be both interesting and informative. This article describes the evolution of our research and some of the novel approaches we attempted in order to deal with the perennial problem of technology transfer.

Early R&D

NetTop began as a research initiative responding to a challenge of NSA's Scientific Advisory Board. The intent of the project was to explore new concepts for security architectures. It was not envisioned



as the first stage of a new but traditional product development. Historically, NSA's product developments have mainly focused on link encryption solutions, many fielded in excess of 20 years but still capable of providing protection from cryptographic attacks even years after being removed from service. This approach to system security had worked well for decades, but it wasn't delivering the kinds of solutions needed to protect modern information systems against the new, active threats encountered in networking environments. A new model for Information Assurance (IA) solutions was being sought—one better suited to today's IT environment and capable of providing incremental security improvements over time.

The IA Research Group accepted the Advisory Board's challenge and established a senior-level tiger team to respond. To meet the one-year deadline, the group quickly focused its attention on some relatively well-understood technologies rather than defining a totally new research activity. Within several months we identified an approach using an interesting "technology cocktail" that looked very promising. The first component of the cocktail was a *refreshed* version of 1960s era virtual machine (VM) technology that had emerged from DARPA-sponsored research, and was

brought to market by a start-up company known as VMware. The second ingredient was an NSA prototype operating system—Security Enhanced Linux (SELinux)—that was gaining traction in the Open Source community. The combination of these technologies seemed to offer interesting possibilities for combining the COTS hardware and software that users wanted with the transparent security controls they needed.

Because of the unusual events that unfolded during the course of our work on NetTop, it became necessary for us to take our concept demonstration to a much more advanced stage of development than usual, and we found ourselves in the uncomfortable driver's seat of product developers. As we worked through the issues associated with developing NetTop for operational use, we had to answer a number of important questions:

"Can we do this?" – Will the technology work for the kind of applications that users want?

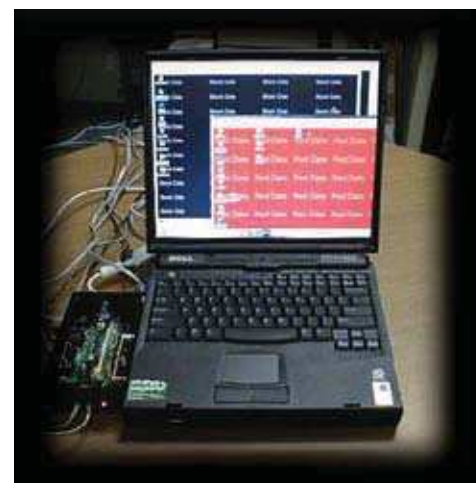
"Should we do this?" – Does the technology protect against expected attacks without introducing new and more serious problems?

"Will we do this?" – Can we deploy this new solution and sustain it in the field?

Our experiences as we attempted to answer these questions are the real story behind NetTop.

Can we do it?

After developing a crude first prototype of NetTop, most of our time was spent trying to determine if virtualized components were practical for use in systems that solved important user security problems. Although we encountered many technical challenges as we tried to integrate SELinux and VMware into a secure configuration, this portion of the project proved to be the shortest phase of the overall effort.



NetTop prototype

Does virtualization help to solve real problems?

One of the first uses for NetTop that we considered was as a Remote Access solution that would allow users to connect to a secure enclave remotely using a commercial laptop computer. Within several months we had a rudimentary prototype demonstrating this capability, and we soon realized that we should be able to do much more with system architectures using virtualized components. If we could efficiently support multiple virtual machines running concurrently on a single workstation, then numerous other types of solutions would be possible. The NetTop prototype we constructed turned out to be just one instance of a general architectural approach to building solutions that could be used to solve many different security problems.

Remote Access Solution Lessons Learned

Working with our first NetTop prototype helped us to distill a number of important characteristics and benefits of virtualization as an isolation mechanism.

▶ Isolating a security critical component like an IPsec encryptor in a separate container shielded from the behavior (or misbehavior) of the user's commodity operating system and applications was helpful in restricting the impact of software attacks. For example, an IPsec encryption stack

installed within a Windows OS can only be as trusted as Windows itself. But installing the same IPsec stack in its own container and linking it by a network connection to the Windows container provides an Inline Network Encryptor (INE) architecture that limits the avenues of attack to just the network interfaces. See Figures 1a and 1b.

▶ Providing multiple virtual machines for a user gave us an opportunity to create multiple single security level environments running simultaneously. This capability was sometimes confused with the traditional notion of Multi-level Security (MLS) in which a single environment protects information objects with multiple security levels. To help avoid this confusion we coined the term “Multiple Single Level” (MSL) to describe the capability that the NetTop architecture provided. See Figures 2a and 2b for a side-by-side comparison of MLS and MSL architectures. Using an MSL approach to architect a solution would involve using a collection of single security level VM's that could communicate with each other using network connections. This approach to designing solutions gave rise to the name NetTop—a **Network on a deskTop**. Each of the individual VMs would contain only data at one security level. Even security critical VMs such

as encryptors could be limited to processing data at one security level, thereby reducing their complexity compared to devices that manage data and keys at multiple security levels. Another approach that used the concept of isolated containers known as MILS (Multiple Independent Levels of Security) was promoted for use in embedded systems. While MILS technology provided very high assurance isolation, it didn't have NetTop's capability of hosting Windows or other legacy operating systems and applications, and so it wasn't as useful for typical user needs.

▶ Using partitions to separate information based upon integrity levels provided another capability that was useful in some applications. One such solution we developed—BoxTop—permitted the execution of suspicious, and potentially malicious, programs in a confined space and ensured that any harmful activity within that space could not spread further. We used one-way network connections to transfer content into the container, but no connection was provided for transferring data out of the container. This “virtual blast cage” could fall victim to an attack, but the damage was blocked from propagating further.

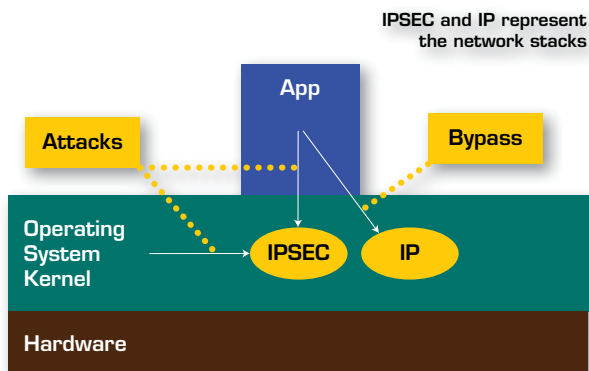


Figure 1a: No virtualization layer. Attacks against the IPSEC can come from an application or through the OS. An application can bypass the IPSEC and attack the IP directly.

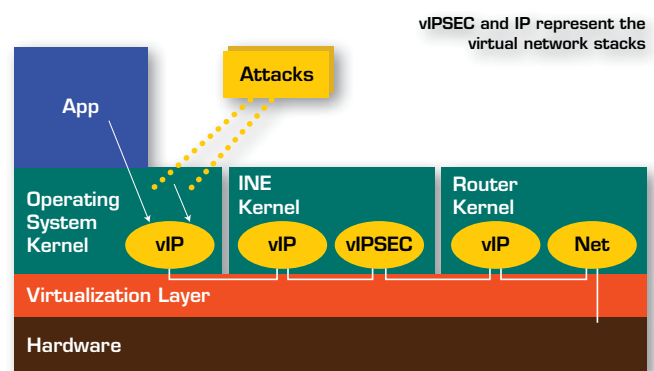


Figure 1b: With a virtualization layer. Attacks from the OS and application are confined.

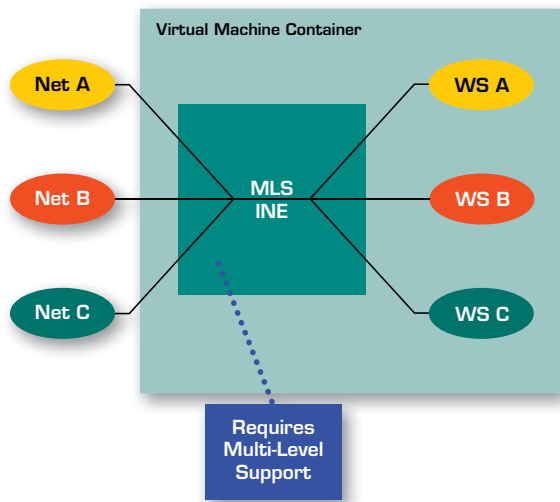


Figure 2a: Graphic of traditional Multi-level Security (MLS)

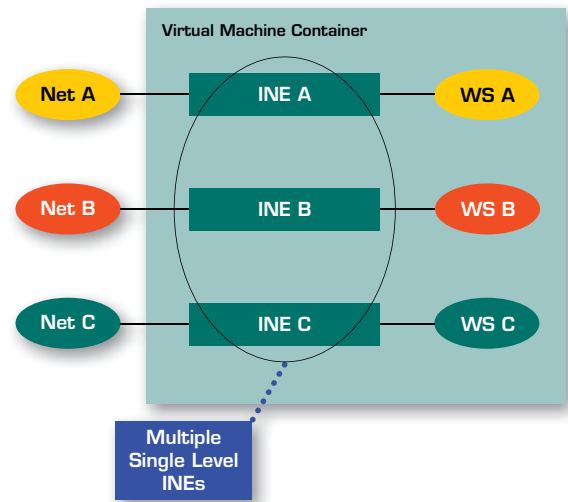


Figure 2b: Graphic of Multiple Single Level (MSL), a capability that the NetTop architecture provided

► The isolation provided by the use of separate virtual machines ensured that changes to system components that were not security critical would not impact security-critical components, and should reduce, if not eliminate, the need for lengthy re-evaluation of the overall system. This use of isolation improved NetTop’s agility by allowing it to quickly incorporate new commercial capabilities without disturbing security-critical components. We were able to quickly create versions of NetTop that used direct ethernet connections, modems, or wireless network adapters since no changes were required to the encryptor VM.

On July 13, 2000, two months earlier than requested, we met with the Advisory Board to describe the architecture and demonstrate the prototype we developed in response to their challenge. At the end of the briefing the Board surprised us with an unprecedented round of applause for what they saw as a creative and useful first step to dealing with security in COTS technology. Subsequent phases of the project would prove to be much more difficult.

Should we do it?

Study 1: Does NetTop introduce more problems than it solves?

We soon convinced ourselves that a variety of useful solutions could be designed using the NetTop architecture. The next important question that we had to answer was whether our approach would introduce more problems than it was solving. To answer this question we sponsored a workshop using some of NSA’s best security evaluators to assess our prototype. Using seven analysts over a ten-week period and with some limited input from VMware developers, we explored the ability of the core NetTop technologies—VMware running on a Linux host—to maintain isolation among virtual machines and to maintain isolation of the Linux host from the virtual machines.

The results of this first study were encouraging—no apparent show-stopping flaws were identified. The analysts were given full access to a VM and were able to write any program they wished in an attempt to crash another VM or the host OS. VMware workstation reliably withstood the attacks that were attempted and although a VMware virtual machine could be crashed, the host OS and other VMs were unaffected. Following these experiments, we expanded our relationship

with VMware through a Cooperative Research and Development Agreement (CRADA) to facilitate further NetTop development.

Study 2: What kinds of network attacks does NetTop prevent?

Our first NetTop study investigated the security robustness of a standalone workstation, but we still needed to address issues that might arise when network connections were allowed. So the next question we wanted to answer was whether there might be any unique remote attacks against a NetTop virtual machine solution that didn’t exist in an identical system using real machines. This was the focus of our next experiment.

In the summer of 2001 we were able to take advantage of a high profile NSA intern program established to develop network security experts. As the final project for the graduating class we devised an exercise to study network attacks against the NetTop virtual platform. The intern group was composed of fifteen network security specialists who worked over a period of twelve weeks and were led by one of NSA’s most talented and respected evaluators. Motivation in the group was high. They were eager to show that our solution was flawed. Some of the bolder analysts were confident that

they would uncover “holes big enough to drive a truck through.” The class project was scheduled to run through the end of September 2001.

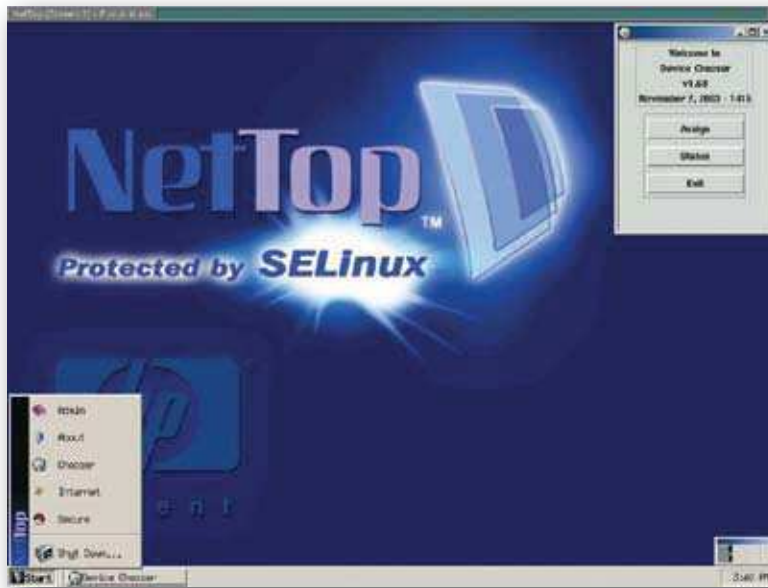
For this exercise we created a specific NetTop configuration that used two virtual machines running simultaneously, with each VM attached to a different network—in other words we created a “virtual KVM switch.” The user interface experience was one familiar to VMware users—two different desktops in two separate windows. We used SELinux as the host OS with a security policy crafted to provide maximum protection for the host. The sole function of the SELinux host was to provide an execution environment for the virtual machines. No user accounts or user applications were allowed on SELinux. It was a very tightly controlled configuration. In fact it was so tightly controlled that the interns requested a relaxation in the SELinux policy in order to allow an initial toehold for their attacks.

The group identified a number of areas that could be improved and they pointed out some lifecycle issues that should be considered in future deployments. But in the end this study reached conclusions very similar to the first, and no show-stopping problems were found.

Will we do it?

From the outset our goal for NetTop was as much about developing new technology transfer approaches as it was about developing new technology. The motivation behind the NSA Advisory Board’s challenge was that government needed more innovative approaches for leveraging commercial technology so that it could be used for sensitive applications. Government-unique IT developments (government off-the-shelf or GOTS) were far too costly to create and maintain, and frequently lacked capability compared with COTS offerings. It was

clear that users were simply not willing to pay a premium for higher assurance. Our previous experiences with MISSI and Fortezza were lingering reminders of this. It seemed to us that the most direct approach for dealing with this problem was to develop technology that was useful for government applications but that also had broad appeal outside of government. We believe that if we could stimulate the development of a large commercial market for this technology, the government could benefit from the cost advantages of large-scale COTS production. In effect we were conducting research in market development as much as in new security technology.



NetTop and SELinux

We used a relatively new NSA program—the Domestic Technology Transfer Program (DTTP)—to help us with our attempts at market development. This program had been established in response to US Code Title 15, Chapter 63, Section 3710, “Utilization of Federal technology,” to promote the transfer of government sponsored research to the public sector. The DTTP provided experts to assist us in numerous areas related to tech transfer including identifying candidate technologies, technology valuation, acquisition of ownership rights, finding transfer partners, establishing partnering agreements, negotiating transfer agreements, and overseeing relationships.

One of our first steps after demonstrating our prototype in March 2001 was to file a patent application to gain control of the intellectual property (IP) embodied in NetTop. In a somewhat unusual move for NSA, we also decided to seek a trademark for the name “NetTop,” since it had gained a fair amount of recognition and therefore seemed useful to control. We wanted to avoid the unfortunate situation encountered in the MISSI program when their flagship Fortezza token had to change its name from Tessera because of a trademark-filing oversight. The NetTop trademark proved to be very useful in later phases of the marketing program. Having protected NetTop’s IP and name, we began a search for industry partners capable of commercializing it.

While our main effort was to transfer our technology to a commercial partner, we knew that an NSA support group for NetTop was needed outside of the research organization. We believed that NetTop’s long-term success and its commercial development strategy needed a program office within the IAD—NSA’s arm responsible for developing security solutions. Unfortunately there wasn’t any pull from the IAD for NetTop or its component technologies—SELinux and VMware. NetTop was seen as lower assurance than the solutions that the IAD normally produced or endorsed. From our point of view, NetTop offered an approach that could deliver “high impact” to the assurance of customer missions rather than just “high assurance.” Our belief was that it provided a mix of functionality and value that users would embrace, rather than the high assurance products that were often developed but not widely used. We also believed that NetTop provided much better assurance than the COTS alternatives that customers were adopting. Furthermore we saw a migration

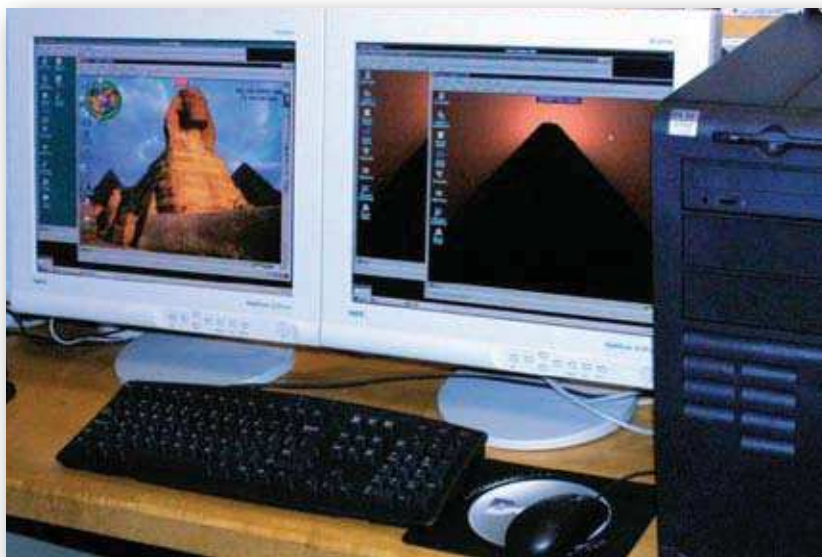
path for NetTop to even higher assurance solutions in the future. NSA's Advisory Board seemed to agree, and in a July 2001 meeting they recommended that the IAD establish a program office to manage the development of NetTop solutions.

The IAD approached the establishment of a NetTop program office very cautiously. A number of studies were undertaken to suggest a product roadmap, an assurance improvement roadmap, a business plan, and an assessment of NetTop's Total Cost of Ownership. The outline that was developed for productizing NetTop focused on the simple virtual-KVM configuration that was built for our ongoing security evaluation, with the added restriction that only adjacent security levels be permitted (e.g. Top Secret to Secret) and that cross domain data movement be prohibited. A management meeting to decide on the way forward was held on September 10, 2001—but no decision could be reached. Ironically, the terrorist attacks of the following day—September 11, 2001—and the US military response did determine the course of the program.

The US Central Command (CENTCOM), headquartered in Tampa, Florida, led the US military response to the terrorist attacks along with a large collection of coalition partners. IT support for this immense operation was a daunting task, and the crush of equipment required to access numerous coalition networks strained available space and power. On a visit to CENTCOM shortly after 9/11 an IAD representative noted this problem and wondered if NetTop might offer relief as a desktop reduction solution. This idea was suggested to IAD management as a unique opportunity for technology insertion. The intern class that was evaluating NetTop was still working when the idea of using it at CENTCOM surfaced. Because of their recent security evaluation

experience, the class was viewed as a useful sounding board, so they were polled regarding NetTop's suitability for use at CENTCOM. The consensus was that NetTop was sufficient for separating adjacent security levels (TS and S for example), but the group was reluctant to give an unqualified recommendation for its use to simultaneously access Unclassified networks. This endorsement was a significant milestone for the project, NetTop had been able to transform a highly motivated group of skeptics into supporters, albeit cautious ones.

IAD management agreed that NetTop could improve CENTCOM's operations, and that it should be quickly retooled



CenTop Workstation

for use in an operational environment. A NetTop Program Manager (PM) was named within the IAD's product development organization, but because of the urgency of the CENTCOM requirement, the research group was given responsibility for developing and fielding the production equipment.

The NetTop prototype needed extensive hardening and refinement to make it suitable for use by typical military operators who weren't hard-core system developers. To accomplish this, NetTop received a complete face-lift to remove the most visible signs of its Linux heritage. The user interface for CenTop—CENTCOM's custom NetTop implementation—was

redesigned from the top down to give it the familiar look and feel of a standard Windows workstation.

Other design changes targeted to CENTCOM's operational needs included an ability to access six networks simultaneously, dual monitor support for more desktop workspace, and the ability to run CENTCOM's standard Windows software load in each VM.

By the end of December 2001 a small team had been assembled to help the NetTop PM manage the myriad activities required to advance NetTop's development. Most important was the transition of NetTop to CENTCOM, which required the initiation of a security evaluation and generation

of the large body of documentation required by the accreditation process. Other important activities included establishing a small NSA NetTop pilot, collecting feedback on user and administrator acceptability, and providing support to NetTop developers.

The eventual transition of NetTop to CENTCOM was well intentioned but, unfortunately, not well executed. While the technology was well received by users,

the various support groups that had to administer it were not equally enthusiastic. A number of operational difficulties were encountered, and while most weren't due to design problems they nevertheless contributed to an overall negative initial impression of the technology. One of our major oversights was not having 24/7 NetTop support available from the outset. This problem was eventually corrected, but the delay proved to be a costly misstep in NetTop's first high-profile deployment. In retrospect, we should have ensured that NetTop had a high-level CENTCOM advocate as well as buy-in from the IT support groups.

After the initial intensity of the coalition military effort subsided, the

urgency to field NetTop at CENTCOM diminished as well, and with it the pace of the evaluation activities. Even with the urging of the NetTop project management office, it took over 18 months before formal decisions were reached about appropriate uses for the technology. Based upon a growing body of technical evidence, including the results of the two earlier evaluations and another evaluation on the CENTCOM-specific NetTop solution, the IAD Director finally approved NetTop for classified applications, but its use was limited to applications needing to separate Top Secret and Secret networks. Having taken over three and a half years to reach this point was disappointing, but the research team was never the less euphoric at reaching this important milestone. Unfortunately the feeling was short lived as we came to the realization that the use of NetTop throughout the Intelligence Community (IC) would require yet another extensive, bureaucratic accreditation process – DCID 6/3. This meant we had to socialize NetTop’s concepts

to yet another group of accreditors and Designated Approving Authorities (DAAs) that had never before seen this type of solution. Furthermore, the DCID 6/3 evaluation criteria used by the DAAs were not designed to address solutions like NetTop that had multiple operating systems running con-currently. DAAs rarely decide issues concerning operational use of security solutions without involving their peers, since individual decisions often create shared security risk across the community. So we once again found ourselves having to educate numerous decision makers about why NetTop should be approved for operational use. Eventually our efforts succeeded, and in the following five years NetTop began to

find use in various deployments. Although painfully slow at the outset, NetTop has continued to gain acceptance as a security solution within the IC and the DoD.

Finding Tech Transfer Partners

Shortly after our decision to pursue a tech transfer path for NetTop through the licensing of intellectual property, the research team initiated a series of meetings with potential commercial partners. The most promising partner initially was the Federal Division of Compaq Computers. Compaq management saw potential in NetTop to help them build a market in

a competitive market would be even better for the government. After two more years of discussions with other potential partners we negotiated a second NetTop license with Trusted Computer Solutions (TCS). TCS was much smaller than HP but very well established in the government market for security products, and they were highly experienced at working with the security accreditation process. TCS’s strengths seemed like an excellent complement to HP’s for developing a significant market for NetTop.

Several months after licensing NetTop, HP was able to generate some interest in the technology from their government customers, and they have continued to steadily grow their sales. Ironically, NSA did not become a strong customer because IT support at NSA had been outsourced to an industry consortium, and NetTop’s approach wasn’t consistent with the terms they had bid in their contract. TCS also began to see some interest in their NetTop offering shortly after they licensed the technology. Unfortunately we

didn’t see the dramatic uptake of the technology that we expected. After several years reflecting on this situation we began to understand why our expectations weren’t being met. Both of our NetTop partners drew their customers from the high assurance DoD and IC market space rather than the broader commercial market, and their revenues were heavily dependent upon selling services rather than products. There was little incentive for them to drive product costs down since they weren’t anticipating a mass consumer market for NetTop. In the high assurance government market, NetTop sales may grow but probably only at the pace of IT infrastructures replacement; so while we may eventually see increased deployments,



HP and TCS NetTop marketing material

security-related IT. Our discussions with Compaq were very positive, but they were soon interrupted because of the prospects of a merger with Hewlett-Packard. After completion of the merger in September 2001, discussions resumed with the new Federal Division of HP. But it was not until November 2002, almost two years after the start of discussions, that a NetTop license was finally negotiated. We viewed this milestone as a tremendous accomplishment and were certain that we would soon see a large commercial market for NetTop that the government could leverage. We were only partially correct.

While we worked with HP to help them refine NetTop, we continued to seek other commercial partners, since we believed that

they are likely to be over a much longer period of time than we expected. It's also unlikely that we will see costs drop as significantly as we had hoped.

Checking Our Projections

In our Fall 2000 NetTop article, "NetTop—A Network on Your Desktop," we speculated about the potential of virtualized architectures to deliver many more capabilities than the remote access solution we started with. Two of the solutions we described included a multi-domain access system and a coalition support system that provided dynamic collaboration environments. In the years since our original prototype, we went on to develop systems very similar to those we described. The workstation that we developed for CENTCOM, in fact, implemented our concept of a multi-domain access solution. We later developed a proof-of-concept system called the Intelligent Infrastructure Demonstration (IID) that showed how dynamic, private collaboration environments could be created rapidly to support mission needs for information sharing. In the following years, we created several other security solutions that we had never imagined, but which helped solve some of NSA's unique IT problems. These solutions further proved the adaptability and flexibility of NetTop's architecture.

Collaboration on Demand – Intelligent Infrastructure Demonstration

The original NetTop prototype was developed with a very tightly controlled configuration and lacked the flexibility to respond quickly to changing user needs. In short, it had the same limitations as the physical systems that it replaced. But in our *Tech Trend Notes* article we suggested the possibility of using virtualization technology to rapidly create new systems of various types and distribute them electronically wherever they might be needed. In 2003 we developed a demonstration of this concept in the Intelligent Infrastructure Demonstration (IID) system, shown in Figure 3. This system used a centralized server that could provision and deploy vir-

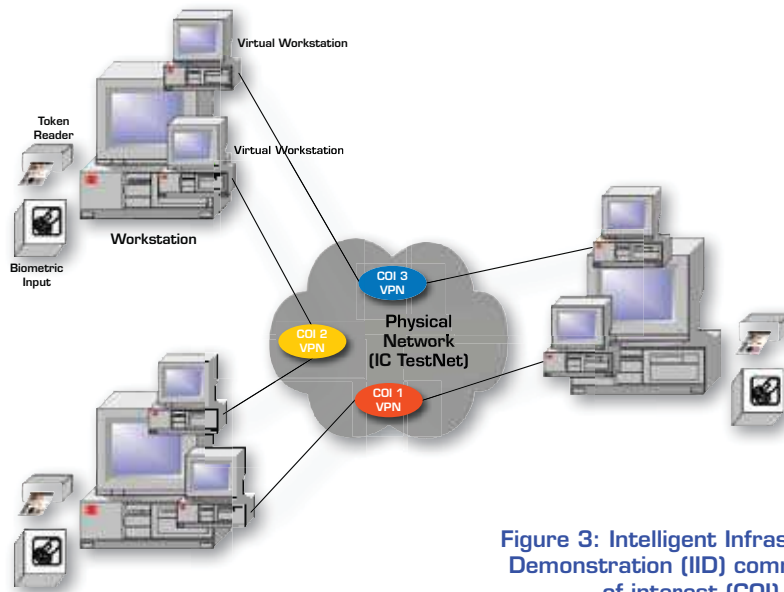


Figure 3: Intelligent Infrastructure Demonstration (IID) communities of interest (COI)

tual components including workstations, encryptors, firewalls, servers, etc., and then interconnect them to form private, collaborative workgroups or communities-of-interest (COI).

Our goals for the IID were to demonstrate how secure, collaborative environments could be set up easily and quickly to support the type of multi-party activities being performed at CENTCOM and other government organizations supporting the war effort. We wanted a capability that would enable the average analyst to set up a COI within minutes, tailor it to the needs of a particular group, and require no administrative support. As a model for IID

operation we used the Internet USENET system, which allowed individuals to easily create news groups for information exchange.

In our prototype each IID workstation was a NetTop that implemented a multi-factor (e.g. fingerprint, password, token, etc.) user authentication system and a special virtual machine dedicated to COI establishment and management functions. If a new COI were needed, a user could specify the members of the COI, the COI security level, the operating system to be used, and the set of application programs to be included. The management service would establish network servers to sup-



BoxTop interface

port the COI and invitations would be sent to the participants to join the COI. When a user accepted the invitation his workstation would receive software to configure itself to participate in the COI. IID users could participate in multiple COIs simultaneously and move easily among them via window selection.

Malware Protection: BoxTop

A different problem came to the NetTop research team from analysts who dealt regularly with documents, emails, and multimedia files that might contain malicious code such as viruses, worms, or other malware. Their management's security policy wisely required them to move any potentially dangerous content to a standalone system in order to ensure the integrity of enterprise operations. Unfortunately, the inconvenience of transferring data and the time-consuming cleanup process following an infection made the analysts' jobs unworkable. To deal with this problem we developed a NetTop spin-off called BoxTop that could safely and easily handle malicious content.

BoxTop used two virtual machines—one to replace the analyst's normal workstation and a second that operated as a sacrificial quarantine zone for processing malicious data. Analysts used a simple one-way data pump to move files into the quarantine zone and a special printing mechanism that allowed information to be exported safely for reports. To further improve analyst efficiency, we provided a mechanism to restore the quarantine zone to a sanitized state with the push of a button. SELinux provided us with an extensive set of security controls that we used throughout BoxTop's design to guarantee that it operated safely.

Protecting the Enterprise: ClearRealm

Following our work on BoxTop we discovered another enterprise IT problem that required a quick and innovative security solution. Unlike the case with BoxTop, where we were dealing with malicious data, this problem involved the use of enterprise software whose pedigree

was questionable. The normal software review and approval process was far too slow to meet mission needs, so managers were considering just installing the software and accepting the risk. We felt that we could leverage the flexibility of NetTop's architecture to quickly develop a solution that would allow the needed software to be used safely. In our ClearRealm design we encapsulated the questionable software as a web service and sandwiched it between two virtual firewalls to protect the integrity of the operational network. The firewalls were configured to ensure that the web service could not access the operational network and that it responded appropriately to user queries. ClearRealm proved to be very successful at meeting an urgent operational need.

ClarifyMind: Wireless NetTop

One of the significant benefits of NetTop's architecture is that it allows communication interfaces to be decoupled and isolated from components that provide security functionality such as firewalls, guards, encryptors, etc. Changes in network interfaces can then be made simply since they do not involve changes to security-critical code. We used this architectural approach in our original remote access solution to switch it between ethernet and modem connections, and we used it several years later in a wireless mobile solution called ClarifyMind. The prohibitive cost of new, wired connections to the desktop had denied Internet access to many analysts, so we proposed using NetTop to provide trusted wireless connections instead. To replace the security and access control of a wired connection we proposed using encryption of the wireless link. ClarifyMind's architecture combined a COTS laptop, an 802.11x card, MobileIP functionality, and an IP encryptor VM to deliver cost effective Internet access for analysts with the added benefit of allowing them to roam wirelessly. NetTop's architecture allowed us to isolate and protect the security-critical IP encryptor from other COTS components, and to ensure that all network traffic passed through the encryptor.

Application Partitioning

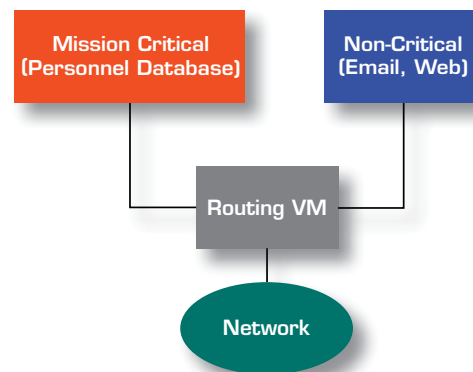


Figure 4: Using NetTop to isolate mission critical programs

Other Interesting Applications

Isolating Mission Critical Software

During the course of our work we met with many different users to try to understand their operational IT needs, and we spotted several additional usage scenarios for NetTop architectures that hadn't originally occurred to us. One scenario involved a need to protect the integrity of mission critical applications from the potential misbehavior of non-critical Internet application programs. The prototype solution we developed used one virtual machine to run important mission applications (in our case a program to track troop deployment) and a separate virtual machine to run non-critical programs such as web browsers. By separating applications in this way we

Legacy Migration

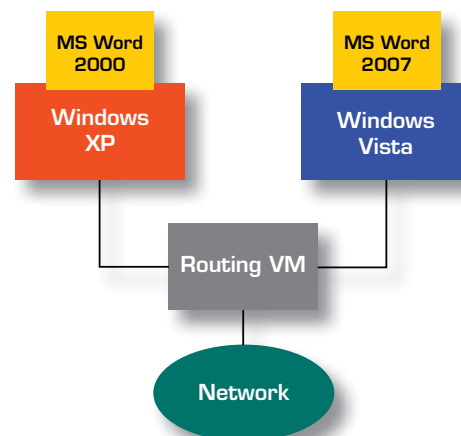


Figure 5: Using NetTop for incremental upgrade of new applications

showed that we could enhance mission integrity without losing desirable functionality. (See Figure 4).

Dealing with Software Migration

A second IT problem we discovered involved delays in deployment of current software versions across an enterprise. Users were often frustrated that they had to continue using outdated program versions until their entire suite of programs was updated to the latest operating system. This problem seemed like an ideal application for a NetTop solution that could provide multiple OS versions running simultaneously. This approach would permit applications to migrate individually and give users immediate access to the latest software versions. (See Figure 5).

Criticisms and Concerns

A healthy dose of criticism is not uncommon with any new technology, and it is par for the course with new security technology. But NetTop attracted a particularly broad set of vocal critics inside and outside of government. In part we believe this was because it represented a fairly radical paradigm change in its technical approach and also in its strategic approach to delivering security solutions via commercial partnerships.

The changes represented by NetTop were particularly uncomfortable for the IA community's traditional culture. NetTop attempted to strike a balance between the low assurance COTS technology being widely adopted by users and the very high assurance security solutions traditionally developed by government. It should not be surprising that neither community was totally satisfied with the result—but we believe that NetTop delivers a useful and credible blend of functionality and security.

Does NetTop Lower the Bar for Security?

One of the first criticisms of NetTop concerned its use of Linux as the host operating system. Linux was selected (actually SELinux) because we were able to customize it and rebuild the kernel to

provide tailored protection for the host OS. We were also looking ahead to the protection that SELinux' mandatory access controls could provide. To deal with potential vulnerabilities in Linux we used a number of design principles to minimize its risk of being exploited. First, we treated SELinux as an embedded OS and prohibited the use of any native user applications. VMware was the only application permitted. Our next step in reducing NetTop's attack surface, was to configure the system to make the SELinux host unreachable from any network connection. Since the time of the original NetTop evaluations SELinux has earned the same certification level (Common Criteria EAL 4+ LSPP, CAPP, and RBAC) as most of the host operating systems used in approved cross-domain solutions.

The environments originally intended for NetTop deployment only allowed access to users having the highest level of clearance of all connected networks in order to help deal with physical attacks. NetTop was often used in sensitive compartmented information facilities (SCIF) where all personnel were cleared to the highest level. Thus the environments intended for NetTop use were considered low risk.

Despite concerns expressed over the years about NetTop's security, it has held up well so far against sponsored evaluations as well as attempts to exploit weaknesses found in VMware's virtualization software. The combination of SELinux' mandatory access controls, VMware's isolation capabilities, and NetTop's tightly controlled architecture have proven effective at blocking attacks. While this is no guarantee that future problems won't emerge, continuing reviews by government analysts help to ensure that serious user problems are avoided.

Does NetTop Diminish the Market for High Assurance Technology?

In 1999 at the start of the NetTop project, users seeking trusted operating systems with a robust set of applications software faced a bleak situation. Those systems that offered the most functionality,

such as Microsoft Windows, fell short in security, and the several systems that offered very high assurance were lacking in functionality and interoperability. The NSA Advisory Board recognized that NSA's arguments for using the highest assurance technologies had long since ceased to be effective in the face of the COTS revolution, and users had opted for functionality over security. The Board was seeking a way to accommodate users' desire for fully COTS platforms while providing adequate assurance for sensitive applications, and they saw in NetTop a path for re-establishing a market for more secure products.

Prior to our work on NetTop, others had suggested a number of technical approaches to marry Microsoft applications with high assurance operating systems such as Trusted Solaris, but none of these products delivered acceptable performance and usability. Within NSA a high-assurance, thin-client architecture was developed as one possible way to give users a multiple security level capability, but it too had similar performance and usability problems. NetTop's architecture provided users with good performance from their Microsoft applications while at the same time keeping them isolated and protected in virtual containers. By interconnecting single-security-level containers, we could create solutions tailored to meet different users' needs.

We also saw in the NetTop architecture the potential to increase assurance over time by improving the security of its component parts. One way to increase assurance was by improving the host OS and virtualization environment that comprised NetTop's isolation infrastructure. A second way was to improve the assurance of individual virtual components. This second approach seemed to be particularly useful for creating specialized components that were hidden from users such as routers, firewalls, and encryptors. We had discussions with several high assurance OS vendors about migrating the NetTop architecture to their products, but concerns about the impact to their own products proved too difficult to overcome.

Does Virtualization Create a Security Problem?

One of the problems NetTop faced shortly after it was prototyped was that neither users nor system accreditors had a good understanding of the virtualization technology it was using. Although virtualization technology was originally developed in the 1960s, primarily for use with large-scale computers, its re-emergence on desktop computers in the late 1990s made it a novelty once again. Lack of understanding led to suspicion and fear about problems that might be lurking, but over a period of several years the issues associated with virtualization became better understood within NSA's IA community. Designing NetTop with security as a primary concern taught all of us a great deal about how to use virtualization prudently.

The commercial IT benefits of virtualization have been amply demonstrated by the dramatic growth of VMware over the past several years. But as the technology has become mainstream, the same security concerns that we encountered years ago re-emerged—this time from security professionals outside of government. Some of the recent security concerns with virtualization have been used as justification for dismissing NetTop's ability to provide robust protection. What many critics fail to appreciate is that most powerful tools can be used wisely or blindly with respect to security. We believe that NetTop's design offers a good example of how to use virtualization wisely.

Could We Do It Again?

The circumstances surrounding the development of NetTop were unprecedented for a research project. NSA's most senior Advisory Board identified a major security challenge and four of the Agency's most senior IA researchers, with over 80 years combined experience, were called upon to craft a solution. They, in turn, leveraged two of NSA's premier analyst development programs to perform in-depth security evaluations. The terrorist attacks of 9/11 were the catalyst that created a high priority military customer

and a committed NSA program to deliver an operational system from a research prototype. Over the course of several years, these events led us to some unique opportunities and to some useful insights into the business of research.

One of the unexpected benefits of our NetTop work was that it generated interest in partnering with research groups from a number of prominent IT companies. They sought to use our security expertise in operating systems and virtualization as a way to help them with their own technology developments. We saw an opportunity to use cooperative research as a way to gain significant leverage from our limited resources. We also saw potential in using cooperative research as a general technique for raising the bar in the assurance of commercial technologies. In effect we were developing a new COTS Security strategy based upon IA research collaboration.

A second benefit we derived from our work on NetTop and its spin-offs was that it served as a breeding ground for new areas of research. One interesting example was our early investigation of integrity checking for NetTop. This activity eventually blossomed into an important new area known as Measurement & Attestation, which deals with assured techniques for measuring the integrity of a computing platform and conveying this information to an enterprise health authority. This work could have future widespread use in the management of enterprise security, as well as more general application in developing trust among systems connected across cyberspace.

NetTop's developers had high expectations that the product's COTS-based blend of security, functionality, and flexibility would quickly generate a large market in public and private organizations that valued information assurance—unfortunately this didn't happen, and has been a major disappointment. What we came to realize was that sometimes technology changes are so dramatic that they require changes in organizational culture in order to succeed, and that

cultural changes often require a very long period of time.

Unfortunately for us, the cultural changes associated with NetTop involved changing not just one culture but two. The first was the crypto-centric, high-assurance product culture that was responsible NSA's long-standing reputation in security. NetTop used technologies unfamiliar and unproven to this culture, so they were considered unacceptable. NetTop was also built from COTS components incapable of delivering the assurance levels of GOTS products. It took years of experience with NetTop to build confidence to the point where it was accepted, at least for some applications.

The second culture that NetTop had to deal with was in the IAD's business community. In many ways this business culture was more difficult to influence than the high-assurance product culture. The bedrock of the business culture was the traditional, large-scale, FAR (Federal Acquisition Regulation) contract typically used for developing security products for customers in the national security community. Getting technologies like NetTop to customers involved a different approach, one similar to what industry would use. The new approach required aggressive practices in the creation and control of intellectual property, in marketing, and in developing and managing partnering relationships. We found little appreciation within IAD's business culture for the value of patents, trademarks, licenses, or open source developments because the use of these techniques were not deep-seated in the government business psyche. Contending with the business culture issues associated with NetTop required a major effort on our part, and added further delay to transfer of the technology. While we were somewhat successful in handling NetTop's unique business issues, to the IAD's business community it remains somewhat of an aberration rather than a useful, alternative business strategy.

We learned from experience that there is often a critical relationship between IA


technology and IT infrastructure, and that if a security technology isn't friendly to both users and to the infrastructure in which it operates, it just won't be accepted. We learned some hard lessons about this in our first NetTop deployment at CENTCOM. One unfortunate lesson occurred when one of the Windows VMs encountered the infamous Microsoft Windows "blue screen of death." The veteran operator reacted instinctively by hitting the machine's reset button. Unfortunately, this rebooted the entire set of virtual machines that were running and created a messy cleanup situation. We should have anticipated this would happen and ensured that only the crashed VM was rebooted. Other infrastructure management issues such as centralized auditing and remote platform configuration were also handled poorly in the original NetTop deployment. While these issues and many more have been addressed over time in improvements made to commercial NetTop products, they resulted setbacks for NetTop early in its development.

But—Would We Do It Again?

NetTop has not yet found widespread use outside of government, and this is a disappointment because we believed it would have many commercial uses. More importantly we hoped that commercialization would drive down the cost of the technology for government use, but this hasn't happened either. It isn't clear if a commercial market failed to materialize because of lack of user interest or because of inadequate marketing. Today NetTop is only available from vendors that focus on technology services for government rather than equipment sales. It remains to be seen if this will change in the future.

Although it would be impossible to recreate the extraordinary circumstances surrounding our work on NetTop, we have thought about whether we would undertake a similar effort in the future if we knew it would have a similar outcome. In short the answer is yes. The potential to have a major impact on customer mission assurance would still make such an effort

worth our investment. Through our work on NetTop we gained valuable expertise in an important, new technology area that allowed us to significantly advance NSA's acceptance of the technology and introduce new business strategies for product development. Another important consequence of our work was the ability to attract major industrial partners in collaborative research. These relationships have been very helpful in our research and particularly in developing next generation versions of NetTop through NSA's High Assurance Platform (HAP) project and our Secure Virtual Platform (SVP) research program.

Several thousands of NetTops have been fielded across elements of the IC and are being used operationally every day. Encouragingly, we have recently seen indications that NSA's own infrastructure upgrade initiatives are considering large-scale deployment of NetTop technology. While eight years is a long time to wait for this development, it is gratifying that our perseverance may finally be rewarded. 

Text Extraction From Color Images



1. Introduction

Color images mixing text and graphics offer an effective and popular method of information sharing. The Internet explosion bears full testimony to the imaginative ways vendors and communicators exploit the medium to flood the bandwidth with their chosen material. Automobile license plates are becoming yet another source of textual graphics, as states increasingly display color alphanumeric overlaid on various backgrounds or logos. In addition, magazine covers and video screen captures of broadcast news banners provide potentially huge databases to users interested in content retrieval using commercial search engines.

The mathematical complexity of multichannel information images—collections of shapes and colors arranged in unpredictable ways—has made automated text segmentation a difficult task. Most graphic designers do follow sensible guidelines so the characters stand out in some way, facilitating legibility. Such disparities will be exploited here to automatically determine the presence of text in a color image.

2. The Method

Text extraction from color images involves three proposed steps: a contrasting operation that converts the tri-chromatic (RGB) original to a grayscale version highlighting the text, thresholding the resulting image to bi-level black-or-white, followed by image character recognition (ICR) in any commercial software. The process is depicted schematically in Figure 1.

The key operation—contrasting—does not convert the color original to grayscale luminance by averaging, but instead combines the three RGB channels into one in which the text is enhanced. Unfortunately, no single such operation has been found to work in all cases; a parallel approach is therefore recommended to yield successful results for the wide class of candidate images expected in practice. The thresholding component of the proposed process also involves a multitude of (straightforward) operations.

2a) Contrasting

Given a tri-chromatic RGB image, many operations are available to separate text from its color background; the really successful ones tend to concentrate textual picture elements (pels) to the lower (or darker) end of the bit-level range, while background pels are shifted to the upper range. For example

$$C_1 = \text{Min}(R,G,B)/\text{Mean}(R,G,B)$$

and

$$C_2 = \text{Min}(R,G,B)/\text{Max}(R,G,B)$$

exploit the fact that color text overlaid on a complex background is typically saturated to enhance legibility; after all, the goal of the graphic designer is to use color to please the eye, not conceal content. A set of transformations that separate textual foreground from a more uniform background is:

$$C_3 = \{R/G, R/B, G/R, G/B, B/R, B/G\}$$

Other useful transformations of the RGB tri-chromatic channels include

$$C_4 = \{R/\text{Max}(R,G,B), G/\text{Max}(R,G,B), B/\text{Max}(R,G,B)\}$$

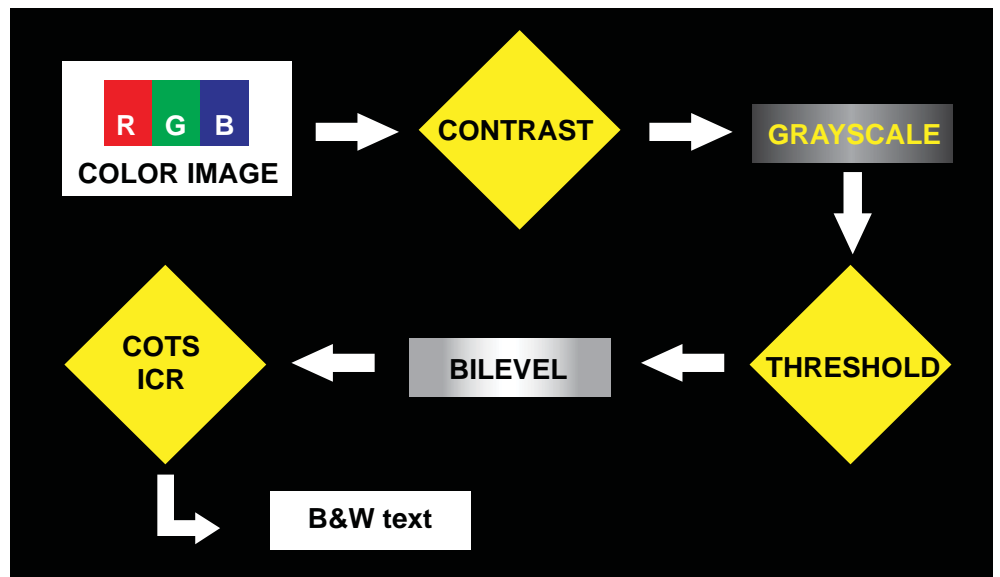


Figure 1: Color text extraction process

Logarithmic separators, e.g.,

$$C_5 = \text{Max}(\log(R), \log(G), \log(B))$$

and

$$C_6 = \text{Min}(\log(R), \log(G), \log(B))$$

also offer powerful contrasting opportunities.

Operations $C_1 - C_6$ (and many others) generate grayscale images from the color original, each of which may (or may not) contain unique text, or in fact any text at all; however, in view of the parallel approach advocated here, selected contrasting output must be thresholded and passed through the ICR software for final text extraction. Alternatively, bi-level results could be combined into one final image prior to ICR processing; however, this approach may risk mixing textual and non-textual output, thus obscuring character structure.

2b) Thresholding

Once contrasted grayscale images have been obtained from various operations, interval thresholding operations convert these images to bi-level [i.e., (0,1)] B&W. Once again, it is perhaps less than satisfying that no single operation has been found to work in all cases. For truly complex mixtures of text and graphics, one threshold estimate is

$$T_1 = [\text{mean}(C_r) - \sigma(C_r), \text{mean}(C_r) + \sigma(C_r)]$$

where C_r is the contrasted image whose values are restricted to the lowest quarter of the pel range, and σ denotes the standard deviation (or dispersion) of the image C_r . Pel values inside interval T_1 are then set to zero (black), and all others to one (white), producing an image ready for automated ICR processing.

Alternatively, since successful contrasting relegates textual pels to values well below the mid-range, histogram-based thresholding also generates a useful B&W image. Indeed, when restricted to this lower range, the most frequently occurring pel value—or histogram mode M_r —generates a threshold interval as (for example)

$$T_2 = [.5M_r, 1.5M_r]$$

The corresponding contrasted image range $C(T_2)$ restricted to this interval is then set to zero, while its complement containing background components is set to one. Of course, many color images are multi-modal in each channel—let alone as contrasted versions—so that their histograms show several strong peaks. In such cases interval thresholding around each maximum would generate a set of candidates for ICR processing.

For images with less complex backgrounds, practical threshold intervals are

$$T_3 = [\text{min}(C), \text{min}(C) + \sigma(C)]$$

$$T_4 = [\text{min}(C), \text{max}(C) - \sigma(C)]$$

$$T_5 = [\text{min}(C), \text{mean}(C)]$$

3. Examples

The process of contrasting, thresholding, and B&W reduction of an image containing color text will first be carried out for the RGB image in Figure 2 (left panel) below; the right panel displays the destructive effect that simple averaging

has on textual content.

Text colors were picked at random but character placement was intentionally chosen to provide a (perhaps unnecessarily) complex background and pose a realistic challenge to the proposed extraction method. Figure 3 below displays the three RGB channels in grayscale; the

greater the contribution in the color image of a given character string, the higher the intensity of that string in a particular channel.

Thus the word 'BEFORE' has high intensity in the Green channel, while 'MATCHES' is strong in Blue; however, color purity is not required, and, in fact, the word 'PLEASE', a mixture of Red and Blue, is also readily extracted by the proposed method. Now, to give a quantitative idea of the character 'blending' produced by channel averaging, Figure 4 below (left panel) displays a histogram counter for the averaged image; the right panel shows the histogram resulting from contrasting operation C_1 . The averaging procedure spread out the pel values and produced useless maxima below the desired mid-range level; contrasting, on the other hand, yielded an image with a clear maximum, so modal thresholding would retain textual content.

The result of applying contrasting operation C_1 to the tri-chromatic image is shown in Figure 5 below (left panel). Thresholding results from type T_1 and T_2 are displayed in the adjoining panels, the output completely processable by ICR engines. As emphasized above, it is not usually known beforehand which contrasting operation or thresholding type will produce textual separation; a parallel approach is generally necessary to successfully process a color text image for content.

More practical sources of color text images that should also benefit from automated extraction are automobile license plates. In recent years, such initially bland black-on-white metal tags have been adorned not only with color text but various logos and complex backgrounds, presumably for various promotional purposes. The family of contrasting operations detailed above produced successful reductions of these color plates to grayscale preparatory to thresholding and bi-level imaging.

The final example, involving a screen-captured news banner image, presents a significant processing challenge because



Figure 2



Figure 3

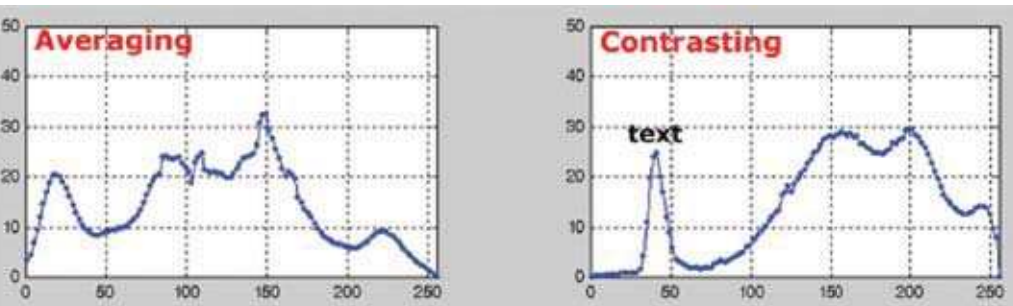


Figure 4



Figure 5

the text is in fact colorless white—see Figure 6; the right panel shows the channel average.

A histogram of the grayscale reduction (Figure 7) shows multi-modality, with peaks at pel values of 63, 100, and 248.

Interval thresholding around each maximum then produced three B&W im-

ages, but only the highest mode led to retention of textual image content (Figure 8), along with some artifacts that most ICR engines typically ignore.

4. Conclusion

The processing methods presented here allow automatic recovery of text from

color images by channel operations that (mostly) reduce only alphanumeric characters in the image to black, and then rely on actual recognition by commercially available ICR software. Although a parallel approach is advocated to handle the wide class of images expected in practice, the mathematical simplicity of the operations should pose no implementation problems.

ICR engines could in fact execute several contrast-then-threshold computations opaquely to the user, collate results internally, and then display unique textual output. This approach seems ideal for search engine or database matching applications. For example, automated license plate readouts leading to vehicle identification would allow more rapid detection of traffic infractions, either for automated ticketing or potential interdiction by the authorities. On the other hand, power users may instead wish to gain insight from each specific operation and then tailor composite processing to the particular family of target images encountered most frequently. A graphic user interface module within any commercial ICR engine would clearly facilitate such exploration.

The notion of feature extraction through channel operations need not be restricted to text detection applications. As medical imaging, airport screenings, and cargo inspections increasingly generate true color images, empirical channel combinations may indeed highlight tell-tale signs of diseased tissue, concealed weapons, or illegal materials not otherwise readily visible. 🇺🇸



Figure 6

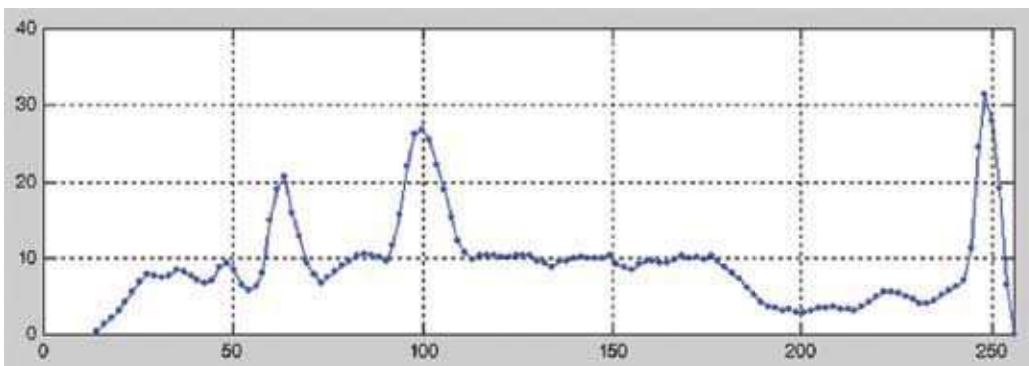


Figure 7



Figure 8



Web 3.0

“Are we there yet?”

The family road trip was a highlight of my childhood. My brother Harry and I would pile into the back of the old Dodge station wagon as the family headed out from Kansas on month-long treks across the country. Between rounds of 20 Questions—and the occasional squabble—we would ask with predictable frequency, “Are we there yet?”



For destinations such as the Ozarks or Badlands or Everglades, the answer was never certain. Even a *Welcome* sign posted to announce an official boundary seemed somehow arbitrary. The terrain always looked much the same through the windshield as it did from the tailgate window.

Contemplating the next generation of the Internet, popularly referred to as Web 3.0, I think back on those family road trips. We have maps to show us where we are and how far we've come. We read glowing accounts of what things will be like when we arrive. But when, exactly, *are we there yet?*

Web 3.0. There are numerous views of the coming Web era, depending on your vantage point. It has been described as *implicit, mobile, 3D, distributed, personalized*—the *metaverse*. Some critics argue that the name Web 3.0 is nothing more than a marketing term. But whatever name the next-generation Web eventually goes by, the change will transform the ways we experience life every day.

Reaching the next-generation *anything* is more often a journey than a destination. Along the road to Web 3.0 stand numerous milestones—Mosaic, XML, FaceBook. On the horizon lie OWL and SPARQL and Twine. These technologies serve more as filling stations than rest stops, fueling the journey along the Internet highway.

When Tim Berners-Lee (now Sir Berners-Lee) inaugurated the World Wide Web, in 1991, his aim was “to allow links to be made to any information anywhere.” The marriage of hypertext with the Internet made the original version of the Web a reality. Web sites and HTML pages began popping up like aspens along a mountain stream.

The Web's phenomenal growth in popularity during the 1990s—hosting more than 20 million websites by the end of the decade—fueled the dot.com boom that closed out the last century. When the technology bubble burst in 2001, pundits called for a reinvention of the Web—a *Web 2.0*—to carry the world into the new millennium.

While the first generation of the World

Wide Web had provided access to vast amounts of data, Web 2.0 took on a more personal tone. During the aught years, Web-based communities have coalesced around social networking sites, and youth everywhere turn to blogging and file sharing as their social lifeblood.

In the Web that is to come, the virtual and the real will seamlessly interact, informing each other across a shrinking digital divide.

While “Netheads” began downloading podcasts and mySpace pages, industry leaders were already busy mapping out the route to the next-generation Web. In 2001, Tim Berners-Lee proposed a *Semantic Web*, where machines not only find what we're looking for, but they also understand what we want.

For many people, Web 3.0 and the Semantic Web are nearly synonymous. But a broader vision for the next-generation Web foresees our virtual and offline worlds increasingly merge—virtual environments, virtual markets, virtual experiences, and even virtual *selves* paralleling those in the real world. In the Web that is to come, the virtual and the real will seamlessly interact, informing each other across a shrinking digital divide.

Even though the age of Web 2.0 only now is starting to mature, we are already catching our first glimpses of the digital landscape we will next inhabit. Over the next few years, people in many parts of the world will be entering the foothills of what inevitably will be called Web 3.0. If the timeline for the first iterations of the Web is an indicator for the future, we will have scaled the summit of Web 3.0 by the end of the next decade. What will we see when we look back over the road we traveled to

get there? And what will we see when we look forward to the digital mountain range that lies ahead?

Web 2.0 provided a two-dimensional platform for creating and exchanging information. A more three-dimensional Web 3.0 weaves straight-forward transactions into intricate patterns of relationships. The 3D Web provides for complex interactions among people, systems, and information in a do-se-do like dance of communication.

The emergence of Web 3.0 is being propelled by the evolution of technological, market, and social systems. Its use will be characterized by the integration of all three.

- *New technologies are driving the development the Semantic Web, APIs and Web services, and rich applications.*
- *A changing market model is placing more emphasis on an implicit Web—one that is able to grab and hold our attention.*
- *Social networking has already changed how we view ourselves and interact with other people and the world at large.*

Although these systems are interrelated and they reinforce each other in a variety of ways, each provides a unique perspective for conceiving of Web 3.0.

The Semantic Web

For the next generation of the Web, Tim Berners-Lee envisions a Semantic Web, a model for the evolution of his original Web design. Interoperability is the backbone of the Semantic Web, where independent applications can access the same data and reuse it in “unexpected ways.” The Semantic Web benefits cyber citizens—or *netizens*, by unobtrusively carrying out imaginative and sophisticated tasks.

For the Semantic Web to work, data has to carry metatags that compose a description of a variety of characteristics. These metatags are expressed by RDF (Resource Description Framework) and OWL (Web Ontology Language)—formal languages endorsed as Semantic Web specifications by the World Wide Web Consortium (W3C).

The Semantic Web uses metatags to define the semantics of data structures, map between them, and publish data records. SPARQL, a new query language, can then be used to search across the records to deliver what Berners-Lee calls the *intelligent web*.

The Semantic Web goes beyond connecting information and people. The *Semantic Wave 2008 Report*, published by Project10X, concludes that the goal of

the Semantic Web is to make our online experience “more relevant, useful, and enjoyable.” As Josh Catone, lead writer for *ReadWriteWeb* blog put it, “When machines understand things in human terms, and can apply that knowledge to your attention data, we’ll have a web that knows what we want and when we want it.”

APIs and Web Services

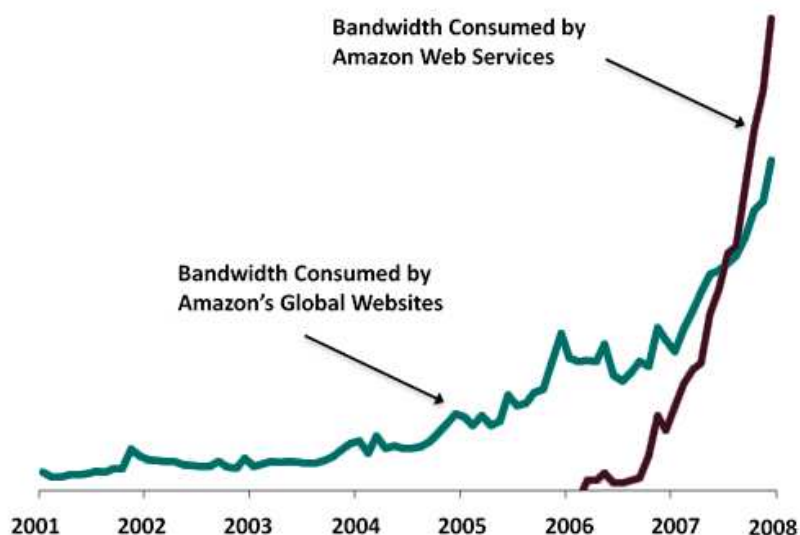
The implementation of the Semantic Web depends on the extent to which authors metatag information. But a more immediate way to extend flexibility to the Web is through the use of APIs and Web services. An API (application programming interface) is a source code interface that supports requests made by computer programs. Closely related to APIs are Web services, which provide ways for different software applications to work together, even when they are running on a variety of platforms. APIs and Web services make it possible for developers to integrate popular online services such as YouTube and Google Maps with their websites.

Web services join up to create a *computing cloud*—the term applied to nearly limitless high-speed Internet access to a variety of Web applications. Web services often generate more traffic than the web site itself. For example, the API for Twitter has 10 times more traffic than the web site. Open-source proponents foresee Web 3.0 being ushered in on a wave of public Web services and mashups that are used to create novel Web applications.

Mobile Access

Netizens increasingly expect to have Internet access from any device in any place at any time. Although the desktop computer may be a mainstay for accessing the Web well into the foreseeable future, people around the globe are demanding greater flexibility.

Logging onto the Internet through a computer modem is already an anachronism. Wireless access to the Web has moved beyond office and home networks to include mobile phones, pagers, tablet PCs, location devices, and even game systems. As the number of options for how we communicate has increased, so has



Within two years after Amazon introduced Web Services, in 2006, consumer demand for Amazon Web services significantly outstripped demand for the Amazon website. Source: Amazon Web Services, 2008

the demand for a single mobile device to handle them all.

WiFi and satellite Internet have effectively cut the tether that bound Web browsing to the personal computer. Numerous coffee shops, restaurants, hotels, airports, and even entire communities are joining the ranks of Internet cafés to provide Web access to anyone within range, often for free.

Web access is even being extended to world travelers. In the spring of 2008, Thalys International began providing broadband Internet access for passengers onboard its high-speed trains on routes across Europe—the first service of its kind for international travel. Net access has also taken to the skies. Lufthansa and Singapore Airlines have been offering passengers broadband Internet access on some routes since 2005, and airlines in the US introduced in-flight Internet in the summer of 2008.

Future generations will expect continuous connection to the IT cloud. In Web 3.0, each netizen will expect constant and total immersion in a variety of social networks. The idea of “logging on to the Net” will seem as quaint as using a payphone is today.

User-created Content

User-created content is the hallmark of Web 2.0. Millions of Internet users today regularly contribute to blogs to express their opinions, vote on the quality of products and comments, build personal profiles that reveal their hopes and desires, and produce videos to share with the world. Development tools and bandwidth are becoming more readily available to provide every nethead with an outlet for their digital creations.

The trend for self expression is sure to carry on into Web 3.0. The difference will be in how intellectual property is created and distributed, with authors having the option to control what they create or open up their work for collaboration. Web 3.0 will provide the tools to turn every Web contributor into an entrepreneur.

Products such as the Geenius entrepreneurial portal already are showing up that combine a social networking platform with

The Web experience will be so customized that distinctions among pages, posts, SMSs, maps, and graphs will become irrelevant. These elements and more will be integrated and arranged into a personalized Web design.

a user-created content revenue model. The strategy is to empower entrepreneurs and subject matter experts with tools to build global e-businesses and communities like Facebook and sell their intellectual property content while being linked through a portal like Amazon. The goal is to launch thousands of subject-specific communities to meet the demands of growing user-created content industries.

While users are creating and adopting a rapidly growing number of Web apps, they are also responsible for much of their propagation. Email, social networks, blogs, and other personal outlets serve as hosts for virally distributing these applications.

Personalization

The Web is becoming less an online resource and more a digital extension of each user. *New York Times* journalist John Markoff suggests that Web 3.0 will do away with the notion of pages altogether. The Web experience will be so customized that distinctions among pages, posts, SMSs, maps, and graphs will become irrelevant. These elements and more will be integrated and arranged into a personalized Web design.

ReadWriteWeb, in 2007, sponsored a contest to define Web 3.0. The winning entry in the Serious category was submit-

ted by Robert O’Brien, who proposed that Web 3.0 is “A decentralized asynchronous me.”

O’Brien went on to characterize the evolution of the Web in this way:

Web 1.0 was a centralized *them*.

Web 2.0 is a distributed *us*.

Web 3.0 will be a decentralized *me*.

The coming Web is emerging as highly personalized. The intimacy of each user’s Web experience has led to what has been described as an *attention* economy. Attention economy theory is attributed to US economist Michael H. Goldhaber, who, in 1997, proposed that the *currency* in the new Web economy is human attention. Getting and holding people’s attention is the aim of every enterprise, marketer, and blogger.

The economic and social forces vying for our attention have contributed to the emergence of what has been called an *implicit web*. Our attention has become an increasingly scarce commodity. With burgeoning amounts of emails—spammed and otherwise—we have been forced to be more selective about where our attention is directed. The things we pay attention to are those we consider to be of the highest value to us. We attend to them naturally, subconsciously—*implicitly*.

Many web sites determine the value of a product, comment, or service by the amount of attention it receives. Most search results are ranked based on how frequently items are viewed. Products are rated by the number of purchases made. Blogs are promoted and demoted according to popular vote.

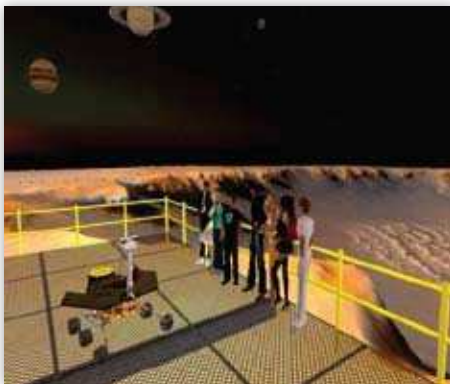
Methods for tracking attention will become increasingly sophisticated in Web 3.0. Each Web experience will be accompanied by a personal history of who we contact, where we click, and what we copy. A complex log of how we have spent our attention in the past will guide where we are directed in the future. An Implicit Web will operate in the background as a kind of digital subconscious.

Virtual Environments

Web 3.0 will likely take advantage of the visual metaphors afforded by increas-



Deacon Lunasea is the author's personal avatar in *Second Life*.



Second Life avatars tour Victoria Crater, a Martian landscape created above. NASA's CoLab Island.

ingly realistic virtual worlds. Dramatic gains in computer processor and graphics card speeds have yielded amazingly authentic looking 3D environments for the PlayStation and Xbox entertainment systems. As even handheld devices move into Terahertz territory, virtual worlds will become ubiquitous in Web 3.0.

The generation that grew up playing *World of Warcraft* is accustomed to following visual cues to guide an avatar—the digital character that represents the gamer—to explore a virtual world. Navigating the Web using mouse pointers and buttons could easily give way to wandering through a virtual landscape and “picking up” the things you want to have or examine.

MUVEs (multiuser virtual environments) such as *Second Life* are likely to be the forerunners of the type of user interface that will replace the countless pages that clutter the current Web. Whether for

shopping or information searches, clicking links to open Web pages will evolve into teleporting to different rooms and islands. Items will be selected from virtual shelves instead of from lists. Your shopping cart will be just that—a virtual basket loaded with the items you have selected.

Numerous universities, businesses, and government organizations have established a presence in *Second Life*. Some residents painstakingly replicate their existing brick-and-mortar infrastructures to provide a real sense of presence. More recently, the trend has been to create fanciful spaces that take advantage of the freedom from the constraints of mechanical systems, structural integrity, and even gravity.

These virtual communities in *Second Life* serve as shopping centers, lecture halls, dance clubs, conference rooms, and experimental labs. In Web 3.0, MUVEs could bring about the redefining of *tele* in *teleconference*, *telecommute*, and *telecourse* to mean *teleport* rather than *televise*.

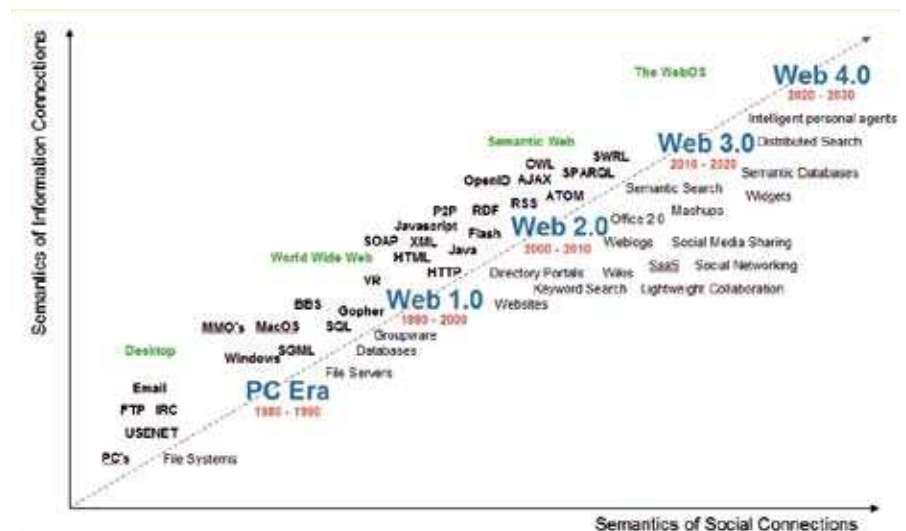
In addition to teleporting from place to place, avatars will be able to walk, swim, and fly, providing users with a sense of continuity of space and time. More importantly, in Web 3.0 your avatar can supply a sense of continuity of self, bridging the real and virtual worlds we will increasingly divide our time between.

Persistent Avatars

Citizens of *Second Life* are able to customize their avatars to generally resemble themselves. As character fidelity increases in Web 3.0, virtual experiences will become more authentic. Clothing retailers already let avatars try on clothing so shoppers can visualize how outfits might look on them. With more precise body measurements, avatars can serve as models for custom tailored clothing that is shipped to your home.

The avatar of the future could also possess a physiological profile that parallels real-world characteristics. Burn patients in some hospitals are now being treated effectively for pain simply by interacting in a virtual arctic setting. For future avatars that more accurately represent their owners, virtual medical check-ups could lead to real-world treatments.

Some psychologists have set up their virtual couches in *Second Life*. They get to know their patients through the alternate personae of their avatars. MUVEs are also being used to create environments for clients to develop social interaction skills and work through traumatic experiences. As artificial intelligence grows more sophisticated in Web 3.0, non-player characters—the autonomously intelligent agents in a virtual world—will know an avatar's personal history and be able to probe for greater understanding, so it can



The productivity of the Web is expected to rise steeply as the semantic Web gains a foothold and we enter the era of Web 4.0. Source: Radar Networks & Nova Spivack, 2007 www.radarnetworks.com

model preferable behaviors. Virtual world experiences can then transfer to real world gains.

Web 3.0 will accommodate persistent avatars. Just as phone number portability makes it possible to keep the same contact information for life, avatars will represent their real-life owners across various platforms. The customizable character that represents you in one MUVE will possess the same attributes and experiences in others. Persistent avatars can build on prior conditions and events. As an avatar gains experience, the beefier database capabilities of the next-generation Web will be able to store and retrieve that information for future encounters.

Beyond Web 3.0

The semantic Web can lay the groundwork for a truly intelligent Web. Whereas Web 3.0 links massive amounts of information to create a single, worldwide database, Web 4.0 will change how we access and process that information. In Web 4.0, desktop computing will give way to *Webtop* computing. A host of Web services will be available through a Web-based operating system, or WebOS.

This “really smart” Web will serve as a personal assistant. Web 4.0 agents will be able to reason and make decisions. The future Web will automatically filter our email, file documents, tag information, and dispatch with a host of other bothersome tasks that plague us today. But Web 4.0 will do more than menial chores. A Web that can access data from a multitude of sources and evaluate the information it finds will be able to help us choose the cars we buy, the medical procedures we undergo, and even the people we date.

Nova Spivack, founder and CEO of Radar Networks and developer of the semantic Web service Twine, foresees the emergence of Web 4.0 as early as 2020. In a 2008 interview for the video documentary series *Learning from the Future*, Spivack says we will have left Web 3.0 when “...the web moves from just a knowledge base to a kind of global mind, an intelligent entity comprised of billions of pieces of software and billions of people working together to make some new form of intelligence that transcends human or machine intelligence on its own.”

Living in Web 3.0

Web 3.0. “Are we there yet?”

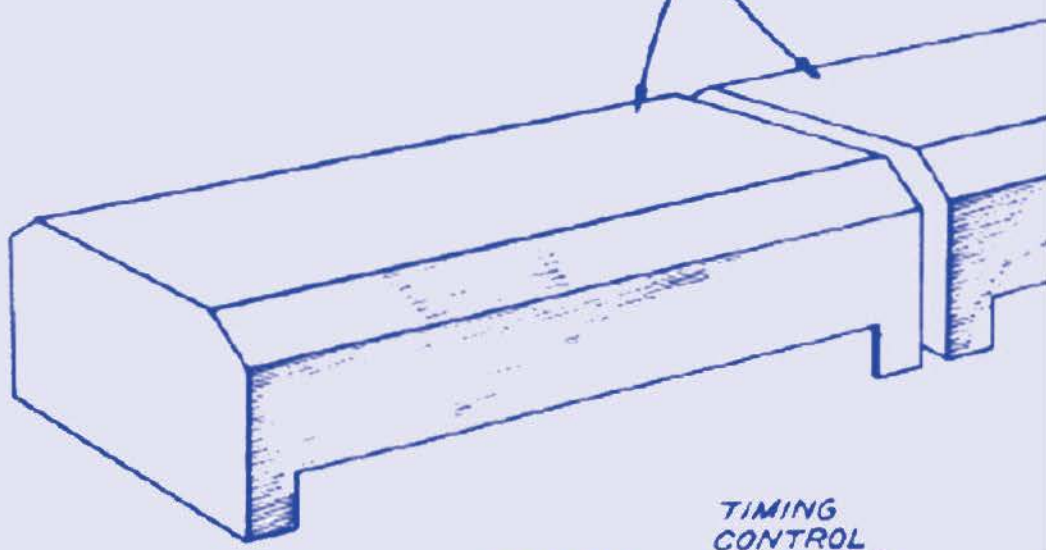
Some pundits will argue that we barely have a foothold in a Web 2.0 world. Others blister at the very idea of trying to erect arbitrary boundaries by ascribing versions to our collective cyber experience.

But our cyber experience has already transformed the world in a matter of only a few brief decades—a transformation we have passionately embraced, despite its negative consequences. By imagining in the virtual world of our minds what lies ahead, we stand a better change to prepare for the coming generation of experiences.

Like the family vacation, it’s time to pack up the station wagon with the gear we think we might need as we set off for a new and exotic destination. It might be another decade before we can look back and realize that we aren’t in Kansas anymore. And from the mountaintops of that new world, we’ll look to the next horizon and try to imagine what life will be like as we set off for Web 4.0. 🗺



DRUM COVERS



TIMING CONTROL CONTACTS

DRUM #2 (KA)

BASE 10 TO BASE 2

BRUSHES

TEMPORARY ONE-CYCLE SWITCH

BASE 2 IN

BASE 2 OUT

FILAMENT TRANSFORMER

