Table of Contents

Nuclear Fusion Power – An Overview of History, Present and Future

Stewart C. Prager
Department of Physics
University of Wisconsin – Madison
Madison, WI 53706, USA
E-mail: scprager@wisc.edu

***Summary*—Fusion power offers the prospect of an almost inexhaustible source of energy for future generations, but it also presents so far insurmountable engineering challenges. The fundamental challenge is to achieve a rate of heat emitted by a fusion plasma that exceeds the rate of energy injected into the plasma. The main hope is centered on tokamak reactors and stellarators which confine deuterium-tritium plasma magnetically.**

Keywords-Fusion Energy; Hydrogen Power; Nuclear Fusion

I. INTRODUCTION

Today, many countries take part in fusion research to some extent, led by the European Union, the USA, Russia and Japan, with vigorous programs also underway in China, Brazil, Canada, and Korea. Initially, fusion research in the USA and USSR was linked to atomic weapons development, and it remained classified until the 1958 Atoms for Peace conference in Geneva. Following a breakthrough at the Soviet tokamak, fusion research became 'big science' in the 1970s. But the cost and complexity of the devices involved increased to the point where international co-operation was the only way forward.

Fusion powers the Sun and stars as hydrogen atoms fuse together to form helium, and matter is converted into energy. Hydrogen, heated to very high temperatures changes from a gas to plasma in which the negatively-charged electrons are separated from the positively-charged atomic nuclei (ions). Normally, fusion is not possible because the strongly repulsive electrostatic forces between the positively charged nuclei prevent them from getting close enough together to collide and for fusion to occur. However, if the conditions are such that the nuclei can overcome the electrostatic forces to the extent that they can come within a very close range of each other, then the attractive nuclear force (which binds protons and neutrons together in atomic nuclei) between the nuclei will outweigh the repulsive (electrostatic) force, allowing the nuclei to fuse together. Such conditions can occur when the temperature increases, causing the ions to move faster and eventually reach speeds high enough to bring the ions close enough together. The nuclei can then fuse, causing a release of energy.

II. FUSION TECHNOLOGY

In the Sun, massive gravitational forces create the right conditions for fusion, but on Earth they are much harder to achieve. Fusion fuel – different isotopes of hydrogen – must be heated to extreme temperatures of the order of 50 million degrees Celsius, and must be kept stable under intense pressure, hence dense enough and confined for long enough to allow the nuclei to fuse. The aim of the controlled fusion research program is to achieve 'ignition', which occurs when enough fusion reactions take place for the process to become self-sustaining, with fresh fuel then being added to continue it. Once ignition is achieved, there is net energy yield – about four times as much as with nuclear fission. According to the Massachusetts Institute of Technology (MIT), the amount of power produced increases with the square of the pressure, so doubling the pressure leads to a fourfold increase in energy production.

With current technology, the reaction most readily feasible is between the nuclei of the two heavy forms (isotopes) of hydrogen – deuterium (D) and tritium (T). Each D-T fusion event releases 17.6 MeV (2.8×10^{-12} joule, compared with 200 MeV for a U-235 fission and 3-4 MeV for D-D fusion).[a] On a mass basis, the D-T fusion reaction releases over four times as much energy as uranium fission. Deuterium occurs naturally in seawater (30 grams per cubic metre), which makes it very abundant relative to other energy resources. Tritium occurs naturally only in trace quantities (produced by cosmic rays) and is radioactive, with a half-life of around 12 years. Usable quantities can be made in a conventional nuclear reactor, or in the present context, bred in a fusion system from lithium.

DOI: 10.21307/ijanmc-2019-064

[b]Lithium is found in large quantities (30 parts per million) in the Earth's crust and in weaker concentrations in the sea.

In a fusion reactor, the concept is that neutrons generated from the D-T fusion reaction will be absorbed in a blanket containing lithium which surrounds the core. The lithium is then transformed into tritium (which is used to fuel the reactor) and helium. The blanket must be thick enough (about 1 metre) to slow down the high-energy (14 MeV) neutrons. The kinetic energy of the neutrons is absorbed by the blanket, causing it to heat up. The heat energy is collected by the coolant (water, helium or Li-Pb eutectic) flowing through the blanket and, in a fusion power plant, this energy will be used to generate electricity by conventional methods. If insufficient tritium is produced, some supplementary source must be employed such as using a fission reactor to irradiate heavy water or lithium with neutrons, and extraneous tritium creates difficulties with handling, storage and transport.

The difficulty has been to develop a device that can heat the D-T fuel to a high enough temperature and confine it long enough so that more energy is released through fusion reactions than is used to get the reaction going. While the D-T reaction is the main focus of attention, long-term hopes are for a D-D reaction, but this requires much higher temperatures.

In any case, the challenge is to apply the heat to human needs, primarily generating electricity. The energy density of fusion reactions in gas is very much less than for fission reactions in solid fuel, and as noted the heat yield per reaction is 70 times less. Hence thermonuclear fusion will always have a much lower power density than nuclear fission, which means that any fusion reactor needs to be larger and therefore more costly, than a fission reactor of the same power output. In addition, nuclear fission reactors use solid fuel which is denser than thermonuclear plasma, so the energy released is more concentrated. Also the neutron energy from fusion is higher than from fission – 14.1 MeV instead of about 2 MeV, which presents significant challenges regarding structural materials.

At present, two main experimental approaches are being studied: magnetic confinement and inertial confinement. The first method uses strong magnetic fields to contain the hot plasma. The second involves compressing a small pellet containing fusion fuel to extremely high densities using strong lasers or particle beams.

III. MAGNETIC CONFINEMENT

In magnetic confinement fusion (MCF), hundreds of cubic metres of D-T plasma at a density of less than a milligram per cubic metre are confined by a magnetic field at a few atmospheres pressure and heated to fusion temperature.

Magnetic fields are ideal for confining a plasma because the electrical charges on the separated ions and electrons mean that they follow the magnetic field lines. The aim is to prevent the particles from coming into contact with the reactor walls as this will dissipate their heat and slow them down. The most effective magnetic configuration is toroidal, shaped like a doughnut, in which the magnetic field is curved around to form a closed loop. For proper confinement, this toroidal field must have superimposed upon it a perpendicular field component (a poloidal field). The result is a magnetic field with force lines following spiral (helical) paths that confine and control the plasma.

There are several types of toroidal confinement system, the most important being tokamaks, stellarators and reversed field pinch (RFP) devices.

In a tokamak, the toroidal field is created by a series of coils evenly spaced around the torus-shaped reactor, and the poloidal field is created by a system of horizontal coils outside the toroidal magnet structure. A strong electric current is induced in the plasma using a central solenoid, and this induced current also contributes to the poloidal field. In a stellarator, the helical lines of force are produced by a series of coils which may themselves be helical in shape. Unlike tokamaks, stellarators do not require a toroidal current to be induced in the plasma. RFP devices have the same toroidal and poloidal components as a tokamak, but the current flowing through the plasma is much stronger and the direction of the toroidal field within the plasma is reversed.

In tokamaks and RFP devices, the current flowing through the plasma also serves to heat it to a temperature of about 10 million degrees Celsius. Beyond that, additional heating systems are needed to achieve the temperatures necessary for fusion. In stellarators, these heating systems have to supply all the energy needed.

The tokamak *(toroidalnya kamera ee magnetnaya katushka*–torus-shaped magnetic chamber) was designed in 1951 by Soviet physicists Andrei Sakharov and Igor Tamm. Tokamaks operate within limited parameters outside which sudden losses of energy confinement (disruptions) can occur, causing major thermal and mechanical stresses to the structure and

walls. Nevertheless, it is considered the most promising design, and research is continuing on various tokamaks around the world.

Research is also being carried out on several types of stellarator. Lyman Spitzer devised and began work on the first fusion device – a stellarator–at the Princeton Plasma Physics Laboratory in 1951. Due to the difficulty in confining plasmas, stellarators fell out of favour until computer modelling techniques allowed accurate geometries to be calculated. Because stellarators have no toroidal plasma current, plasma stability is increased compared with tokamaks. Since the burning plasma can be more easily controlled and monitored, stellerators have an intrinsic potential for steady-state, continuous operation. The disadvantage is that, due to their more complex shape, stellarators are much more complex than tokamaks to design and build.

RFP devices differ from tokamaks mainly in the spatial distribution of the toroidal magnetic field, which changes sign at the edge of the plasma. The RFX machine in Padua, Italy is used to study the physical problems arising from the spontaneous reorganisation of the magnetic field, which is an intrinsic feature of this configuration.

IV. Inertial confinement

In inertial confinement fusion, which is a newer line of research, laser or ion beams are focused very precisely onto the surface of a target, which is a pellet of D-T fuel, a few millimetres in diameter. This heats the outer layer of the material, which explodes outwards generating an inward-moving compression front or implosion that compresses and heats the inner layers of material. The core of the fuel may be compressed to one thousand times its liquid density, resulting in conditions where fusion can occur. The energy released then would heat the surrounding fuel, which may also undergo fusion leading to a chain reaction (known as ignition) as the reaction spreads outwards through the fuel. The time required for these reactions to occur is limited by the inertia of the fuel (hence the name), but is less than a microsecond. So far, most inertial confinement work has involved lasers.

Recent work at Osaka University's Institute of Laser Engineering in Japan suggests that ignition may be achieved at lower temperature with a second very intense laser pulse guided through a millimetre-high gold cone into the compressed fuel, and timed to coincide with the peak compression. This technique, known as 'fast ignition', means that fuel compression is separated from hot spot generation with ignition, making the process more practical.

A completely different concept, the 'Z-pinch' (or 'zeta pinch'), uses a strong electrical current in a plasma to generate X-rays, which compress a tiny D-T fuel cylinder.

V. Magnetized target fusion

Magnetized target fusion (MTF), also referred to as magneto-inertial fusion (MIF), is a pulsed approach to fusion that combines the compressional heating of inertial confinement fusion with the magnetically reduced thermal transport and magnetically enhanced alpha heating of magnetic confinement fusion.

A range of MTF systems are currently being experimented with, and commonly use a magnetic field to confine a plasma with compressional heating provided by laser, electromagnetic or mechanical liner implosion. As a result of this combined approach, shorter plasma confinement times are required than for magnetic confinement (from 100 ns to 1 ms, depending on the MIF approach), reducing the requirement to stabilize the plasma for long periods. Conversely, compression can be achieved over timescales longer than those typical for inertial confinement, making it possible to achieve compression through mechanical, magnetic, chemical, or relatively low-powered laser drivers.

Several approaches are underway to examine MTF, including experiments at Los Alamos National Laboratory, Sandia National Laboratory, the University of Rochester, and private companies General Fusion and Helion Energy.

R&D challenges for MTF include whether a suitable target plasma can be formed and heated to fusion conditions while avoiding contamination from the liner, as with magnetic confinement and inertial confinement. Due to the reduced demands on confinement time and compression velocities, MTF has been pursued as a lower-cost and simpler approach to investigating these challenges than conventional fusion projects.

VI. Hybrid fusion

Fusion can also be combined with fission in what is referred to as hybrid nuclear fusion where the blanket surrounding the core is a subcritical fission reactor. The fusion reaction acts as a source of neutrons for the surrounding blanket, where these neutrons are captured, resulting in fission reactions taking place. These fission reactions would also produce more neutrons, thereby assisting further fission reactions in the blanket.

The concept of hybrid fusion can be compared with an accelerator-driven system (ADS), where an accelerator is the source of neutrons for the blanket assembly, rather than nuclear fusion reactions (see page on **Accelerator-driven Nuclear Energy**). The blanket of a hybrid fusion system can therefore contain the same fuel as an ADS – for example, the abundant element thorium or the long-lived heavy isotopes present in used nuclear fuel (from a conventional reactor) could be used as fuel.

The blanket containing fission fuel in a hybrid fusion system would not require the development of new materials capable of withstanding constant neutron bombardment, whereas such materials would be needed in the blanket of a 'conventional' fusion system. A further advantage of a hybrid system is that the fusion part would not need to produce as many neutrons as a (non-hybrid) fusion reactor would in order to generate more power than is consumed – so a commercial-scale fusion reactor in a hybrid system does not need to be as large as a fusion-only reactor.

VII. Fusion research

A long-standing quip about fusion points out that, since the 1970s, commercial deployment of fusion power has always been about 40 years away. While there is some truth in this, many breakthroughs have been made, particularly in recent years, and there are a number of major projects under development that may bring research to the point where fusion power can be commercialized.

Several **tokamaks** have been built, including the Joint European Torus (JET) and the Mega Amp Spherical Tokamak (MAST)in the UK and the tokamak fusion test reactor (TFTR) at Princeton in the USA. The ITER (International Thermonuclear Experimental Reactor) project currently under construction in Cadarache, France will be the largest tokamak when it operates in the 2020s. The Chinese Fusion Engineering Test Reactor (CFETR) is a tokamak which is reported to be larger than ITER, and due for completion in 2030.Meanwhile it is running its Experimental Advanced Superconducting Tokamak (EAST).

Much research has also been carried out on **stellarators**. A large one of these, the Large Helical Device at Japan's National Institute of Fusion Research, began operating in 1998. It is being used to study the best magnetic configuration for plasma confinement. At the Garching site of the Max Planck Institute for Plasma Physics in Germany, research carried out at the Wendelstein 7-AS between 1988 and 2002 is being progressed at the Wendelstein 7-X, which was built over 19 years at Max Planck Institute's Greifswald site and started up at the end of 2015.Another stellarator, TJII, is in operation in Madrid, Spain. In the USA, at Princeton Plasma Physics Laboratory, where the first stellarators were built in 1951, construction on the NCSX stellerator was abandoned in 2008 due to cost overruns and lack of funding[2].

There have also been significant developments in research into **inertial fusion energy** (IFE). Construction of the $7 billion National Ignition Facility (NIF) at the Lawrence Livermore National Laboratory (LLNL), funded by the National Nuclear Security Administration, was completed in March 2009. The Laser Mégajoule (LMJ) in France's Bordeaux region started operation in October 2014. Both are designed to deliver, in a few billionths of a second, nearly two million joules of light energy to targets measuring a few millimeters in size. The main purpose of both NIF and LMJ is for research to support both countries' respective nuclear weapons programs.

VIII. ITER

In 1985, the Soviet Union suggested building a next generation tokamak with Europe, Japan and the USA. Collaboration was established under the auspices of the International Atomic Energy Agency (IAEA). Between 1988 and 1990, the initial designs were drawn up for an International Thermonuclear Experimental Reactor (ITER, which also means 'a path' or 'journey' in Latin) with the aim of proving that fusion could produce useful energy. The four parties agreed in 1992 to collaborate further on engineering design activities for ITER. Canada and Kazakhstan are also involved through Euratom and Russia, respectively.

Six years later, the ITER Council approved the first comprehensive design of a fusion reactor based on well-established physics and technology with a price tag of $6 billion. Then the USA decided pull out of the project, forcing a 50% reduction in costs and a redesign. The result was the ITER Fusion Energy Advanced Tokomak (ITER-FEAT) – initially expected to cost $3 billion but still achieve the targets of a self-sustaining reaction and a net energy gain. The envisaged energy gain is unlikely to be enough for a power plant, but it should demonstrate feasibility.

In 2003, the USA rejoined the project and China also announced it would join. After deadlocked discussion, the six partners agreed in mid-2005 to site ITER at Cadarache, in southern France. The deal involved major concessions to Japan, which had put forward Rokkasho as a preferred site. The European

Union (EU) and France would contribute half of the then estimated €12.8 billion total cost, with the other partners – Japan, China, South Korea, USA and Russia – putting in 10% each. Japan will provide a lot of the high-tech components, will host a €1 billion materials testing facility – the International Fusion Materials Irradiation Facility (IFMIF) – and will have the right to host a subsequent demonstration fusion reactor. India became the seventh member of the ITER consortium at the end of 2005. In November 2006, the seven members – China, India, Japan, Russia, South Korea, the USA and the European Union – signed the ITER implementing agreement. The total cost of the 500 MW ITER comprises about half for the ten-year construction and half for 20 years of operation.

Site preparation works at Cadarache commenced in January 2007. First concrete for the buildings was poured in December 2013. Experiments were due to begin in 2018, when hydrogen will be used to avoid activating the magnets, but this is now expected in 2025. The first D-T plasma is not expected until 2035.ITER is large because confinement time increases with the cube of machine size. The vacuum vessel will be 19 m across and 11 m high, and weigh more than 5000 tonnes.

The goal of ITER is to operate with a plasma thermal output of 500 MW (for at least 400 seconds continuously) with less than 50 MW of plasma heating power input. No electricity will be generated at ITER.

An associated CEA facility at Cadarache is WEST, formerly Tore Supra, which is designed to test prototype components and accelerate their development for ITER. It is focused on the divertor structure to remove helium, testing the durability of tungsten materials used.

A 2 GW Demonstration Power Plant, known as Demo, is expected to demonstrate large-scale production of electrical power on a continual basis. The conceptual design of Demo was expected to be completed by 2017, with construction beginning around 2024 and the first phase of operation commencing from 2033. It has since been delayed, with construction now planned for after 2040.

IX. JET

In 1978, the European Community (Euratom, along with Sweden and Switzerland) launched the Joint European Torus (JET) project in the UK. JET is the largest tokamak operating in the world today. A similar tokamak, JT-60, operates at the Naka Fusion Institute of Japan Atomic Energy Agency in Japan, but only JET has the facilities to use D-T fuel.

Following a legal dispute with Euratom, in December 1999 JET's international contract ended and the United Kingdom Atomic Energy Authority (UKAEA) took over the management of JET on behalf of its European partners. From that time JET's experimental programme has been co-ordinated by the European Fusion Development Agreement (EFDA) parties.[c]

JET produced its first plasma in 1983, and became the first experiment to produce controlled fusion power in November 1991, albeit with high input of electricity. Up to 16 MW of fusion power for one second and 5 MW sustained has been achieved in D-T plasmas using the device, from 24 MW of power injected into its heating system, and many experiments are conducted to study different heating schemes and other techniques. JET has been very successful in operating remote handling techniques in a radioactive environment to modify the interior of the device and has shown that the remote handling maintenance of fusion devices is realistic.

JET is a key device in preparations for ITER. It has been significantly upgraded in recent years to test ITER plasma physics and engineering systems. Further enhancements are planned at JET with a view to exceeding its fusion power record in future D-T experiments. A compact device – Mega Amp Spherical Tokamak (MAST) – is also being developed alongside JET, partly to serve the ITER project.

X. KSTAR

The KSTAR (Korean Superconducting Tokamak Reactor) at the National Fusion Research Institute (NFRI) in Daejeon produced its first plasma in mid-2008. It is a pilot device for ITER, and involves much international collaboration. It will be a satellite of ITER during ITER's operational phase from the early 2020s. The tokamak with 1.8 metre major radius is the first to use Nb3Sn superconducting magnets, the same material to be used in the ITER project. Its first stage of development to 2012 was to prove baseline operation technologies and achieved plasma pulses of up to 20 seconds. For the second phase of development (2013-2017), KSTAR was upgraded to study long pulses of up to 300 seconds in H mode–the 100s target was in 2015–and embark upon high-performance AT mode. It achieved 70 seconds in high-performance plasma operation in late 2016, a world record. In addition, KSTAR researchers also succeeded in achieving an alternative advanced plasma operation

mode with the internal transport barrier (ITB). This is a steep pressure gradient in the core of the plasmas due to the enhanced core plasma confinement. NFRI said this is the first ITB operation achieved in the superconducting device at the lowest heating power. KSTAR Phase 3 (2018-2023) is to develop high performance, high efficiency AT mode technologies with long-pulse operation. Phase 4 (2023-2025) will test DEMO-related prior arts. The device does not have tritium handling capabilities, so will not use D-T fuel.

XI. K-DEMO TOKAMAK

In collaboration with the US Department of Energy's Princeton Plasma Physics Laboratory (PPPL) in New Jersey and South Korea's National Fusion Research Institute (NFRI) K-DEMO is intended to be the next step toward commercial reactors from ITER, and would be the first plant to actually contribute power to an electric grid. According to the PPPL, it would generate "some 1 billion watts of power for several weeks on end", a much greater output than ITER's goal of producing 500 million watts for 500 seconds by the late 2020s. K-DEMO is expected to have a 6.65m diameter major radius tokamak, and a test blanket module as part of the DEMO breeding blanket R&D. The Ministry of Education, Science and Technology plans to invest about KRW 1 trillion (US$ 941 million) in the project. About KRW 300 billion of that spending has already been funded. The government expects the project to employ nearly 2,400 people in the first phase, which will last throughout 2016. K-DEMO is expected to have an initial operational phase from about 2037 to 2050 to develop components for the second stage, which would produce electricity.

XII. EAST

In China the Experimental Advanced Superconducting Tokamak (EAST) at China Academy of Sciences Hefei Institute of Physical Science (CASHIPS) produced hydrogen plasma at 50 million degrees Celsius and held it for 102 seconds in 2017. In November 2018 it achieved 100 million degrees Celsius for 10 seconds, with input of 10 MW of electric power.

A. TFTR

In the USA, the Tokamak Fusion Test Reactor (TFTR) operated at the Princeton Plasma Physics Laboratory (PPPL) from 1982 to 1997.[d] In December 1993, TFTR became the first magnetic fusion device to perform extensive experiments with plasmas composed of D-T. The following year TFTR produced 10.7 MW of controlled fusion power – a record at that time. TFTR set other records, including the achievement of a plasma temperature of 510 million degrees centigrade in 1995. However, it did not achieve its goal of break-even fusion energy (where the energy input required is no greater than the amount of fusion energy produced), but achieved all of its hardware design goals, thus making substantial contributions to the development of ITER.

B. ALCATOR

At the Massachusetts Institute of Technology (MIT) since the 1970s a succession of small ALCATOR (Alto Campus Torus) high magnetic field torus reactors have operated on the principle of achieving high plasma pressure as the route to long plasma confinement. Alcator C-Mod is claimed to have the highest magnetic field and highest plasma pressure of any fusion reactor, and is the largest university-based fusion reactor in the world. It operated 1993-2016. In September 2016 it achieved a plasma pressure of 2.05 atmospheres at a temperature of 35 million degrees Celsius. The plasma produced 300 trillion fusion reactions per second and had a central magnetic field strength of 5.7 tesla. It carried 1.4 million amps of electrical current and was heated with over 4 MW of power. The reaction occurred in a volume of approximately 1 cubic metre and the plasma lasted for two seconds. Having achieved this record performance for a fusion reactor, government funding ceased.

A scaled-up version planned to be built at Triotsk near Moscow in collaboration with the Kurchatov Institute is Ignitor, with 1.3 m diameter torus.

XIII. LARGE HELICAL DEVICE– STELLARATOR

The Large Helical Device (LHD) at Japan's National Institute for Fusion Science in Toki, in the Gifu Prefecture, was the world's largest stellarator. LHD produced its first plasma in 1998 and has demonstrated plasma confinement properties comparable to other large fusion devices. It has achieved an ion temperature of 13.5 keV (about 160 million degrees) and plasma stored energy of 1.44 million joules (MJ).

XIV. WENDELSTEIN 7-X STELLARATOR

Following a year of tests, this started up at the end of 2015, and helium plasma briefly reached about one million degrees centigrade.In 2016 it progressed to using hydrogen, and using 2 MW it achieved plasma of 80 million degrees centigrade for a quarter of a second. W7-X is the world's largest stellarator and it is planned to operate continuously for up to 30 minutes. It cost €1 billion ($1.1 billion).

XV. HELIAC-1 STELLARATOR

At the **Australian Plasma Fusion Research Facility**at the Australian National University the H-1 stellarator has run for some years and in 2014 was upgraded significantly. H-1 is capable of accessing a wide range of plasma configurations and allows exploration of ideas for improved magnetic design of the fusion power stations that will follow ITER.

XVI. NATIONALIGNITIONFACILITY– LASER

The world's most powerful laser fusion facility, the $4 billion National Ignition Facility (NIF) at Lawrence Livermore National Laboratory (LLNL), was completed in March 2009. Using its 192 laser beams, NIF is able to deliver more than 60 times the energy of any previous laser system to its target[e]. LLNL announced in July 2012 that in "an historic record-breaking laser shot, the NIF laser system of **192 beams delivered more than 500 TW of peak power and 1.85 megajoules**(MJ) of ultraviolet laser light to its (2mm diameter) target" for a few trillionths of a second. It was reported that in September 2013 at NIF for the first time the amount of energy released through the fusion reaction exceeded the amount of energy being absorbed by the fuel, but not the amount supplied by the giant lasers. Publication of this in 2014 said 17 kJ was released.

An earlier high-power laser at LLNL, Nova, was built in 1984 for the purpose of achieving ignition. Nova failed to do this and was closed in 1999, but provided essential data that led to the design of NIF. Nova also generated considerable amounts of data on high-density matter physics, which is useful both in fusion power and nuclear weapons research.

In connection with NIF, LLNL is developing the Laser Inertial Fusion Engine (LIFE), a hybrid fusion system where neutrons resulting from laser fusion would drive a subcritical nuclear fission blanket to generate electricity. The blanket would contain either depleted uranium; used nuclear fuel; natural uranium or thorium; or plutonium-239, minor actinides and fission products from reprocessed used nuclear fuel[4].

XVII. LASER MÉGAJOULE

Meanwhile, the French Atomic Energy Commission (Commissariat à l'énergieatomique, CEA) has operated a similar size laser – the Laser Mégajoule (LMJ) – near Bordeaux since 2014. Its 240 laser beams are able to generate 1.8 MJ pulses for a few billionths of a second, concentrated on a small deuterium and tritium target. A prototype laser, the Ligned'Integration Laser (LIL), commenced operation in 2003.

XVIII. SG-II

China's National Laboratory of High-Power Laser and Physics, associated with the China Academy of Science, has a laser inertial confinement experiment in Shanghai – the Shenguang-II eight-beam laser facility (SG-II), similar to the National Ignition Facility in the USA and Laser Mégajoule in France. It is the only high power neodymium-glass solid laser facility with an active probe light in China. In 2005 a ninth beam was added, advancing the capacity for fusion research. The SG-II facility is China's high-power laser technology international demonstration base.

XIX. PETAL AND HIPER LASERS

The Petawatt Aquitaine Laser (PETAL) laser facility is a high energy multi-petawatt laser (3.5 kJ energy with a duration of 0.5 to 5 ps) under construction near Bordeaux, on the same site as LIL. PETAL will be coupled with LIL to demonstrate the physics and laser technology of fast ignition. First experiments are expected in 2012.

The High Power Laser Energy Research Facility (HiPER) is being designed to build on the research planned at the PETAL project. HiPER will use a long pulse laser (currently estimated at 200kJ) combined with a 70kJ short pulse laser. A three-year preparatory phase that commenced in 2008 has direct funding or in-kind commitments amounting to around €70 million from several countries. The detailed engineering phase is projected to begin in 2011, with a six-year construction phase possibly commencing by 2014.

XX. Z MACHINE

Operated by Sandia National Laboratories, the Z machine is the largest X-ray generator in the world. As with NIF, the facility was built as part of the country's Stockpile Stewardship Program, which aims to maintain the stockpile of nuclear weapons without the need for full-scale testing.

Conditions for fusion are achieved by passing a powerful electrical pulse[f](lasting less than 100 nanoseconds) through a set of fine tungsten wires inside a metal hohlraum[g]. The wires turn into a plasma and experience a compression ('Z-pinch'), forcing the vapourized particles to collide with each other, thereby producing intense X-ray radiation. A tiny cylinder containing fusion fuel placed inside the hohlraum would therefore be compressed by the X-rays, allowing fusion to occur.

In 2006, Z machine had achieved temperatures of over 2 billion degrees[6],considerably higher than what

is needed for fusion, and in theory high enough to allow nuclear fusion of hydrogen with heavier elements such as lithium or boron.

XXI. OTHER FUSION PROJECTS

There is a considerable amount of research into many other fusion projects at various stages of development.

Lockheed CFR. Lockheed Martin at its so-called 'skunk works' is developing a **Compact Fusion Reactor**(CFR) which uses conventional D-T plasma in evacuated containment but confines it differently. Instead of constraining the plasma within tubular rings, a series of superconducting coils will generate a new magnetic-field geometry in which the plasma is held within the broader confines of the entire reaction chamber. The energy is supplied by radiofrequency heating. Superconducting magnets within the coils will generate a magnetic field around the outer border of the chamber. The aim is to go to plasma pressure being as great as confining pressure at high enough temperature for ignition and net energy yield. Heat exchangers in the reactor wall would convey energy to a gas turbine. It has progressed to a magentised ion confinement experiment, but has some way to go before any prototype, which they claim will be very much smaller than conventional designs such as the ITER tokamak.

Italy's National Agency for New Technologies, Energy and Sustainable Economic Development (ENEA) is developing a small tokamak reactor by the name of **Ignitor**. Under an Italian-Russian agreement signed in May 2010, a reactor will be assembled in Italy and installed at the Kurchatov Institute's Troitsk Institute for Innovation and Fusion Research (TRINITI) near Moscow[7].

An alternative to using powerful lasers for inertial confinement fusion is '**heavy ion fusion**', where high-energy particles from an accelerator are focused using magnetic fields onto the fusion target. The **NDCX-II**(Neutralized Drift Compression Experiment II) accelerator has been used for heavy ion fusion experiments since 2012 at Lawrence Berkeley National Laboratory. It is being expanded to deliver short intense pulses of ion beams with kinetic energy of 1.2 MeV. High-energy-density physics (HEDP) experiments with laboratory plasmas is a growing part of inertial fusion energy (IFE) physics.

LPP Fusion (Lawrenceville Plasma Physics) is a US enterprise developing aleuronic fusion using a dense plasma focus device (DPF or **focus fusion**) and hydrogen-boron fuel. The hydrogen and boron (B-11) as plasma fuse at high temperature to form a pulsed beam of helium nuclei without emitting neutrons. (The boron and hydrogen combine to produce a brief intermediate carbon-12 atom which rapidly decays to three alpha particles.) This charged high-energy ion beam generates electricity as it passes through a series of coils similar to a transformer, at 80% efficiency. The balance of energy is as by-product X-rays which are captured in an array of photoelectric receptors. LPP Fusion has achieved electron energies of 400 keV.

Another line of fusion research using lasers also involves fusing hydrogen and boron-11 (HB11) to produce helium nuclei, which continue the chain reaction from boron. One laser generates a powerful magnetic confinement field in a coil to trap the fusion reaction in a small area for a nanosecond, while a second more powerful laser triggers the nuclear fusion process. Early HB11 fusion trials at the Prague Asterix Laser System, using high-energy iodine lasers, have generated more energy than needed to trigger the fusion process.

The **Polywell**('polyhedron' combined with 'potential well')deviceconsists of magnetic coils arranged in a polyhedral configuration of six sides, forming a cube. A cloud of electrons is confined in the middle of the device so as to be able to accelerate and confine the positive ions to be fused. This electrostatic confinement concept differs from traditional magnetic confinement because the fields do not need to confine ions– only electrons. **EMC2 Fusion Development Corporation** has been researching the Polywell concept and looking at hydrogen and boron as fuel for aneutronic fusion. This followed some years of development by the US Navy, using deuterium fuel.

General Fusionis one of a number of private efforts to develop a commercial fusion power plant. The company's magnetized target fusion (MTF) approach generates a compact toroid plasma in an injector, containing and compressing it using a magnetic field before injecting it into a spherical compression chamber. The chamber holds a liquid lead-lithium liner which is pumped to create a vortex, into which the plasma target is injected. A synchronized array of pistons firing simultaneously creates a spherical compression wave in the liquid metal, compressing the plasma target and heating it to fusion conditions. Founded in Canada in 2002, General Fusion is funded by a syndicate of private investors, energy venture capital companies, sovereign wealth funds and the Canadian government's Sustainable Development Technology Canada (SDTC) fund. A

further government grant was announced in October 2018, from the Strategic Innovation Fund. The company has demonstrated milestones including creating 200-300 eV magnetized spheromak plasmas and confining them for over 500 μs.

Much of current work underway on MTF is derived from programmes at the Soviet Kurchatov Institute of Atomic Energy, under E. P. Velikhov, circa 1970. This inspired the LINUS project at the Naval Research Laboratory in the USA, and later the fast-liner project at Los Alamos.

Tokamak Energyin the UK is a private company developing a spherical tokamak, and hopes to commercialize this by 2030. The company grew out of Culham laboratory, home to JET, and its technology revolves around high temperature superconducting (HTS) magnets, which allow for relatively low-power and small-size devices, but high performance and potentially widespread commercial deployment. Its first tokamak with exclusively HTS magnets – the ST25 HTS, Tokamak Energy's second reactor – demonstrated 29 hours' continuous plasma during the Royal Society Summer Science Exhibition in London in 2015, a world record. The next reactor is the ST40 at Milton Park in Oxfordshire, which achieved first plasma in April 2017. It isexpected to produce plasma temperatures of 15 million degrees Celsius– hotter than the centre of the Sun in 2017 after the commissioning of further magnetic coils. "The ST40 is designed to achieve 100 million degrees C and get within a factor of ten of energy break-even conditions. To get even closer to break-even point, the plasma density, temperature and confinement time then need to be fine-tuned." The company is working with Princeton Plasma Physics Laboratory on spherical tokamaks, and with the Plasma Science and Fusion Centre at MIT on HTS magnets. It aims to achieve commercial scale fusion power by 2030.

XXII. COLD FUSION

In March 1989, spectacular claims were made for another approach, when two researchers, in the USA (Stanley Pons) and the UK (Martin Fleischmann), claimed to have achieved fusion in a simple tabletop apparatus working at room temperature. 'N-Fusion', or 'cold fusion', involves the electrolysis of heavy water using palladium electrodes on which deuterium nuclei are said to concentrate at very high densities. The researchers claimed that heat – which could only be explained in terms of nuclear processes – was produced, as well as fusion byproducts, including helium, tritium and neutrons. Other experimenters failed to replicate this, however, and most of the scientific community no longer considers it a real phenomenon.

XXIII. LOW-ENERGY NUCLEAR REACTIONS (LENR)

Initiated by claims for 'cold fusion', research at the nanotechnology level is continuing on low-energy nuclear reactions (LENR) which apparently use weak nuclear interactions (rather than strong force as in nuclear fission or fusion) to create low-energy neutrons, followed by neutron capture processes resulting in isotopic change or transmutation, without the emission of strong prompt radiation. LENR experiments involve hydrogen or deuterium permeation through a catalytic layer and reaction with a metal. Researchers report that energy is released, though on any reproducible basis, very little more than is input. The main practical example is hydrogen plus nickel powder evidently giving more heat than can be explained on any chemical basis.

The Japanese government is sponsoring LENR research– notably a nano-metal hydrogen energy project (MHE)– through its New Energy and Industrial Technology Development Organization (NEDO), and Mitsubishi is also active in research.

XXIV. CONCLUSION

The use of fusion power plants could substantially reduce the environmental impacts of increasing world electricity demands since, like nuclear fission power, they would not contribute to acid rain or the greenhouse effect. Fusion power could easily satisfy the energy needs associated with continued economic growth, given the ready availability of fuels. There would be no danger of a runaway fusion reaction as this is intrinsically impossible and any malfunction would result in a rapid shutdown of the plant.

However, although fusion does not generate long-lived radioactive products and the unburned gases can be treated on site, there would a short- to medium-term radioactive waste problem due to activation of the structural materials. Some component materials will become radioactive during the lifetime of a reactor, due to bombardment with high-energy neutrons, and will eventually become radioactive waste. The volume of such waste would be similar to the corresponding volumes from fission reactors. However, the long-term radiotoxicity of the fusion wastes would be considerably lower than that from actinides in used fission fuel, and the activation product wastes would be

handled in much the same way as those from fission reactors with some years of operation.[8].

There are also other concerns, principally regarding the possible release of tritium into the environment. It is radioactive and very difficult to contain since it can penetrate concrete, rubber and some grades of steel. As an isotope of hydrogen, it is easily incorporated into water, making the water itself weakly radioactive. With a half-life of about 12.3 years, the presence of tritium remains a threat to health for about 125 years after it is created, as a gas or in water, if at high levels. It can be inhaled, absorbed through the skin or ingested. Inhaled tritium spreads throughout the soft tissues and tritiated water mixes quickly with all the water in the body. Although there is only a small inventory of tritium in a fusion reactor – a few grams – each could conceivably release significant quantities of tritium during operation through routine leaks, assuming the best containment systems. An accident could release even more. This is one reason why long-term hopes are for the deuterium-deuterium fusion process, dispensing with tritium.

While fusion power clearly has much to offer when the technology is eventually developed, the problems associated with it also need to be addressed if it is to become a widely used future energy source.

REFERENCES

[1] Fusion Research: An Energy Option for Europe's Future, Directorate-General for Research, European Commission, 2007 (ISBN: 9279005138)

[2] Statement by Dr. Raymond L. Orbach, Under Secretary for Science and Director, Office of Science, US Department of Energy (22 May 2008)

[3] National Ignition Facility achieves unprecedented 1 megajoule laser shot, Lawrence Livermore National Laboratory news release (27 January 2010)

[4] LIFE: Clean Energy from Nuclear Waste page on Lawrence Livermore National Laboratory website (www.llnl.gov)

[5] Z produces fusion neutrons, Sandia scientists confirm, Sandia National Laboratories news release (7 April 2003)

[6] Sandia's Z machine exceeds two billion degrees Kelvin, Sandia National Laboratories news release (8 March 2006)

[7] New project aims for fusion ignition, Massachusetts Institute of Technology news (10 May 2010) New record for fusion, MIT News (14 October 2016)

[8] Safety and Environmental Impact of Fusion, I. Cook, G. Marbach, L. Di Pace, C. Girard, N. P. Taylor, EUR (01) CCE-FU / FTC 8/5 (April 2001)

Research on Improved Adaptive ViBe Algorithm For Vehicle Detection

Kun Jiang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: 416578078@qq.com

Jianguo Wang
School of Computer Science and Engineering
Xi'an Technological University
Xi'an, China
E-mail: wjg_xit@126.com

***Abstract*—Vehicle detection is an important step in vehicle tracking and recognition in video environment. Vibe algorithm is a moving target detection algorithm based on background difference method. Based on the traditional ViBe algorithm, this paper introduces three-frame difference method combined with ViBe algorithm to speed up the elimination of ghosts, and proposes an adaptive Vibe algorithm, which defines two kinds of vehicle detection errors and their corresponding error functions. Then, according to the range of these two errors, a set of reasonable judgment methods are determined to adjust the unreasonable threshold, which ensures the adaptive updating of the background model. It improves the environmental adaptability of vehicle detection and ensures higher accuracy of vehicle detection under complex illumination conditions.**

Keywords-Vehicle Detection; Background Difference Method; Vibe Algorithm; Three-Frame Difference Algorithm

I. INTRODUCTION

Vehicle detection is the key step of video vehicle recognition, which aims to obtain the location of vehicle for further recognition. For each pixel, its background can usually be built using a model. At present, there are three methods for moving object detection: optical flow method[1], frame difference method[2], background subtraction method[3]. Based on the motion vector of pixels, the optical flow method can detect and track the target, but it has a large amount of computation and poor real-time performance. Moreover, the method lacks sensitivity to noise, illumination change and background interference. Frame difference method detects moving objects according to the difference between two or three consecutive frames. It has strong adaptability to the background change, but it does not perform well in detecting the contour of moving objects. In addition, it is very sensitive to the speed of moving objects, so it cannot effectively detect slow moving objects; background difference method is a commonly used moving object detection algorithm. The main idea is to make a distinction between each frame and background model to build the background model and get the moving foreground objects. Background difference method has the ability to adapt to the scene changes in the video background, but if the background model contains foreground objects, it may generate ghosting.

Background difference method is one of the most widely used vehicle detection methods because of its fast and accurate. Traditional background subtraction algorithms include Gaussian mixture model (GMM)[4] and codebook[5]. GMM method is simple and low cost. However, the initialization time is too long and the algorithm is complex to meet the real-time requirements. Cookbook has the advantages of dealing with time fluctuations well, but its memory consumption is quite high. In 2011, Olivier barnechand

DOI: 10.21307/ijanmc-2019-065

Marc van droogenbroeck proposed the vibe algorithm [6]. Because the algorithm only needs the first frame to complete the model initialization, it can meet the real-time requirements compared with GMM model. In addition, in the process of execution, the algorithm only needs to record the corresponding sample set for each pixel, so it has smaller memory consumption. In addition, vibe algorithm has goodanti noise ability. Although the foreground object can be mixed in pixel initialization, it will produce ghost phenomenon.

In this paper, an improved adaptive algorithm of vibe is proposed, and a moving target detection method based on three frame difference method is introduced. Through experimental verification, the algorithm proposed in this paper effectively solves the problems of "ghost" existing in traditional VIB algorithm and insufficient adaptability to complex light environment. It has the advantages of simple algorithm, good real-time performance and high detection accuracy.

II. VIBE BACKGROUND MODELING

The vibe background modeling algorithm was proposed by Olivier barnech et al in 2011.[7] can be used for fast background extraction and moving object detection. Vibe algorithm uses two mechanisms of random selection and neighborhood propagation to build and update the background model, which includes three steps: background modeling and initialization, foreground detection and background model update. In this paper, based on the traditional vibe algorithm, an adaptive background model of vibe is added. According to the range of vehicle detection error, a set of judgment method is determined to evaluate the rationality of the current threshold. When the threshold is not reasonable, adjust according to a certain step to ensure that the background model is updated automatically, and finally get more accurate vehicle detection results.

Initialization of background model

v_x represents the pixel value at point x, and each pixel builds the number of background sample sets N ($N = 20$):

$$M(x) = \{v_1, v_2, \ldots, v_N\} \tag{1}$$

As shown in Fig. 1, the gray space makes the region centered on X, the radius of the gray space $S_R(v(x))$ is R, and the threshold $\#\min$ is set($\#\min = 2$). Then find the intersection of $M(x)$ and $S_R(v(x))$, C is the total number of elements in the intersection:

$$C = \#M(x)\left[\{S_R(v(x)) \cap \{v_1, v_2 \ldots, v_n\}\}\right] \tag{2}$$

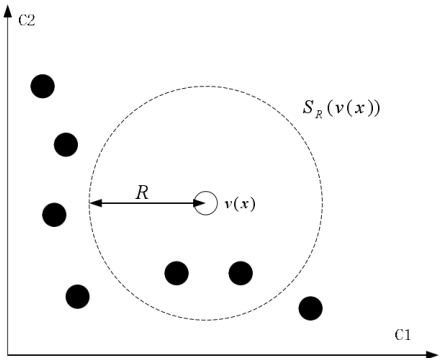


Figure 1. ViBe background model.

The initialization of vibe model only needs the first image, but only one image does not contain the spatial information of pixels. According to the similar spatial characteristics of the close pixels, the sample set is filled with the approximate pixels. When the first image is input, the background model of the pixels in the image is as follows (3):

$$M_0(x) = v_0\{y \mid y \in N_G(x)\} \tag{3}$$

Where, $N_G(x)$ represents the neighborhood pixel; v_0 is the currently selected pixel. The probability that the pixels in $N_G(x)$ are selected in N initialization is $(1-N)/N$.

A. Foreground detection

After initialization, vehicle detection starts from the second image. Separating foreground target and background is the process of moving target detection. At time t, the pixel value of random pixel x is v_t, According to formula (2) C is judged according to formula (4):

$$v_t = \begin{cases} C < \#\min & (\text{foreground}) \\ C \geq \#\min & (\text{background}) \end{cases} \quad (4)$$

T ($T = 20$) is a preset threshold. When the number of times for the background is greater than the threshold $\#\min$, the pixel x are considered to belong to the background area; if not, it belongs to the foreground target area. According to formula (4), the binary image obtained after vehicle detection is the initial vehicle detection binary image.

B. Dynamic update of background model

In the process of updating the vibe background model, not only the relationship between the current pixel and its historical samples, but also the relationship between the current pixel and other pixels in other spatial neighborhood should be considered. In other words, the updating of vibe background model is a random process in both time and space.

In the background model in the previous frame, if the current pixel v_t is marked as background, its background model $M_t(x)$ is updated in time. If the current pixel is marked as vehicle, the model is not updated. This update strategy is called conservative update strategy. The method of updating the background model is to randomly select the sample m in the sample set $M_t(x)$.

The method to update the background model in space is to calculate the gradient amplitude of the current pixel v_t, If the gradient amplitude is greater than 50, the space update is not implemented. Otherwise, in the 8 neighborhood of the current pixel v_t, randomly select the pixel marked as v_t, in the background model $M_t'(x)$ of pixel v_t' randomly select a sample m_j, and the characteristic value of the current pixel v_t is assigned to m_j. If the current pixel v_t is at the edge of the image, it is randomly selected in its incomplete 8 neighborhood. The spatial update strategy ensures the continuity of spatial information in the background image.

III. ADAPTIVE VIBE BACKGROUND MODEL

In the vibe background model, threshold R represents the range of background eigenvalues (as shown in Figure 1). Threshold R has a great influence on vehicle detection results. If the fixed threshold value is greater than the expected threshold value, it should be the vehicle's area, which will lead to inaccurate detection of the vehicle area.

A. Error functions of vehicle

There are several situations of vehicle detection error: the detection background area is mistakenly

regarded as the vehicle area, or the vehicle area is mistakenly regarded as the background area. The former belongs to the error vehicle area; the latter is the error background area. The size of these two types of error areas will change with the change of threshold R. When the threshold R is very small, the fluctuation range of the sample eigenvalues in the vibe background model is also very small, which helps to improve the detection accuracy of the vehicle area. The noise area may be mistakenly regarded as the wrong vehicle area, and the small noise area can be removed by morphological method, but the large noise area will not be easily removed. Therefore, in order to prevent this situation, it is necessary to minimize the wrong vehicle area.

In theory, the wrong vehicle area is part of the background and it is static. Therefore, if the connected region does not overlap all the moving regions in the binary mapping of frame difference, the connected region is considered as the wrong vehicle region, and the error function of the wrong vehicle region can be defined as:

$$Err_1(R) = \frac{\sum_{i=1}^{n_1} a_1^i}{L \times W} \tag{5}$$

Where, a_1^i is the ith connected domain, which does not overlap all moving regions in the binary mapping of frame difference, and n_1 is the number of such connected domains. Is the total area of the wrong vehicle area, L and W represent the length and width of the image respectively, and their units are pixels. The higher the resolution of the image, the more accurate the value of $\sum_{i=1}^{n_1} a_1^i$. The wrong vehicle area function is defined as the ratio of the total area of the wrong vehicle area to the total area of the image.

When the threshold value R is large, the fluctuation range of the sample eigenvalues in the vibe background model is also large, which is conducive to improving the detection accuracy of the background region. The second type of error area is to detect the area originally belonging to the vehicle as the background area [9].

Firstly, error background area error a_2^i is defined, which represents the difference between the area of the smallest external rectangle containing the ith vehicle and the area of the same vehicle detected. s^i is the area of the smallest external rectangle containing the first vehicle, and v^i is the area of the ith vehicle detected.

$$a_2^i = s^i - v^i \tag{6}$$

Therefore, error background area error function can be further defined as:

$$Err_2(R) = \frac{\sum_{i=1}^{n_2} a_2^i}{n_2} \tag{7}$$

Where n_2 is the number of vehicles. $\sum_{i=1}^{n_1} a_2^i$ is the total area of the wrong vehicle area, take the average value to the error $Err_2(R)$ of the wrong background area.

From the above, the error $Err_1(R)$ of the error background area and the error $Err_2(R)$ of the error vehicle area are obtained. The total error of the error area can be defined as:

$$Err(R) = Err_1(R) + Err_2(R) \tag{8}$$

B. Adaptive adjustment of threshold

If the current threshold R is too small, the area of the area originally belonging to the background and mistakenly detected as the vehicle is too large, which means that the area $Err_1(R)$ of the wrong vehicle area is relatively large. If the current threshold is too large, the area $Err_2(R)$ of the area originally belonging to the vehicle and mistakenly detected as the background is too small, which means that the area of the wrong background is relatively large. According to this situation, we use the following adaptive scheme[10]:

$$\begin{cases} R = R - N, \ if \ \ Err(R) > T_2 \\ R = R + N, else \ if \ Err(R) > T_1 \ and \ Err(R) \le T_2 \\ R = R, \qquad else \end{cases} \quad (9)$$

T_1 and T_2 is the parameter to judge the rationality of threshold, and N is the adjustment step of threshold R. After a lot of experiments, the range of T_1 is 0.02 to 0.04,the range of T_2 is 0.13 to 0.26, the range of N is 1 to 3. A large number of experiments show that these values can ensure that the total error of the error background area and the error vehicle area proposed in this paper can be minimized.

IV. IMPROVED ADAPTIVE VIBE ALGORITHM OF THREE FRAME DIFFERENCE METHOD

This paper presents an improved adaptive vibe background modeling algorithm, which uses the vibe algorithm to model the background, adjusts the threshold adaptively, and then introduces the three frame difference method to improve. Vibe background model algorithm is based on the first image to establish the background model[10], but the traditional vibe algorithm will appear the phenomenon of "ghost". At present, many domestic and foreign literatures have carried out relevant research on the problem of "ghost". At present, the more commonly used method is to combine the traditional vibe algorithm with other algorithms, or to change the initialization of the first image of the original algorithm to a multi frame image Initialization.

The traditional vibe algorithm needs hundreds of frames to completely eliminate the "ghost" in the first frame. Using the improved adaptive vibe background modeling algorithm, the speed of eliminating the "ghost" is accelerated, and the "ghost" can be eliminated within dozens of frames. At the same time, the proposed adaptive vibe algorithm greatly improves the traditional vibe background modeling algorithm for complex light environment detection. After the introduction of three frame difference method, the speed of eliminating "ghost" is obviously speeded up, and the problem of "hole" existing in the three frame difference method itself is solved. Finally, the accuracy of moving object detection is improved by morphological processing of detection results. The flow chart of the improved vibe algorithm based on the three frame difference method is shown in Figure 2, and the specific implementation steps are as follows:

1) Input video image, and carry out image pre-processing such as graying and binarization.

2) Background modeling of three frame difference method and background modeling of vibe algorithm are carried out respectively for the image preprocessed. The final image is the "and" of the image calculated by the two methods.

3) Background modeling of three frame difference method and background modeling of vibe algorithm are carried out respectively for the image preprocessed. The final image is the "and" of the image calculated by the two methods.

4) Through the adaptive threshold adjustment algorithm proposed in this paper, the appropriate threshold is calculated to update the current background.

5) The updated image is processed by morphology to get the final detection results.

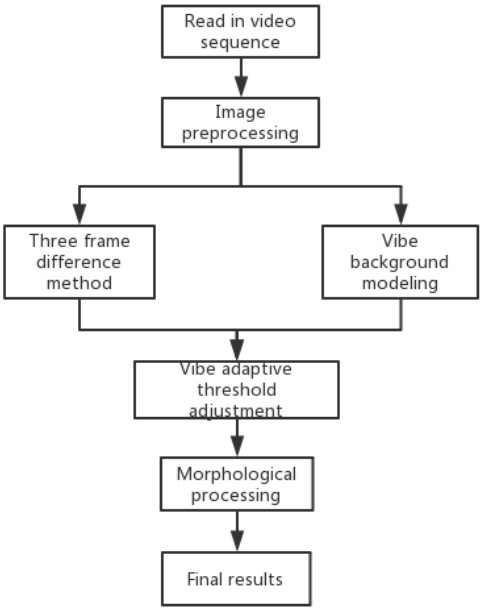


Figure 2. Improved ViBe algorithm of three frame difference method.

V. EXPERIMENTAL RESULT

Based on the above theory and processing flow, the algorithm is tested in the following environment: operating system: Microsoft Windows 10, experimental platform: Visual Studio 2019, CPU: intel5, RAM: 8g, third-party open source library: openCV 4.1.0. In order to verify that the method proposed in this paper can accurately detect moving objects in complex environment, the video selected in this experiment is the road monitoring video with more vehicles. This video is the traffic situation of a certain intersection at a certain time, with 650 frames in total, and the frame size is 640*480. An improved three frame difference algorithm based on vibe background modeling is used to detect moving vehicles. During the experiment, the 20th frame, 122 frame and 380 frame of video image sequence are randomly selected to analyze the detection effect.

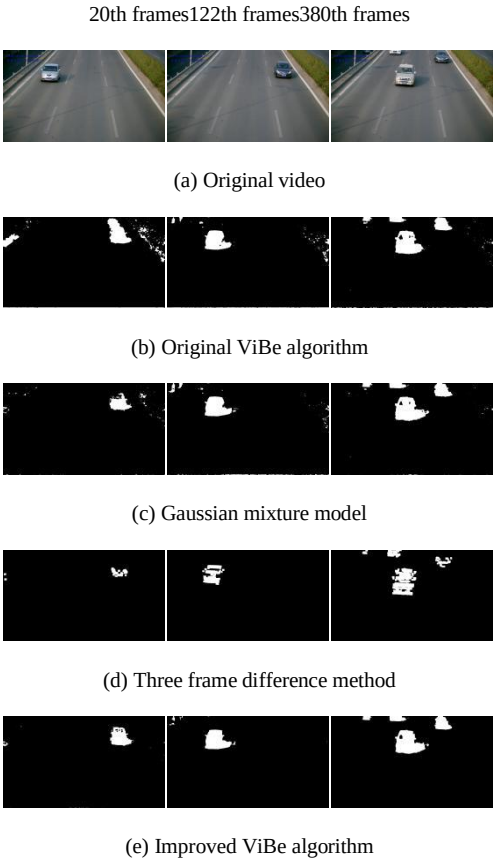


Figure 3. Comparison between the algorithm in this paperand the traditional target detection algorithm.

It can be seen from Figure 3 that the original vibe algorithm of frame 20 has obvious ghost phenomenon and interference of complex lighting environment factors, and the effect of Gaussian mixture model is good, but the calculation of Gaussian mixture model is too complex to meet the real-time requirements of vehicle detection, the real-time performance of three frame difference method is good but the accuracy is not high enough, and there is an obvious "empty" phenomenon. In contrast, this algorithm effectively solves the ghost phenomenon and reduces the impact of complex environmental factors.

TABLE I. EVALUATION RESULTS OF VEHICLE INSPECTION

	GMM	Three frame difference	Original vibe	Improved ViBe
Recall	72.1%	70.0%	80.5%	87.1%
Precision	91.2%	83.1%	90.2%	92.3%
F1	80.1%	80.0%	85.1%	90.0%

VI. CONCLUSION

In order to solve the ghost phenomenon in vibe algorithm and the problem of low detection accuracy in complex illumination environment, this paper first proposes an adaptive threshold vibe algorithm in order to update the background accurately under the condition of complex illumination change. By defining two kinds of vehicle detection error functions, according to the error range calculated by these two functions, an algorithm is used to determine the rationality of threshold. Vehicle detection and background update are performed by using the adaptive algorithm.

In order to solve the problem of ghost, three frame difference method and adaptive vibe algorithm are introduced. Finally, the experimental results show that the improved adaptive vibe algorithm can effectively remove the "ghost" phenomenon and improve the accuracy of vehicle detection in complex lighting environment. Figure 4 shows the change of the number of ghost pixels with the number of frames. The X axis represents the number of frames, and the Y axis represents the number of ghost pixels. It can be seen from the figure that the improved algorithm in this paper greatly accelerates the ghost elimination speed, which is significantly faster than the Gaussian mixture algorithm and the vibe algorithm.

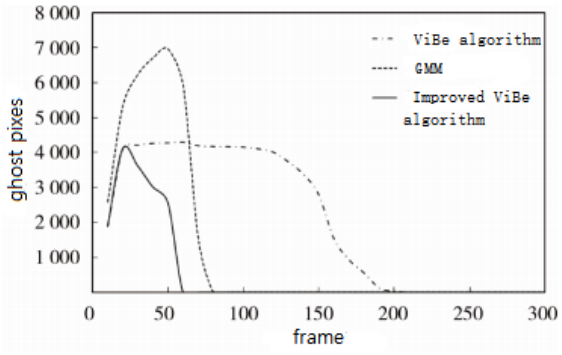


Figure 4. Ghost elimination speed.

REFERENCES

[1] Delpiano J, Jara J, Scheer J, et al. Performance of optical flow techniques for motion analysis of fluorescent point signals in confocal microscopy. Machine Vision & Applications, 2012, 23(4):675-689.

[2] J.-G. Yan, W.-H Xu. Moving object real-time detection algorithm based on new frame difference. Computer Engineering & Design, 2013, 34(12):4331-4335.

[3] Yang W, Zhang T. A new method for the detection of moving targets in complex scenes. Journal of Computer Research & Development, 1998.

[4] Kaewtrakulpong P, Bowden R. An improved adaptive background mixture model for realtime tracking with shadow detection. Springer US, 2002.

[5] Kim K, Chalidabhongse T H, Harwood D, et al. Real-time foreground– background segmentation using codebook model. Real-Time Imaging, 2005, 11(3):172-185.

[6] Barnich O, Van D M. ViBe: a universal background subtraction algorithm for video sequences. IEEE Transactions on Image Processing A Publication of the IEEE Signal Processing Society, 2011, 20(6):1709-1724.

[7] Barnich O, Van Droogenbroeck M. ViBe: A unrsal background subtraction algorithm for video sequences[J]. IEEE Transactions on Image Processing, 2011, 20(6):1709-1724.

[8] Hu Changhui, Lu Xiaobo, Ye Mengjun, Zeng Weili. Singular value decomposition and local near neighbors for face recognition under varying illumination [J]. Pattern Recognition, 2017,64: 60-83.

[9] Z. Qiming and M. ChengQian, A vehicle detection method in tunnel video based on ViBe algorithm,2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, 2017, pp. 1545-1548.

[10] C. Pan, Z. Zhu, L. Jiang, M. Wang and X. Lu, "Adaptive ViBe background model for vehicle detection," 2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, 2017, pp. 1301-1305.

[11] Ekpar F. A framework for intelligent video surveillance. Proceedings of the IEEE 8th International Conference on Computer and Information Technology Workshops. Sydney, QLD, Australia. 2008. 421–426.

Review of Development and Application of Future Network (IPV9)

Lou Peide
Secretary-general of China mobile communication federation
Beijing University of Posts and Telecommunications, Beijing, 100876, China
E-mail: lpd@cmca.org.cn

Liu Zhang
State and Provincial Joint Engineering Lab. of Advanced Network, Monitoring and Control, China
School of Computer Science and Engineering, Xi'an Technological University, Xi'an, 710021, China
E-mail: 604463203@qq.com

I. INTRODUCTION

In 28 June, 2019, general secretary Xi Jianping proposed to accelerate the construction of the "BRICS Institute of Future Networks" at the informal meeting of the BRICS leaders on the eve of the G20 in Osaka, Japan. Which received a unanimous approval by the leaders of the BRICS countries, this is the first time that president Xi mentioned "future network" in his previous public speech. It shows that after more than ten years of researches and practices, the international technology of the "Future Network" which promoted by the International Standards Organization ISO/IEC since 2007 has been highly recognized by the Chinese government. It will soon become the core technology concept that guiding the innovation and development of China's information and communication technology, and will provide a powerful impetus for the development of the international standards of "future network" and related industry chains.

II. THE CONCEPT AND DEVELOPMENT OF "FUTURE NETWORK"

"Future Network" is a specific term, which represents a computer network standardization project in the field of ISO/IEC international standards since 2007. Its mission is to develop a new network system that is independent of the existing Internet with an empty cup design and a new architecture approach. It will achieve safer, more economical, faster, more flexible, and more capable of meeting the technical requirements of the new era, with more than 15 years of research and it will put in preliminary commercial around 2020.

The basis for whether it can be called "future network" internationally is the definition given by the International Standards Organization. One of the conditions is that the address length must be at least longer than 2 to the power of 128. Since 2007, the United Nations International Standards Organization ISO/IEC has been developing and formulating the "Future Network" international standard, and released nine technical standards reports of the ISO/IEC Future Network TR 29181 series(Among them, the report TR 29181-2*Name and Addressing* and TR 29181-5 *Safety* were completely led by Chinese experts and were completely led by Chinese experts and were publicly approved to release in 2014.) discussed the problem of the existing Internet network and proposed the technical requirements of the future network. According to the agreement of the WTO Agreement, in the case of existing ISO/IEC international standards, the relevant policies and regulations of each country should be consistent with the international standard specifications. Therefore, the international standards of

DOI: 10.21307/ijanmc-2019-066

the ISO/IEC future network should be the basis to delimit the type of future networks in each country.

In 2014, the academician of the Chinese Academy of Sciences set up a consulting research topic "Future Network Architecture and Its Security", comprehensively discussed the technical characteristics of the future network security architecture, and demonstrated the feasibility of the future network. After the argumentation by the Chinese Academy of Sciences Committee, the President of the Chinese Academy of Sciences submitted a report to the State Council in July 2015, proposing to set up a major project for the future internet of the thirteenth Five-Year Plan and use the willpower of the state to promote future network research. This report should be the most authoritative statement of the Chinese scientific community on the future network, and should be the basis for defining the future network in China. In June 2016, the Ministry of Industry and Information Technology released the standard of future network/IPV9 electronic industry.

The basic purpose of the future network is to be able to survive independently without parasitic on the Internet (but it can compatible with the Internet), thereby improving the sovereignty, survival and control cyberspace in countries around the world. and there is no "Internet", indicating that the future network does not use the technology of internet, not just the upgrade of the Internet. Some domestic researchers have translated the future network into "Future Internet", and some even translated "Next Generation Internet IPv6" into "Future Internet", which will cause the future network to be mistaken for the Internet high-speed intelligent project. There will be administrative disputes with Internet agencies such as the IETF. The use of the term "Future Network" by the International Standards Organization ISO/IEC indicates that there is no affiliation with the Internet, and there is no dispute over its jurisdiction, avoiding the misunderstanding of "the combination of internet and new generation of information technology such as 5G, internet of Things, cloud computing, big data, and artificial intelligence is future network"

III. CHINA ATTACHES GREAT IMPORTANCE TO "FUTURE NETWORK"

Among the BRICS countries, China is the most active and contributing member of participate the ISO/IEC future network international standards. as the former Ministry of Information Industry, established an expert working group leader unit of "decimal network standard work group "and" Overall Design of the New Generation of Security and Controllable Information Network Technology Platform" around 2001, led the project to be responsible for the future network/IPV9 protocol standard proposal and the implementation of the new generation of secure and controllable network platform design, planning and demonstration projects. , achieved very outstanding achievements. Therefore, the BRICS countries can be given more support and assistance in terms of standards, theory, technology, industrial development and application. The most important thing is that China should provide the future network technology architecture plan that reflects China's wisdom, advanced and feasible, to the BRICS Future Network Research Institute, including technical solutions that can reflect China's complete network sovereignty and security architecture.

Now, President Xi mentioned the "future network" on international occasions for the first time in public, and it is clearly calling for the acceleration of the construction of the "BRICS Institute of Future Networks" project. This is a long-awaited and inevitable event for our research team that has long been engaged in the future of network technology research and development of international standards. For more than 20 years, China’s technical team of "decimal network standard work group" and the "Overall Design of the New Generation of Security and Controllable Information Network Technology Platform", under the long-term planning and guidance

of the party and the government, under the leadership of the expert group leader Xie Jianping, with the spirit of innovation, advancing with the times, and struggling hard. We have made breakthroughs in future network/IPV9 cyberspace, root name server, IPV9 Taishan operating system and legal digital currency anchor technology: The design, planning, construction and the implementation of demonstration projects of the new autonomous safety and controllable network. The engineering implementation technology has been mastered and can be interconnected with the existing IPV4/IPV6 Internet. China has built a computer communication network with independent intellectual property rights and network sovereignty which is independent of the US Internet but compatible with the US Internet. It forms the most basic element of China's fifth territorial cyberspace sovereignty.

In 2000, Shanghai General Chemical Technology Research Institute researched the IPV9 protocol and digital domain name top-level domain name server under the support of the Shanghai Municipal Government, and passed the IPV9 network protocol expert appraisal of the team led by Academician Shen Changxiang. In 2007, as an integral part of China's national ISO membership, participated in the standard formulation of ISO/SC6 future network standard formulation, led the "name and address", "safe" technical reporting standards, and obtained the recognition and support from ISO standard organization, the National Standards Committee and the Ministry of Industry and Information Technology. China has proposed a hybrid communication scheme of IP communication and virtual real circuit and a method of address encryption/first verification and post-communication for the technical defects of IPv4/IPv6.It breaking through the United States' Intellectual property monopoly on the Internet TCP/IP protocol and become an important part of the ISO technical report of the International Standards Organization. In the second vote of the ISO/IEC future network standard in 2014, members of China, the United States, Russia, Canada all voted in favor. In June 2016, the Decimal Network Standards Working Group led to formulate the industry standards of six future network/IPV9 electronics and the Ministry of Commerce issued by the Ministry of Industry and Information Technology and the Ministry of Commerce.

IV. THE PLANNING AND CONSTRUCTION OF "FUTURE NETWORK"

A. Planning of future network/IPV9

Due to the achievements of China's future network and the new generation Internet (IPV9), the State Council has incorporated the future network and the new generation Internet into the medium and long-term (2012-2030) planning of major national science and technology infrastructure construction.

In November 2017, the General Office of the CPC Central Committee and the General Office of the State Council of the People's Republic of China issued a focus statement for the implementation of the demonstration application of the future network:

a) Build an independent technology industry ecology, strengthening the technology innovation of network front-end, and accelerating the construction of major scientific research infrastructure.

b) Accelerate the construction of national future network test facilities, and actively carry out trial verification and application demonstration of new network technologies and new applications.

c) Further accelerating the new internet architecture, and basic technology innovations such as new addressing and routing, endogenous network security, and network virtualization Innovate, strengthen the experimental verification and application demonstration of new technologies and new applications of the network, continuously improve the productivity transformation level of innovation achievements, significantly enhance the independent innovation capability of network

information technology, and form the first-mover advantage of network technology in the future.

Currently there are four communication methods in the world. The first three are invented by Westerners: telegraph, telephone, and IP packet switching. The fourth communication method is a hybrid method of circuit and IP packet switching (TCP/IP/M), which was invented by the Chinese. The first three communication methods are inherently unsafe. The method invented by China has established a new way of secure communication and solved the key technologies of network security communication.

B. Construction of Future Network / IPV9

1) China has built a research system and industrial system of future network/IPV9:

a) Hardware: core router, edge router, IPV9-IPv4 protocol conversion router, embedded router, client, Beidou/GPS network timing server, IPV9 protocol NTP timing service, V9-enabled application terminal, IPV9 firewall,IPV9/ IPV4 VPN server.

b) Software: IPV9 network management system, IPV9 automatic distribution application access system,IPV9 operating system,IPv4 and IPV9 address and network management,IPV9 windows protocol stack, Windows system IE browser IPV9 plug-in.

c) Domain Name System.

d) The parent root server, root name server, the top-level domain name server, the IPV9 reverse resolution server and the second-level domain name server.

At present, China has built demonstration projects in Beijing, Shanghai, Shandong, Jiangsu, and Zhejiang that have IPV9 address space, root name server and IPV9 backbone optical cable system, which can replace the Internet network management signaling imported from the United States through the Pacific Ocean cable. The national military-civilian integration IPV9 backbone optical cable and Gateway MSC is under construction.

2) The construction goals of future network / IPV9:

a) Based on the IPV9 autonomous network, build a self-controllable and secure network that connect the Communist Party of China, Politics, Military and Civilian, safeguard China's cyberspace sovereignty, and accelerate the promotion of domestic controllability of cyberspace Achieve trusted network entity authentication and authentication, trusted management, content supervision and other functions, lay the foundation for the construction of national big data center to build national integration, and provide safe and reliable information services for the decision-making of the Party Central Committee and the Central Military Commission's combat training department.

b) China has already had the independent intellectual property rights of root name server and address space, it has realized a historic transformation from renting to autonomy, it shows that China has the ability to build a second computer network with independent intellectual property rights and a national integrated national big data center, and change the situation that the cyberspace security strategy politically by other countries.

c) Form the national cyberspace territory and network strategic depth system, improve the level of cyber operation management, improve the ability to maintain cyberspace security, and accelerate the promotion of cyberspace domestically controlled alternatives.

3) Introduction to typical applications of future network/IPV9:

a) The application of 5G-future network/IPV9 movie network release application:

Now the 5G network of China Unicom Beijing and China Mobile Suzhou have been directly connected through the IPV9 fiber routing backbone node of Beijing University of Posts and Telecommunications and the IPV9 national backbone optical cable network, and achieved the world's first time End-to-end

500Mbps to 1000Mbps speed on May 21 this year. On the IPV9 national backbone network +5G local access/5G core network, the digital film program network distribution work was successfully carried out (the data capacity of each movie was about several hundred GB), and the national network distribution of Chinese movies was first entered in the new era of "one hour", we will continue to carry out the intelligent work of the theater centered on the IPV9. The first realization of the future network / IPV9 + 5G + broadband ultra-high-speed video business proved the future network / IPV9 can play a major role in the process of solving the amazing problem of our famous entrepreneur Ren Zhengfei "5G development is slow, the current network structure is not good, and the network speed is only 20/30Mbps during the day. "And it can also prove that network/IPV9+5G is also suitable for remote 8K TV service, remote medical service and remote AR/VR broadband ultra-high speed video service.

b) "Health Taian " IPV9 Big Data Platform:

"Health Tai'an "IPV9 big data platform project relies on the existing backbone optical cable and user transmission access network of Shandong Broadcast Network Co. Ltd. Tai'an Branch, using IPV9 network technology to upgrading and construction, cover the medical and health institutions of the city, county, township and village levels and the medical insurance bureau, the administrative department and the finance bureau of Tai'an, and further expand to families and individuals. The bandwidth meets the requirements of healthy Tai'an big data business and can be sustainable. The expansion realizes compatible security operation between IPV9 network and IPV4 network (also realizes logical security isolation between IPV9 and IPV4 and IPV6 networks).At present, two 40G and six 10G IPV9 backbone routers and 300 (by the end of this year it will rise to 3000) IPV9 thousand M user routers have been used, and Feicheng has begun to deploy to the village health center. The project is doing the third-level system certification, the project is applying for the third-level system certification, and the specific network protection measures and operations are in compliance with ISO27001 requirements.

For the Tai'an Municipal Government, on the domestically-controlled and controllable security network, the medical institutions of 3,000 village-level health centers will be connected to the IPV9 Tai'an Metropolitan Area Network (including the Tai'an Finance Bureau and Medical Insurance Department).It is necessary that supervising the bills of medical treatments for all patients, by the bureau of finance. In the past, every October, Tai'an's financial subsidies of one billion yuan per year were used up (because of excessive drug use and opaque drug purchase). Now use the financial subsidy program to block the black hole of medical expenses waste. The Tai'an Municipal Finance Bureau also requires that all the information of each medical unit should be unified in the "Big Health Cloud", which facilitates the access of doctors and patients' information, the sharing and utilization of information, and enables citizens to enjoy convenience and cost-saving benefits. Allowing citizens to enjoy the convenience of medical information sharing and the benefits of cost reduction, and also fully and effectively display the performance of the government-established medical security system. Such as the rapid and real-time verification of medical insurance reimbursement bills and redemption, this protects the quality of citizens' life.

V. CONCLUSION

In the 36th collective study of the Political Bureau of the CPC Central Committee in 2016, General Secretary Xi Jianping emphasized that it is necessary to accelerate the independent innovation of network information technology. He pointed out that network information technology is the world's most concentrated investment in R&D, the most active innovation, the most widely used, and the largest technological innovation field and it can promote technology in other fields, and it is the competitive

high ground for global technological innovation, and it is the competitive high ground for global technological innovation. We must comply with this trend, vigorously develop core technologies, strengthen the security of key information infrastructure, and improve the network governance system. We must firmly control the core technology independent innovation, break through the cutting-edge technologies and core technologies with international competitiveness of network development, accelerate the promotion of domestically controlled independent control plans and build a safe and controllable information technology system.

China is a socialist country. We firmly believe that under the leadership of the Party Central Committee with General Secretary Xi Jianping as the core, we will surely win the final victory of the national cyberspace sovereignty battle and building a community of human destiny. Although the road ahead is difficult, We have the future network intellectual property and root name server system that have the network core, and have built the future network/IPV9 industrial ecosystem and several typical demonstrations engineering.

We hope that China will vigorously develop China's information and communication industry and the national digital economy industry on the basis of the development of the future network/IPV9, and provide services for the development of the BRICS countries and the development of countries along the "Belt and Road".

REFERENCES

[1] Tang Xiaodan etc. Computer Operating System (third edition) [M]. Xi'an: Xidian university press, 2010.

[2] Xie Jianpingetc.A method of assigning addresses to network computers using the full decimal algorithm[P]. CN: ZL00135182.6, 2004.2.6.

[3] Xie Jianpingetc.Method of using whole digital code to assign address for computer [P]. US: 8082365, 2011.12.

[4] RFC - Internet Standard. Internet Protocol, DARPA INTERNET PROGRAM PROTOCOL SPECIFICATION, RFC 791,1981.09.

[5] S. Deering, R. Hinden, Network Working Group.Internet Protocol, Version 6 (IPv6)-Specification, RFC-1883, 1995.12.

[6] M. Crawford. Network Working Group.Transmission of IPv6 Packets over Ethernet Networks.RFC-2464, 1998.12.

Detection of Blink State Based on Fatigued Driving

Lei Chao
School of computer science and engineering, Xi'an Technological University
Xi'an, China
E-mail:callofduty2015@163.com

Li Guang
School of computer science and engineering, Xi'an Technological University
Xi'an, China
E-mail:865413666@qq.com

Wang Changyuan
School of computer science and engineering, Xi'an Technological University
Xi'an, China
E-mail:cyw901@163.com

Shi Lu
School of computer science and engineering, Xi'an Technological University
Xi'an, China
E-mail:892178238@qq.com

***Abstract*—In recent years, with the improvement of the national economy, the penetration rate of automobiles has been increasing, and traffic accidents have also increased. Fatigue driving is the main factor in many traffic accidents. Fatigue driving can cause the driver's inattention, slow response, and make wrong decisions on danger signals, which affect the driver's personal safety. In modern development, driving safety is developing towards intelligence and safety. Therefore, the detection of driver fatigue has become a generally accepted demand. This paper proposes a method to calculate the threshold of blinking, which can detect the blinking state of the driver in real time through video. During the driving process, when the driver is in the closed eye state for a long time, an early warning is issued to avoid the accident. This paper uses Python language to achieve the first, through the digital image technology call Dlib open source library to detect 68 feature points of the face, and then measure the aspect ratio between the length and width of the human eye, and finally through the Kmeans clustering algorithm to collect the ratio The analysis yields the blink threshold. The experimental results show that the recognition rate is 92.5% when the video frame rate is 30, and the recognition accuracy is 92.5%. The experimental results show that the method designed in this paper can quickly detect the fatigue characteristics of the human eye, has a higher recognition rate and accuracy for fatigue driving, and helps reduce the occurrence of traffic accidents.**

Keywords-Blinking Algorithm; Fatigue Detection; Digital Image Processing; Clustering Algorithm; Key Points Of Human Eyes

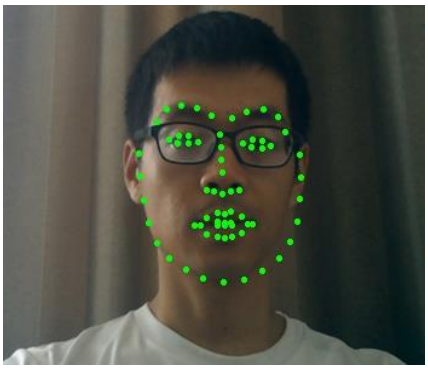
I. INTRODUCTION

With the improvement of people's material living standards, cars have become the main means of transportation for people, but the growing number of vehicles has led to more traffic accidents. According to statistics, fatigue driving is the main cause of traffic accidents[1,2].Under normal circumstances, the medical community believes that there are two reasons for fatigue driving, one is because the driver's attention is too concentrated, and the other is that the body does not rest well. Because of being in this state for a long time, the body will be fatigued, lose concentration, the driver will snoring, lose concentration, decrease the ability to judge dangerous situations, and cause traffic accidents. At present, there are relatively few applications of fatigue driving equipment in China's in-vehicle systems. Fatigue detection mainly through facial features, eye and mouth features, human electrical signal characteristics and convolutional neural network characteristics[3,4,5].The detection of facial features is generally based on the frequency of blinking eyes, the degree of mouth opening, and the frequency of head movements due to fatigue. The fatigue of human body electrical signals is generally the measurement of surface EMG signals, because human fatigue can be expressed by muscle physiological information. Surface EMG signals can reflect real-time physiological processes of muscle information and physiological signals on the skin surface. Convolutional neural networks generally extract facial features through image processing methods, and then extract the main features through convolutional layers, pooling layers, and fully connected layers to analyze and determine whether fatigue. Chen[6] uses the ASM algorithm to accurately locate the eyes and mouth area, calculates the eye's aspect ratio, mouth height value, and black and white pixel ratio near the mouth, and obtains the blink frequency and mouth opening degree. The degree of mouth opening is used as an input to the fuzzy

inference engine to obtain three types of fatigue levels to accurately quantify the degree of fatigue

The method proposed in this paper is to judge the driver's fatigue driving according to the characteristics of the human eye. Because the digital image processing open source visual library OpenCV comes with a human face detection library, but the disadvantage is that the lighting requirements are very high, the lighting slightly changed, it will be difficult to locate or inaccurate positioning[7]. Therefore, this paper chooses Dlib open source library to detect human eye features. Firstly, the 68 face feature points provided by the Dlib open source library are used to accurately calibrate the position of the face and the human eye, and then the aspect ratio between the length and the width of the human eye is measured. Finally, the Kmeans clustering algorithm is used to analyze the collected ratio. The threshold of blinking. Figure 1 below, a is the 68 face feature points marked by Dlib, and b is the feature point on the face of the paper.

(a) 68 feature points of a face annotation



(b) Recognition of face images

Figure 1. Facial feature points

II. RELATED WORK

A. Blink detection and threshold analysis methods

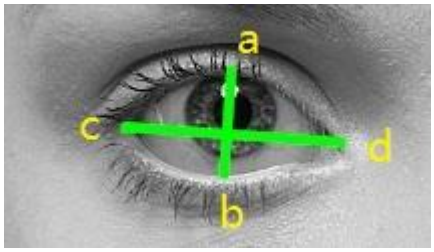
This chapter mainly introduces the blink algorithm formula and blink threshold analysis method. The blink threshold analysis method uses the Kmeans clustering algorithm in machine learning. There are many methods for blink detection, such as support vector machine classification, eye movement sequence analysis, convolutional neural network feature extraction, eye feature point analysis, etc. This article uses the eye feature analysis method. Threshold analysis methods in machine learning usually use regression algorithms, decision tree methods, Bayesian methods, and clustering algorithms. This article uses the Kmeans clustering algorithm in machine learning.

B. Blinking formula

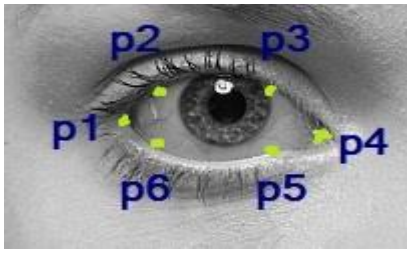
There are currently many methods in the field of blink detection. Andrej Fogelto et al[8] analyzed the relationship between the speed of blinking and the duration of time through the cyclic neural network (RNN), so as to better distinguish the state of blink and blink, through the comparison of one-way and two-way circulating neural networks. One-way neural networks work better in blink detection. However, neural networks have their own limitations. For example, network merging will lose a lot of information. In the field of face recognition, the relationship between part and whole is neglected, and each person's face features are different, so it takes a lot of time to train. parameter. RenAnhu et al[9] trained the blink classifier through the AdaBoost algorithm, but the AdaBoost algorithm is very sensitive to the discrete data of the blink. The detection of the eye part during image processing is easily affected by the speed of light and object movement, just as the blinking behavior is fast. the process of. ZengYouwen et al[10] used the correlation between computer signals and the number of blinks to determine the fatigue state, but the experiment required special equipment and equipment, and the operation was difficult and difficult to implement. Since the blinking algorithm proposed by Soukupov á[11] has measured the blink threshold of the aspect ratio of the eye, using the support vector machine[12] method to analyze the collected threshold of the blink to finally obtain an EAR of 0.2, the algorithm of this paper is the In addition, by measuring the aspect ratio of the eye, the Key clustering algorithm in machine learning is used to obtain the blink threshold. As shown in Figure 2.1a, the absolute value of the longitudinal distance ab of the eye will become smaller when the eye is blinking. At this time, the ratio of cd to ab will suddenly become larger, so we analyze the threshold when the ratio

becomes larger when blinking. Figure 2a shows the distance between ab and cd, and b is the position of the human eye feature point marked by Dlib. This paper proposes the blink threshold formula as follows:

$$Blinkthreshold = \frac{2|p1-p4|}{|p2-p6|+|p3-p5|} \quad (1)$$



(a)



(b)

Figure 2. (a) The lateral distance is cd longitudinally ab; (b) dlib human eye calibration features

C. Kmeans clustering algorithm

The Kmeans algorithm is a relatively common algorithm in clustering algorithms. Its advantage is that it is easy to implement and understand, and the calculation speed is fast. The core idea is to calculate the distance between the sample point and the centroid of the cluster, and divide the calculated result into the same cluster as the sample point with the centroid of the cluster.

The similarity between samples in K-means is determined by the distance between them. The closer the distance is, the higher the similarity is. The common distance calculation methods are Euclidean distance, Euclidean distance and Manhattan distance. European distance. In the cluster analysis, the formulas for two m-dimensional samples xi=(xi1,xi2,xi3...,xim) and xj=(xj1,xj2,xj3...,xjm) are as follows:

$$dist_{ed} = \sqrt{\sum_{k=1}^{m}(x_{ik}-x_{jk})^2} \quad (2)$$

The steps of the k-means algorithm are as follows:

1) First randomly select the centroids of K clusters.

2) Calculate the Euclidean distance from each sample point to each centroid, and classify it into the cluster with the smallest center of mass, and then calculate the centroid of each new cluster.

3) After all the sample points are divided, recalculate the position of the centroid of each cluster, and then iteratively calculate the distance from each sample point to the centroid of each cluster, and then re-divide the sample points.

4) Repeat steps 2 and 3 until after the iteration, the partitioning of all sample points remains unchanged, and K-means gets the optimal solution.

The main problem of the calculation result is to ensure the convergence of the algorithm. Here, the square error is calculated by the following formula, which is used to illustrate that the clustering effect can minimize the sum of squares in each cluster.

$$J(c,u)=\sum_{i=1}^{K}\left\|x^{(i)}-u_{c(i)}\right\|^2 \quad (3)$$

$j(c,u)$ represents the sum of squares of the distance from each sample point to its cluster, $u_{c(i)}$ represents the centroid of the cluster to which the i-th sample belongs, and the smaller $j(c,u)$, all the sample points and their clusters The smaller the distance, the better the quality of the division. The termination condition of the K-means algorithm is that $j(c,u)$ converges to a minimum. In order to achieve clustering, the maximum value of the objective function is obtained. Take a one-dimensional array as an example.

$$J = \sum_{i=1}^{k}\sum_{x_j\in u_i}\left(x^{(i)}-u_{c(i)}\right)^2 \quad (4)$$

Transform the above formula to get:

$$\frac{\partial J}{\partial u_i} = \frac{\partial}{\partial u_i}\sum_{i=1}^{k}\sum_{x_j\in u_i}(x_j-u_i)^2$$

$$= \sum_{i=1}^{k} \sum_{x_j \in u_i} (x_j - u_i)^2$$

$$= (-2) * \sum_{x_j \in u_i} (x_j - u_i)$$

When $(-2) * \sum_{x_j \in u_i} (x_j - u_i) = 0$ $u_i = \frac{1}{|c_i|} \sum_{x_j \in u_i} x_j$

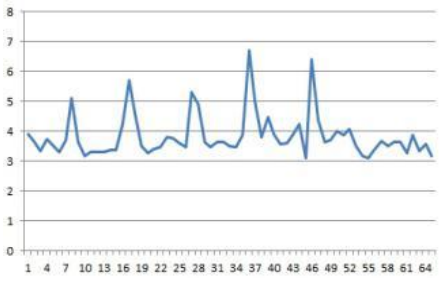
The result of the optimization is to calculate the mean of the cluster.

During the experiment, the algorithm may be too slow to achieve effective results because the data set is too large. Therefore, you can specify the maximum number of convergence times for the K-means algorithm or specify the cluster center transformation threshold. When the algorithm reaches the maximum number of times or When the cluster center rate of change is less than a certain threshold, the algorithm stops updating.

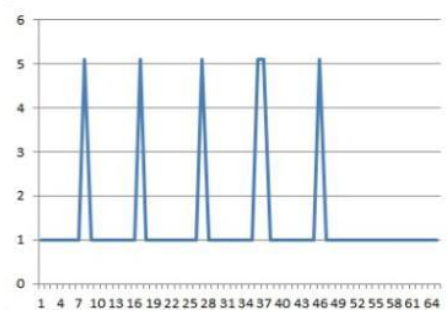
K-means algorithm advantages: easy to understand, easy to implement, high operating efficiency, the disadvantage is that the greedy strategy is used to cluster the sample points, resulting in easy local convergence of the algorithm, slower data processing in big data, and outliers and The noise is very sensitive, and a small number of outliers and noise points can have a significant impact on the averaging of the algorithm.

III. THE EXPERIMENT

This article mainly introduces two aspects, one is to introduce the source of the experimental data set, and the other is to analyze the experimental data.Figure 3 shows the author's own data collection and analysis. From left to right, the first picture is the analysis of the blink value of the author's 5 blinks, and the middle figure is when the blink threshold is 5.1 when Kmeans = 2 The blink data graph, the right graph is the actual record of the author's experiment. Figure 4 shows the data of a random experimental sample in the public data set. From left to right, the first figure analyzes the blinking value of the aspect ratio of the 3 blinks in the experimental sample. The blink data graph at 5.1, the right graph is the actual record of the experimental sample.This article uses the blink data set provided by Zhejiang University[13]. All the data in the data set is collected under the natural light of the room. There is no special illumination. The collected equipment is a common camera that comes with the computer. There are 80 video clips and 20 people participate. , four segments per person, these four segments represent a. front view without glasses; b. front view with thin edge glasses; c. front view with black frame glasses; d. View. Participants spontaneously blink at the front of the camera at normal speed. The size of each video is 320*240, the frame rate of the video is 30fps, and the duration of recording is generally about 5 seconds. The experimenter usually blinks about 1 to 6 Times, a total of 255 blinks. The following are the results of data collection and analysis.



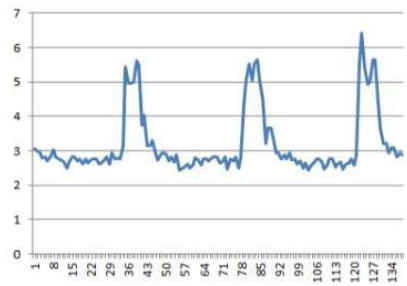
The text person has 5 blink-eye aspect ratio data



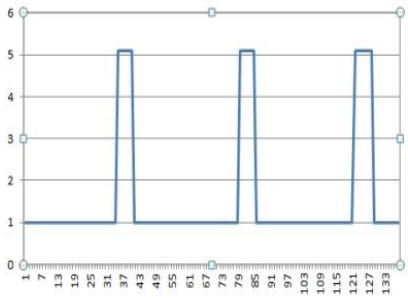
Kmeans=2, the blink data chart

The text person image

Figure 3. The following is the experimental data of the paper



Public data set 3 times blinking eye aspect ratio data



Kmeans=2, blinking data chart

Public data set sample

Figure 4. Public data set sample

The following table a is a comparison of the public data sets provided by Zhejiang University and the experimental results of the text person. Table b is a comparison of other methods with the method of this paper. RenAnhu[9] trained the classifiers of blinking and closed eyes through the Adaboosts algorithm. The person in the video is then tested for blinking. Zhang Wei[14] performed a correlation analysis of the blink of the eye by analyzing the left forehead EEG signals Attention and Meditation and Blink data.

TABLE I. COMPARED WITH PUBLIC DATASETS

Experimental sample	Blink threshold	Blink times	Number of recognition	Recognition rate
Text person	5.1	100	90	90%
Public data set	5.1	255	236	92.5%

TABLE II. COMPARED WITH OTHER LITERATURE

literature	Recognition rate
9	91.5%
14	83.7%
This article	92.5%

IV. CONCLUSION

This paper overcomes the shortcomings of digital image processing and OpenCV vision open source library, and combines the existing open source Dlib machine learning library,The data between the vertical and horizontal ratio of blink is calculated by mathematical method, and the threshold value of the vertical and horizontal ratio of blink is analyzed by means of kmeans clustering algorithm in machine learning. According to the analysis of the public data set of Zhejiang University, when the threshold value of the vertical and horizontal ratio of blink is 5.1, the accurate recognition rate of blink is 92.5%. Through the experimental comparison, this algorithm can effectively detect the fatigue state of blink, which is more important This algorithm is fast, efficient and easy to transplant to various devices, and has great practical value in the field of fatigue driving. The shortcomings of the paper: for fatigue monitoring, not only eyes as a reference point, nose tip shaking, mouth opening and so on have an impact on face fatigue, so the fatigue detection algorithm in this paper needs to be improved.

REFERENCES

[1] M. Hülsmann,D. Donnermeyer,E. Schäfer. A critical appraisal of studies on cyclic fatigue resistance of engine‐driven endodontic instruments[J]. International Endodontic Journal,2019,52(10).

[2] Pierre Thiffault,Jacques Bergeron. Monotony of road environment and driver fatigue: a simulator study[J]. Accident Analysis and Prevention,2003,35(3).

[3] Liu Longfei, Wu Shizhen, Xu Wangming. Real-time detection method of fatigue driving based on face feature point analysis[J]. Television Technology, 2018, 42(12): 27-30+55.

[4] Yan Wang,Rui Huang,Lei Guo. Eye gaze pattern analysis for fatigue detection based on GP-BCNN with ESM[J]. Pattern Recognition Letters,2019,123.

[5] Driver's Fatigue Detection Based on Yawning Extraction[J]. Nawal Alioua,Aouatif Amine,Mohammed Rziza,Aboelmagd Noureldin.International Journal of Vehicular Technology. 2014

[6] Chen Xin, Li Weixiang, Li Wei, Zhang Wenqing, Zhu Yuan. Multi-feature fusion fatigue detection method based on improved ASM [J]. Computer Engineering and Design, 2019, 40 (11): 3269-3275.

[7] Rafael C.Gonzalez,Richard E.Woods.Digital Image Processing,Third Edition[M],2017

[8] Andrej Fogelton,Wanda Benesova. Eye blink completeness detection[J]. Computer Vision and Image Understanding,2018.

[9] Ren Anhu, Liu Bei. Face Recognition Blink Detection Based on Adaboost[J]. Computer and Digital Engineering, 2016, 44(03): 521-524.

[10] Zeng Youwen, Feng Zhen, Zhu Yabing, Li Qi.Relationship between the number of blinks and fatigue based on EEG experiment[J].Journal of Changchun University of Science and Technology(Natural Science Edition),2017,40(01):123-126.

[11] Tereza Soukupova ´,Jan C˘ ech, Eye blink detection using facial landmarks[J]. 21st Computer Vision Winter Workshop(CVWW),2016

[12] J. Manikandan,B. Venkataramani. Study and evaluation of a multi-class SVM classifier using diminishing learning technique[J]. Neurocomputing,2009,73(10).

[13] F.Song, X.Tan, X.Liu and S.Chen, Eyes Closeness Detection from Still Images with Multi-scale Histograms of Principal Oriented Gradients, Pattern Recognition, 2014.

[14] Zhang Wei, He Jian, Zhang Yan, Zhou Ming. A wearable fatigue driving detection system based on EEG and blink frequency[J]. Computer Engineering, 2017, 43(02): 293-298+303.

Research of Network Closed-loop Control System Based on the Model Predictive Control

Xu Shuping
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China
E-mail: 563937848@qq.com

Guo yu
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China

Wang Shuang
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China

Su Xiaohui
School of Computer Science & Engineering
Xi'an Technological University
Xi'an, 710032, China

***Abstract*—Uncertainty latency of the remote closed-loop control system in information transmission through Internet, Analysis delays how to influence closed-loop control system. Based on predictive control method of neural network, research the application of closed-loop control system control methods under a random network delay. Simulation results show that: This method is able to reflect and predict the delay characteristics of between network path represented by the measured data, and can be replace actual network to research in application based on Internet closed-loop control system; the methods used is fast and accurate, it can be used for online learning network model and predict the network delay value, provides a new way of remote closed-loop control based on Internet.**

Keywords-Remote Control; Neural Network; Network Delay; Model Predictive Control

I. INTRODUCTION

The remote control system is an integration of control technology and network communication technology, it applications in many fields more and more common as ocean development, space station maintenance, remote surgery, virtual reality in recent years, and stable, fast, accurate is the highest target remote control system pursue[1].

Closed-loop controller is to control by the disturbances of feedback, which is by comparative behavior of the system output and the deviation between expectations to make the appropriate control action to eliminate the bias in order to achieve the desired system performance. It has the ability to suppress interference, is not sensitive to changes in device characteristics, and can improve the response characteristics of the system.

The delay phenomenon exist in the field of remote control is a common problem exists in the remote closed-loop control applications. Delay does not only exist in the before control channel of system and in feedback channel. The delay in before control channel Makes the control signal unable to act on the controlled object, the delay in the feedback channel makes the controller can not found the change of controlled object immediately[2].

DOI: 10.21307/ijanmc-2019-068

Delay value influence by the inherent properties of control information transmission network such as the network structure, the amount of data transmission, the transmission timing and transmission agreements and other factors[3], the size of the delay values are reasonable from a few hours to several days[4], the approximate value will bring a big problem of the stability and dynamic quality of the remote control system[5], analysis and research the delay problems of the remote control system is a long-term in this area.

II. ISSUES PROPOSE

A. *The influence of delay on the stability of remote control system*

Delay has a lot of influence on real-time, accuracy and stability performance of the remote control system. Kinetic of equation a single-link robotic arm second-order remote control system as follows[6]:

$$\frac{d^2\phi}{dt^2} = -10\sin\phi - 2\frac{d\phi}{dt} + u$$

Among them, ϕ represent the angle of robot arm; μ represent the behalf of DC motor torque. simulink block diagram of Mechanical Arm shown in Figure 1.

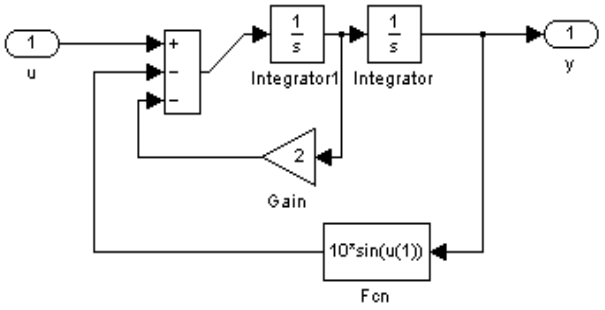
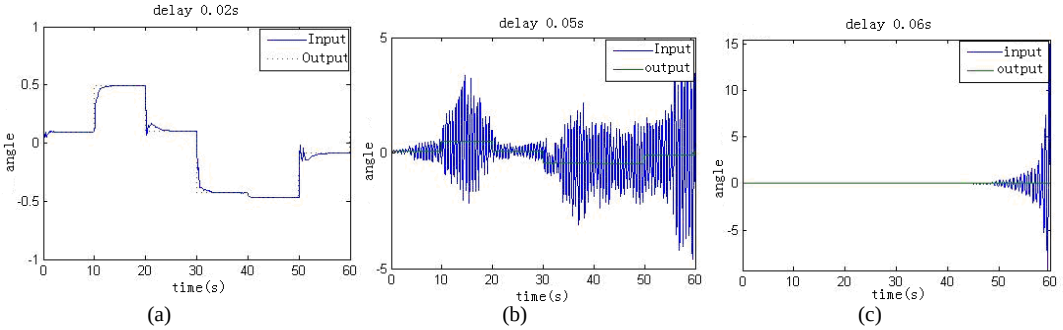


Figure 1. Simulink Block Diagram of Mechanical Arm

Since the forward channel and feedback channel in remote control system is generally the same physical link, the article assumes that the forward channel and feedback channel delay values equal. First set the delay time Delay to 0, that is without delay, adjust the PID parameters (how much) to get the response curve satisfied. Secondly, to maintain the constant of the PID parameters, increase the network delay value gradually when the network delay value is set to 0.02s to get feedback curve in Figure 2 (a), the performance of the system compared with without delay gradual deterioration this time , when the network latency increased to 0.05s, the feedback curve in Figure 2 (b), system becomes a oscillation system, and continue to increase the delay to 0.06s, the feedback curve in Figure 2 (c), system response divergence, that is system becomes unstable. It can be seen that increase the delay gradually make the control systems become increasingly unstable.



(a) (b) (c)

Figure 2. The Influence on the Stability of the Control System from Remote Control Delay

B. Research for the stability of remote control system

There are a lot of research on the stability improvement of the remote control system, in 1965, Ferrel put forward network delay problem of need to pay attention to time-varying in the network control [7]. Halevi and Ray in the literature[8] augmented deterministic discrete-time model for the periodic network delay, Gregory C. Walsh in the literature[9] considered controller and controlled object is nonlinear time-varying assuming no observation noise, based on nonlinear perturbation method theory, the network delay impact the system is described as a perturbation of the continuous-time systems. Walsh and Bushnel in the literature [10] to prove this method and conclusions can applicable equally to linear systems. Goktas in the literature [11] saw τ_{sc} and τ_{ca} as a multiplicative uncertainty bounded perturbation and provided the method of use robust control theory design of NCS controller in frequency domain. Studied robust passive control problem of class of long delay networked control systems in the literature [12], and derive the passive controller design method and proved the validity of the method through simulation. Short delay network control system disturbs by white noise in the literature[13], transformed impact of the random delay on the system into the unknown bounded uncertainties using robust control theory give the H_2 / H_∞ system state observer design method. Literature[14] make delay uncertainty convert to the perturbation of the closed-loop system parameters, propose the conditions for the existence of robust guaranteed cost control law based on the robust control theory and Lyapunov stability principle, and gives the method of design of network control Robust guaranteed cost state feedback controller to solve linear matrix inequality. Research show that actually adjust the controller parameters PID makes the instability of the closed-loop control system becomes stable and meet the requirements of remote control system which need less real-time demand. Modify PID parameters of the control system and set the network delay to 0.06s again get feedback curve shown in Figure 3.Research show that modify the PID control parameters has indeed improved the stability of the control system.

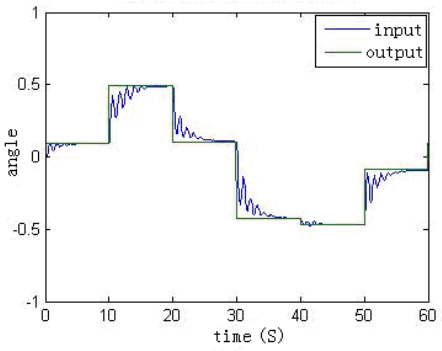


Figure 3. Response after regulate PID parameters at 0.06 seconds time delay

C. Question propose

The adjustments of the PID control parameters need to be dynamic adjustment constantly with the size of the control system delay and values and other parameters of systems, which makes controlled object in the work environment unknown to dynamically adapt adjust PID values become remote intelligent control systems that are experiencing another problem. This paper based on the method of modify the control parameters of PID values to improve the stability of the remote control system, and propose intelligent remote control system design methods with adaptive function under a random delay based on neural network theory.

III. QUESTION ANALYSIS

Remote control single-link manipulator, set the sampling period in figure 1 is 0.05s, and take the delay value of 0.05s, the control information in time k transmitted to the controller after 0.05s, as opposed to the system sampling time 0.05s, the controller receives status information at the moment of k has pass a sampling point, the state of the system has become the state in time k +1, that is state of the k time fed back to the PID controller at time k +1, the PID controller for

time k, the state at time k+1 has not yet come, but this time system status values at k-1 after a sample time delay before it is passed controller, therefore, the controller can only decision at time k should be imposed control value u (k) based on the state of the k-1 times, and this control value can be a real work on the system after a time delay, while at the time k +1 and the state of the system has been turned into a time k+1 the state of X(k+1), while u (k) produce at the state time k-1, so u (k-1)grieved and u (k +1) required difference two sampling cycles. In these two sampling cycle, the state of the system state transition, that is x (k-1) transfer to the x (k+1), x (k-1) and x (k+1) often is different lead to u (k-1) and u (k +1) is different. In other word, the system control value produced offset and the greater delay the greater offset, which is the root source of result in deterioration of the system closed-loop control performance and even instability.

The above analysis shows that the system performance deterioration caused by the remote network delay because of can not correctly calculate the amount of control exerted by the controller to the system ,if the system model is known and the size of delay is known, then forecast the state of system in accordance with the principle of the system predict compensation, and calculate the size of control value need to be added the control system in accordance with the predicted state, that is time k applications to predict the state $\hat{x}$ (k+1) at time k+1 yet not the state of x(k-1) at time k-1 calculation to be applied to the system state at time k, then the control value u (k +1) at actual time k +1, the u (k +1) after a delay transmission in the time k +1 transfer to the system just after a sampling period, the state of the system change into x (k +1) ,

So, if the predicted state $\hat{x}$ (k+1) is infinitely close to the actual state x (k +1), the performance of control network delay closed-loop control system can be infinitely close the performance of the closed-loop control system without delay links, which is the basic idea of the predictive control model. However, the delay of the control network is time-varying and controlled objects are often immediately confounding factors, it is can not use an inconvenience model to predict the state of system and can not use a specific delay time to do the fixed step predictive control, neural network has the advantages of online learning the state of the system, predictive control based on neural network has strong robustness to be adaptive to the change of system status and network delay aspects ,it is a way to solve the network latency closed-loop control.

IV. BUILD THE MODEL PREDICTIVE CONTROL LAW

According to the running state of the system over the past time and present moment, more accurate forecasting system desired output value in the future moment, calculated control value of the system should be added according to output value desired depending on certain optimization algorithm [15] is adaptive computer control of online solving control value [16], the method includes three steps: prediction model, rolling optimization and feedback correction[17].

A. The prediction model

For a module description of the alleged object behavior in the predictive control based on neural network belong to forward model of system, there use the training methods as shown in Figure 4, where dashed box picture shows the actual controlled object, here is the simulink block diagram of the robotic arm, at random input signal u to produce output y. Selected BP neural network with one hidden layer as training model of a controlled object , set the number of hidden layer neurons is 10, using the Levenberg-Marquartdt learning rules, with the group [u, y] data training neural network model of the charged object, the results shown in Figure 5, where Figure 5 (a) is the data used for training, Figure 5 (b)is convergence diagram for training [18].

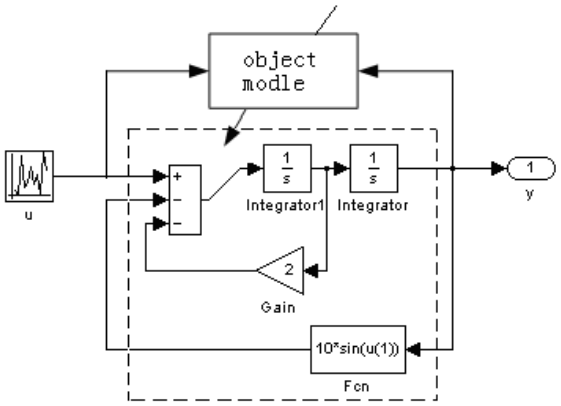
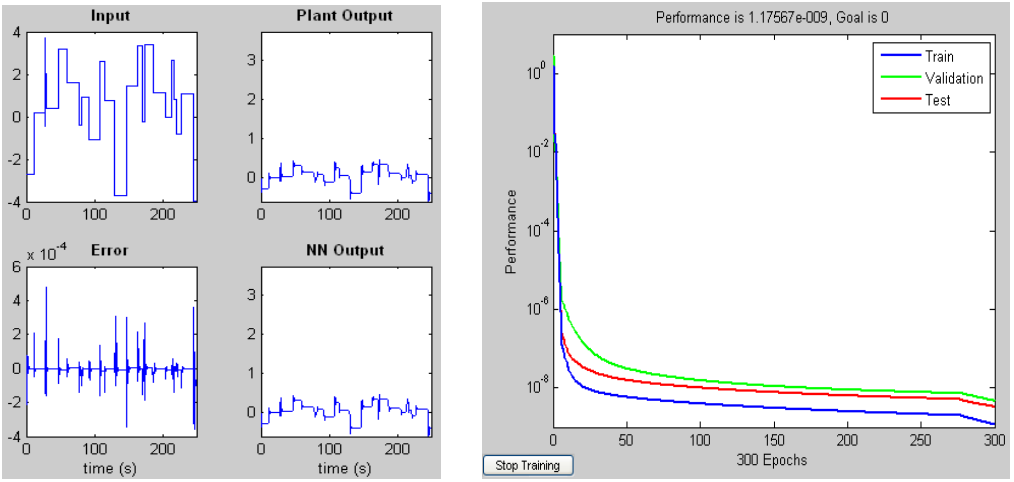


Figure 4. Neural Network Training Block Diagram of the Manipulator

B. *Receding Horizon Optimization*

Rolling optimization is an optimal control algorithm, which uses the output of the rolling finite domain optimization that is the goal of optimization over time. Predictive control proposes optimization index based on the moment in every time instead of using global optimization indexes. Rolling optimization index locality through make it can only get the global optimal solution in the ideal case, but when the model mismatch or time-varying and non-linear or confounding factors can take into account this uncertainty in a timely manner compensate, reducing the deviation, keeping the actual optimal control ,and it is also easy to use input/output value of finite difference time domain to identify rapidly the state of controlled object so as to implement the online adjustment to the control law and need for repeated optimization .



(a) Data for Training (b) Convergence Diagram for Training

Figure 5. Neural Network Model Training Results of Manipulator

Optimization algorithm in this article also uses neural network to achieve, set the time-domain involved in the optimization value of 2, using the BP network neural of hidden layer neuron number 7, the same learning rule Levenberg-Marquartdt do the online training to achieve the control signal to the continuous optimization. Training block diagram is shown in the dashed box in Figure 4. Neural network optimization device in accordance with a given input signal u produce predictable output u1, u1 is imposed to the neural network model of the controlled object to produce predictable output y1, y1 compare with the desired output u of the system, and both the difference to train the neural network optimization. Then, the output u2 of the e2 enough litter as the actual amount of control applied to the actual controlled object. Visible, the optimizer in the regulation system is the inverse model of the charged object. Y1 can also be compared with actual output y2, and the error e1 and

the actual input u2 of charged object, output y2 as the data of training charged object neural network model.

C. Feedback correction

Feedback correction is forecast control to keep the dynamic correction forecasting model to ensure that the prediction model with infinitely close to the actual controlled object, and make optimization algorithm establish on the basis of the correct prediction of the system state then the new optimization. Error e1 in Figure 6 is the amendment process of the neural network model of the controlled object. Neural network prediction model is built on the basis of the past run data in system, the new operating environment and the actual system has the nonlinear, time-varying, interference and other factors make prediction model based on neural networks need to constantly learn to modify their weights and even structure to ensure that it can well represent the actual controlled object to a control signal prediction.

V. SIMULATION ANALYSES

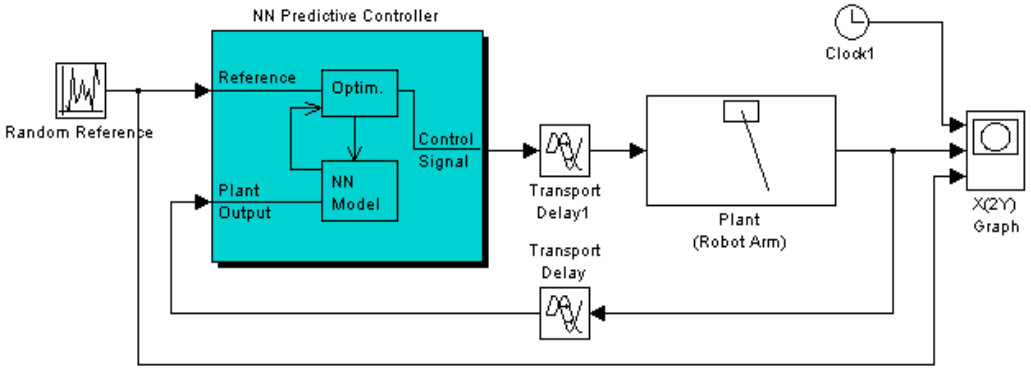
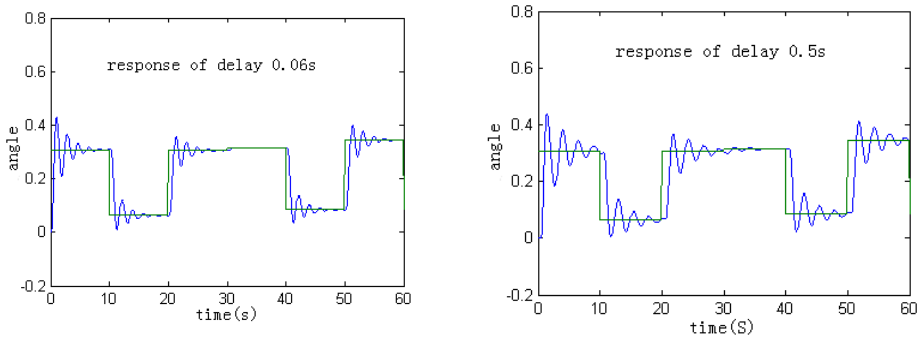


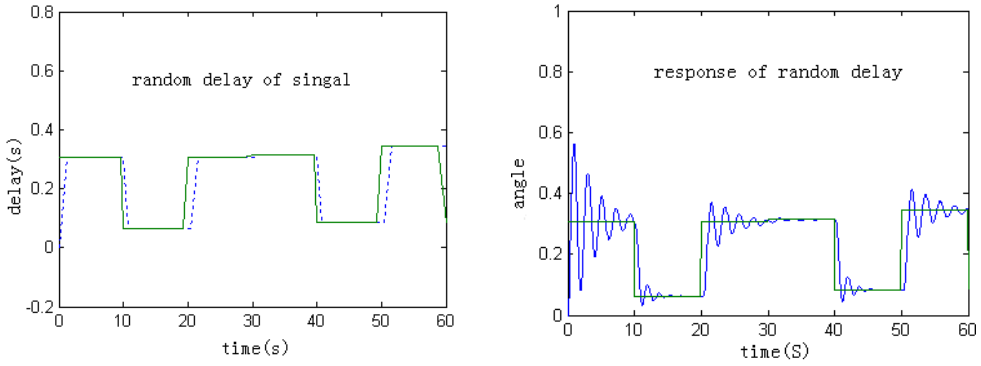
Figure 6. Simulink Simulation of Network Closed-Loop Control System based on Predictive Neural

Build the Simulation block diagram shown in Figure 6 under robotic arm Smulink environment, network training based on neural network predictive control by the steps in Figure 4-Figure5, and at the role of the same random input signal gradually adjust the value of delay to simulation. The results in Figure7 show that the prediction control based on neural network has a good control performance to the fixed delay network. Further used random delay module shown in Figure 8(a) instead of fixed delay module in Figure 5 immediately delay module for delay characteristics of input shown in Figure8 (b), where In.mat file stored random input signal in Figure 5.There are used random input signal stored in this file in order to compare the simulation results in the simulation. Finally, simulation under the random delay conditions and results shows in Figure8 (c). Whether a fixed delay or random delay neural network predictive controller can satisfy the closed-loop control requirements in the network delay conditions.



(a) Response Curve delay 0.06s (b) Response Curve delay0.5s

Figure 7. Predictive Control Random Responses Curve based on Neural Network



(a) Delay Curve under Random Delay (b) Response Curve under Random Delay

Figure 8. Responses under Random Delay

VI. Conclusion

This article discusses the difficulties of remote closed-loop control, that is different from the general control system of the difficulty lies in channel and feedback channel network of system existence uncertain delay which greatly reduces the stability of system and improve the design difficulty of control system. This paper elaborated network closed-loop control problems form uncertain network delay to includes network delay controller design method, and study the impact of network transmission delay on the network closed-loop control system, proposed by predictive control based on neural network to solve feasibility of the network control system which existence random delay closed-loop control, and verified the validity of the method by simulation.

Acknowledgment

The authors wish to thank the cooperators. This research is partially funded by the Project funds in.

National Network Engineering Testing LabFund project(GSYSJ2018012).

References

[1] Zhang Wei, Michael S,Branicky, Stephen M, Phillips, Stability of Networked Control Systems[J], IEEE Control Systems Magazine February,2001,(21):84-99.

[2] Almutairi Naif B, Chow Moyuen, PI Parameterization Using Adaptive Fuzzy Modulation (AFM) for Networked Control Systems-Part I: Partial Adaptation [J].IEEE Proceedings of IECON 2008, Sevilla, Spain, 2008. 3152-3157.

[3] Goodwin C, Juan Carlos Aguero,Arie Feuer, State Estima2tion for Systems Having Random Measurement Delays UsingErrors in Variables[C], The 15th Triennial World Congress Barcelona, Spain, 2002.

[4] Lee Kyung Chang,Lee Suk, Remote Controller Design of Networked Control System Using Genetic Algorithm[C], ISIE 2007, Pusan, KOREA in IEEE, 2007: 1845-1850.

[5] Huang J Q, Lew is FL, Liu K A, Neural predictive control for telerobot swith time delay [J]. Journal of Intelligent and Robotic System s, 2000, 29:1- 25.

[6] Lian Fengli Analysis, Design, Modeling, and Control of Networked Control Systems[D], Ph.D. thesis, The University of Michigan, 2001.

[7] FerrelW R.Remote manipulation with transmission delay[J]. IEEE Transaction on Human Factors in Electronics, 1965,FE6-(1):24-32.

[8] Halevi Y, Ray A.Integrated communication and control systems; Part I analysis[J]. Journal of Dynamic Systems Measurement and control, 1988,110(4):367-373.

[9] Walsh G C,Beldiman O,Bushnell L G. A symptotic behavior of nonlinear networked control systems[J]. IEEE Transactions on Automatic Control, 2001,46(7):1093-1097.

[10] Gregory C Walsh,Octavian Beldiman,Linda Bushnell.Error encoding algorithms for networdked control systems[C]. Proceedings of the 38th Conference on Decision and Control Phoenix, 1999,5:4933-4938.

[11] Göktas F.Distributed control of systems over communication networks[D]. Ph.D Dissertion. Philadelphia, PA, USA:University of Pennsylvania,2000.

[12] Sun Haiyi, Li Ning. Robust passive control of long delay network control system [J]. Computing Technology and Automation, 2007, 26 (4) :5-8.

[13] Zhu Zhangqing, Zhou Chuan, Hu Weili. Robust HH state observer design of short delay network control systems[J]. Control and Decision, 2005,20 (3)

[14] Zhang Ximin, Li Jiandong, Zhang Jianguo. Robust guaranteed cost control of network control systems [J]. Xi'an Electronic Technology University, 2008, 35 (1):96-100.

[15] Huang J Q , Lew is FL, Liu K A, Neural predictive control for telerobot swith time delay [J]. Journal of Intelligent and Robotic System s, 2008, 29:1- 25.

[16] Chen, S, C.F.N. Cowan, and P.M. Grant, Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks[J], IEEE Transactions on Neural Networks, 1991(2): 302-309.

[17] Huang J Q, Lew is FL, Liu K A, Neural predictive control for telerobot swith time delay[J]. Journal of Intelligent and Robotic System s, 2000,29:1- 25.

[18] Chen, S, C.F.N. Cowan, and P.M. Grant, Orthogonal Least Squares Learning Algorithm for Radial Basis Function Networks, IEEE Transactions on Neural Networks, 1991(2): 302-309.

Internet of Things Application in Satellite Communication in Power Transmission, Transform and Distribution

Dong Chuang
Shaanxi Huabiao Network Technology Co., Ltd.
Xi'an, 610041, China
E-mail: james9894@163.com

***Abstract*—In the national grid enterprise system, the geographical environment of transmission, substation lines and equipment distribution is very complicated. Transmission lines and substation equipment built in complex environments such as plateaus, forests, canyons and borders are often subjected to earthquakes, floods and blizzards. The threat of mudslide disasters, the ground communication network is greatly affected by extreme natural disasters, and it is impossible to repair and rescue in time when disasters occur. In response to the needs of facility operation and maintenance under emergency anomalies, Tiantong No. 1 satellite is a communication satellite with independent intellectual property rights in China. It has functions such as voice communication, image transmission, positioning, and data reporting. It can set up satellite Internet of Things to realize real-time monitoring and equipment. Conveniences such as inspections and emergency rescues reduce the operation and maintenance costs of the power transmission and transformation process.**

Keywords-Tiantong Satellite; Audio And Video Transmission; Inspection; Satellite Ground Station; Ground Network

I. INTRODUCTION

The Ubiquitous Electricity Internet of Things connects people , things and power users , power grid enterprises ,power generation enterprises , suppliers and their equipment to generate shared data for users, the grid, power generation, and suppliers And government social services; use the power grid as a hub, play a platform and sharing role, create value for the development of the industry and more market players, and provide value services.

The essence of the ubiquitous electric power Internet of Things is to fully apply modern information technologies such as "big data, cloud platforms, the Internet of Things, mobile Internet, artificial intelligence" to create a smart service system with comprehensive status perception, efficient information processing, and convenient and flexible application

In January 2019, the State Grid first proposed the strategic goal of "three types (hub type, platform type, shared type) two networks (Strong Smart Grid, Ubiquitous Electricity Internet of Things)" in the "two sessions" report, and proposed the construction The important material foundation of a world-class energy Internet company is to build and operate the "two networks." In March 2019, the State Grid clarified the definition of the ubiquitous electric power IoT for the first time at a teleconference on the deployment of ubiquitous electric power IoT construction work, and proposed a two-phase strategic arrangement for the construction of the next five years. The most urgent and important task is to accelerate the construction of the ubiquitous electric power Internet of Things.

At present, a large number of devices and applications have been developed in the application layer, platform layer, network layer and perception layer. Such IoT devices have been deployed in areas covered by terrestrial networks with stable business

capabilities and reasonable use costs. The development of services in areas without ground network signal coverage urgently needs to be improved, and current ground network and IoT equipment cannot support this demand.

Due to the characteristics of the satellite network: covering no dead zones, communication distance has nothing to do with cost, convenient networking, safe and reliable communication, and can be used as a last-guaranteed communication method in the event of ground network paralysis caused by earthquakes, floods, and snowstorms, Particularly suitable as a supplementary application in the power industry. The ubiquitous electric power Internet of Things-related services, in areas where there is no ground network signal coverage, or where new ground networks cover higher cost areas, the most effective technical implementation is achieved by satellite communications using satellites as relays. Ubiquitous power satellite IoT solution.

The application of satellites to electricity is accompanied by the rapid development of satellite networks, combined with the objective communication needs of the power industry in areas without ground network coverage. The ubiquitous power satellite IoT will be an extension and supplement of the ubiquitous power IoT. Add satellite communication capabilities on the basis of existing power IoT equipment to ensure that it meets the needs of power applications and expands the use of locations and scenarios. In short, it extends from areas with terrestrial networks to areas without ground network coverage.

This article will analyze the application of ubiquitous power satellite IoT in the power system's power generation, transmission, substation distribution and power consumption process.

II. OVERVIEW OF SATELLITE COMMUNICATIONS IOT TECHNOLOGY SUPPORT

A. Satellite communications overview

Satellite mobile communication refers to the use of artificial earth satellites as relay stations to forward radio waves used for communication between mobile users or between mobile users and fixed users to achieve mobile communication between two or more points. Satellite communications generally use the L, S, C, X, Ku, and Ka frequency bands, while high-throughput satellites generally use the C, Ku, and Ka frequency bands. However, resources in the C and Ku frequency bands are tight, so currently Qualcomm satellites are increasingly The development of Ka band, the frequency resources of Ka band are more abundant. Typical satellite mobile communication systems include space segment, ground segment and user segment. The space segment is composed of one or more satellite constellations. As a communication relay station, it provides the connection between the network user and the gateway. The ground segment usually includes the gateway, the network control center, and the satellite control center. Operation; The user segment is composed of various user terminals. There are mainly two types of terminals-mobile terminals and handheld terminals.

New satellite technology has developed rapidly, and there are many mature commercial systems, such as Tiantong, High-throughput, Iridium, etc. There are also many satellite newcomers who will launch thousands of satellites and put them into operation in the next few years. Figure 2 Basic composition of

satellite.

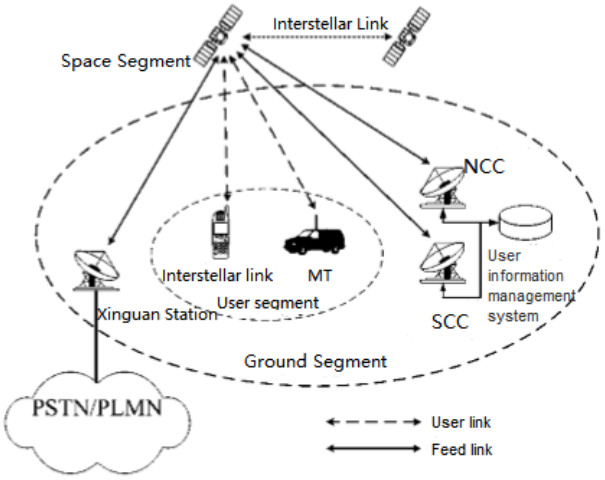


Figure 1. Basic composition of satellite communication link

B. *Satellite communication applications*

Mobile satellite communications are an effective complement to terrestrial cellular mobile communications. In addition to the characteristics of mobile communications, it also has the inherent advantages of satellite communications, including: ① long communication distances, large coverage areas, and multi-address connectivity. A geosynchronous orbit communication satellite can cover 42% of the earth's surface. The large communication distance between two points on the ground is about 18 000 km, and all earth stations in the area covered by the satellite can use the same satellite to communicate with each other. ② Large communication capacity and many types of transmission services. Typical communication satellites have a capacity of tens to hundreds of megabits per second and can serve hundreds of video channels or tens of thousands of voice and data links. ③ Good communication quality and low line bit error rate. Since the radio waves of mobile satellite communications mainly propagate in the space beyond the atmosphere, the propagation characteristics are relatively stable, and they are not easily affected by natural conditions and interference. The normal operating rate of satellite communications is above 99.8%, and the transmission quality is high. ④ The "on-the-move communication" of the mobile platform can be realized. Diverse users of mobile satellites, satellites

Terminals can be mobile carriers located on the ground, at sea, in the air or even in space. It is combined with terrestrial cellular mobile communication systems and other communication systems to form a global coverage seamless communication network. In addition to the above advantages, satellite mobile communication also has the following constraints: ① the size, weight and power consumption of mobile terminal equipment are limited, the size and shape of the antenna are limited by the installed carrier (such as aircraft, cars, ships, etc.), handheld terminals The requirements are even more demanding. ② The satellite antenna beam should be able to adapt to changes in the ground coverage area and keep pointing. The antenna beam of the user's mobile terminal should be able to keep pointing to the satellite as the user moves, or an omnidirectional antenna beam. ③ Due to the movement of the mobile body, when the link between the mobile terminal and the satellite transponder is blocked, a “shadow” effect will occur, causing communication to be blocked. ④ A satellite constellation system composed of multiple satellites requires the establishment of inter-satellite communication links and on-board processing and on-board exchanges, or the establishment of a gateway station with exchange and processing capabilities.

C. *Development Status of Satellite Communication System*

1) *Tiantong Satellite*

The Tiantong-1 satellite mobile communication system is composed of a space segment, a ground segment and a user terminal, and the space segment is planned to consist of multiple geosynchronous orbit communication satellites. As the first star of China’s satellite mobile communication system, Tiantong-1 was launched on August 6, 2016. It belongs to China

Satellite Communications Group Co., Ltd. It is mainly developed by China Academy of Space Technology and uses new plastic antennas. New equipment and technologies such as stand-alone integrated technology, the communication frequency is designed in the S band, and the cellular technology with a bandwidth of 30 MHz can form hundreds of spot beams. The signal transmission loss is small and the communication quality can be effectively guaranteed. At the same time, it uses communication satellite frequency multiplexing technology and a large-scale expandable antenna on board, which greatly improves the sensitivity of satellite receiving signals and increases the capacity of satellite communications.

The area covered by the 01 star of Tiantong No. 1 is mainly China and its surroundings, the Middle East, Africa and other related regions, and most of the Pacific Ocean and the Indian Ocean. There are no restrictions on the coverage of the terrain, and the ocean, mountains, Shangyuan, forests, Gobi, and desert can be seamlessly covered. Covers all types of mobile users including vehicles, airplanes, ships, and individuals. It provides all-weather, all-day, stable and reliable mobile communication services for various fields such as personal communications, marine transportation, ocean fishing, aviation rescue, and tourism research. When natural disasters occur in voice, short message and data services, the emergency communication capabilities of Tiantong One can play a great role. In addition, the main advantages of Tiantong One 01 are reflected in the miniaturization and mobile phone of the terminal, which is easy to carry.

The ground service of Tiantong No. 01 is operated by China Telecom Corporation, which will form a mobile communication network with ground mobile communication systems, and provide voice and data communication coverage for various handheld and small mobile terminals in China's land and surrounding sea areas. It is understood that China Telecom has launched an operation and plans to launch a mobile phone with mobile satellite communication capabilities. Even if it reaches the end of the earth, you can use it to talk to family and friends, send text messages, chat on WeChat, and communicate with video.

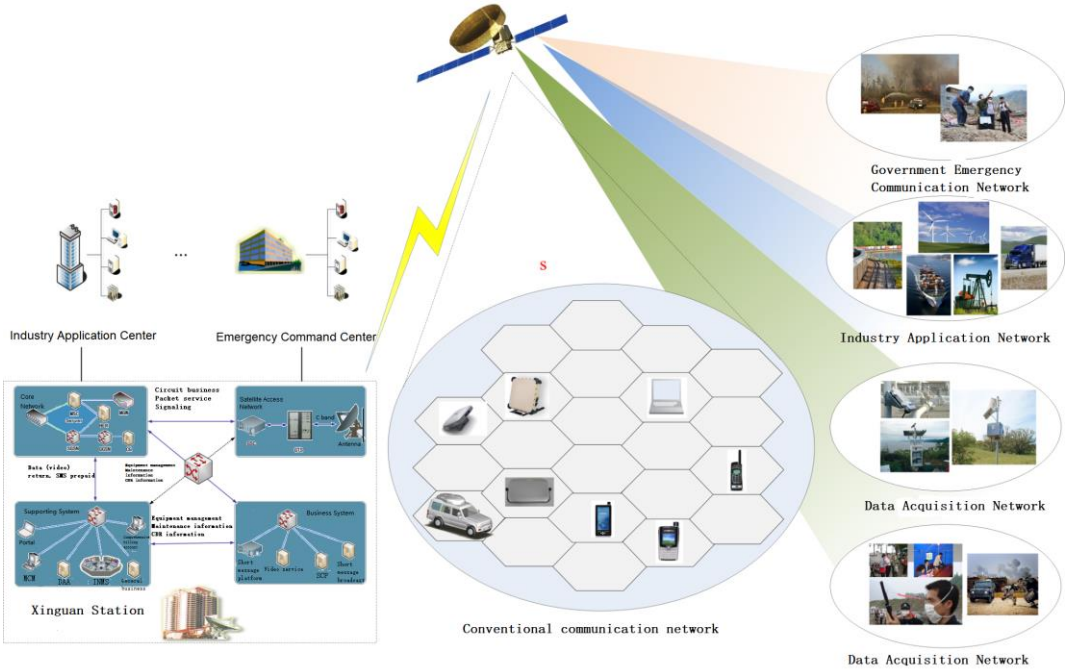


Figure 2. Tiantong-1 satellite mobile communication system

It is estimated that by 2025, China 's mobile communication satellite system will have more than 3 million end users, and its services will cover disaster relief, personal communications, marine transportation, ocean fishing, air passenger transport, bipolar scientific research, and international peacekeeping. During this period, we will also launch multiple Tiantong-1 satellites to further increase the satellite mobile communication service capacity and coverage area, and expand from the surrounding areas of China to form a regional mobile communication system integrating satellite and ground integration to achieve satellite mobile communication. Large-scale application and operation build an important support platform for the country's 'Belt and Road' strategy.

2) Iridium Satellite

1. The Iridium satellite system was the first LEO satellite cellular system with global coverage. It was launched by Motorola in the late 1980s and developed in the early 1990s. The "Iridium" star system includes space segment, ground segment and user segment[9]. The networking and coverage are shown in Figure 2. Space segment The initial design of the constellation consisted of 77 LEO satellites, evenly distributed in 7 polar orbits, and all satellites were moving in the same direction. It is similar to the 77 electrons of iridium orbiting the nucleus, hence the name of the system. The actual constellation consists of 66 satellites distributed on 6 circular near-polar orbital planes with an inclination of 86.4 °, with an interval of 27 ° and an orbital height of 780 km. Each satellite provides 48 spot beams, forming 48 cells on the ground. At a small elevation angle of 8.2 °, the diameter of each cell is 600 km, and the coverage area of each satellite is approximately 4700 km. The constellation forms seamless cellular coverage on the global ground. One spot beam of each satellite supports 80 channels, A single satellite can provide 3840 channels. It was officially operated on November 1, 1998. At present, it has more than 1.5 million users and uses 66 low-orbit satellites to cover the world. The characteristics are low orbit, fast transmission speed, small information loss, and greatly improved communication quality. The ground receiving station, equipment is miniaturized, and it is interconnected with the ground network, which makes the communication in the uninhabited, barren land, remote areas with backward communication, and the scene of natural disasters unobstructed.

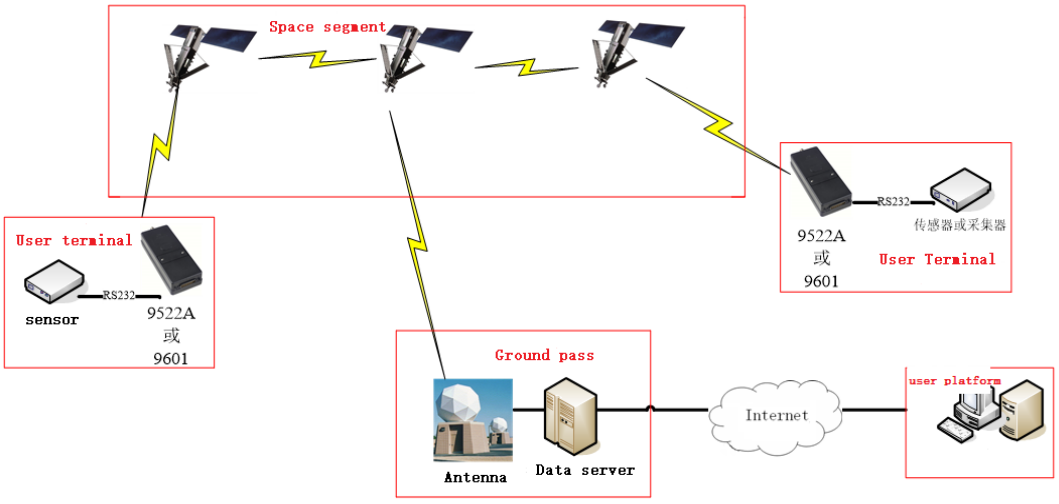


Figure 3. Iridium satellite system

3) High Throughput Satellite

The "Practice No. 13" satellite first applied Ka-band broadband technology to China's communication satellites, with a total communication capacity of 20Gbps, exceeding the total capacity of communication satellites developed and launched in China, marking that China's satellite communications has entered a high-throughput era At the same time, the "Thirteenth Practice" satellite carried out the first international two-way laser communication test between high-orbit satellites and the ground, with a speed of up to 5Gbps, which established China's global leading position in the field of high-speed space information transmission. The "Thirteenth" satellite is China's first electric propulsion satellite. Compared with conventional chemical propulsion satellites, its launch weight is greatly reduced, which reduces the requirements for the carrying capacity of the launch vehicle. The amount of data will no longer be a constraint on the life of the satellite, and the design life of the communication satellite will generally exceed the current 15-year limit and reach 18 to 20 years.

The "Practice No. 13" satellite will mainly provide services for users in China and other regions. It can realize the access of mobile communication base stations in remote areas and be used in the fields of enterprise private networks, distance education, medical treatment, digital news gathering and emergency communications. At the same time, the star can facilitate users to quickly access the network, download and return rates up to 150Mbps and 12Mbps, which can effectively meet the needs of passengers on the high-speed vehicles such as aircraft, high-speed rail to access the Internet anytime, anywhere.

High-throughput satellites are an important part of satellite communications and broadcasting systems in China's integrated space-ground information network. They are an effective complement to terrestrial communication systems and can be widely used in areas where terrestrial communication systems are difficult to cover or where construction costs are high. After the completion of the “Global Coverage, On-Demand Access, On-Demand Services, Safe and Trustworthy” integrated information network in the world, China will have global space-time continuous communication, highly reliable and secure communication, regional large-capacity communication, and high-mobility full-range information transmission. And other capabilities to meet national strategic needs such as expanding national interests, safeguarding national security, safeguarding national economy and people's livelihood, and promoting economic development. In view of the strategic significance of high-throughput satellites in the space and information fields, China will continue to deploy ultra-high-capacity high-throughput satellites. It is expected that by 2020, it will form a communication capacity that can cover the entire territory of China and the Asia-Pacific region with a capacity of nearly 200Gbps. To meet the urgent needs for broadband communications in the construction work of "Broadband China" and the country's "Belt and Road", the relevant industrial chain is expected to usher in rapid development.

4) 'Hongyan' Project

Construction goal: "Communicate and connect everything, the world will never lose touch"

It is planned to invest 100 billion yuan in 300 low-orbit satellites. The first test star was launched on December 29, 2018. The backbone constellation system is expected to be built in 2023. Its satellite data exchange function can provide two-way, real-time data transmission worldwide, as well as multimedia data services such as short messages, pictures, audio, and video. After the system is completed, services such as intelligent terminal communications, Internet of Things, mobile broadcasting, navigation enhancement, aviation and maritime surveillance, and broadband Internet access will be launched. The system has a

global real-time communication capability under all-weather and full-time terrain conditions.

5) 'Hongyun' Project

China Aerospace Science and Industry plans to launch 156 satellites. The satellites will be networked in orbits 1,000 kilometers above the ground to build a space-borne broadband global mobile internet network. The first test star was launched on December 22, 2018.

The entire Hongyun project is divided into three phases. The first phase is to launch the first star at the end of 2018; the second phase is to launch four business test stars at the end of the "Thirteenth Five-Year Plan", or before the end of 2020; the third stage is to "Ten In the middle of the 4th Five-Year Plan period, around 2023, 156 satellites will be launched, and the construction of the heaven-earth integration system will be completed initially, with full operating conditions. After the entire plan is completed, it will realize Internet access at anytime, anywhere, on the "Belt and Road" and even globally, and achieve a high-speed, high-quality Internet / IoT application experience.

6) OneWeb of United States

One Web plans to deploy 648 small satellites in low-Earth orbit to form an Internet satellite network, enabling users around the world to connect to satellites and access 3G, LTE, 5G, and Wi-Fi networks through small, low-cost user terminals. It is characterized by multiple satellites, interconnection with the Internet, and a single ground terminal download rate of 50Mbps.

7) SpaceX of United States

SpaceX Starlink:7, 518 low-Earth orbit satellites not exceeding 346 kilometers in height (Launch of Microsat-2a and Microsat-2b in early 2018)

III. Application of Satellite Communication in the Process of Power Transmission, Transformation and Distribution

A. Transmission fault monitoring system

The transmission fault monitoring system, consisting of tower poles, sensor cluster, satellite wireless transceiver, power monitoring center, and mobile phone of the responsible person, to realize the status monitoring, information reporting, and fault early warning of the transmission line. The transmission fault monitoring system is a supplement to the online monitoring system of the entire smart grid transmission line. By using satellite technology, it provides strong support in areas that cannot be covered by existing monitoring systems.

The business capability is increased to realize the visual real-time monitoring and fault early warning of various states of the transmission line without the coverage of the ground network. The monitoring of the status of important electrical devices on the tower pole, such as circuit breakers, load switches, section switches, and other switch monitoring and fault early warning on 10kV lines. The logic diagram of the high-level scheme of the transmission fault monitoring system is shown below:

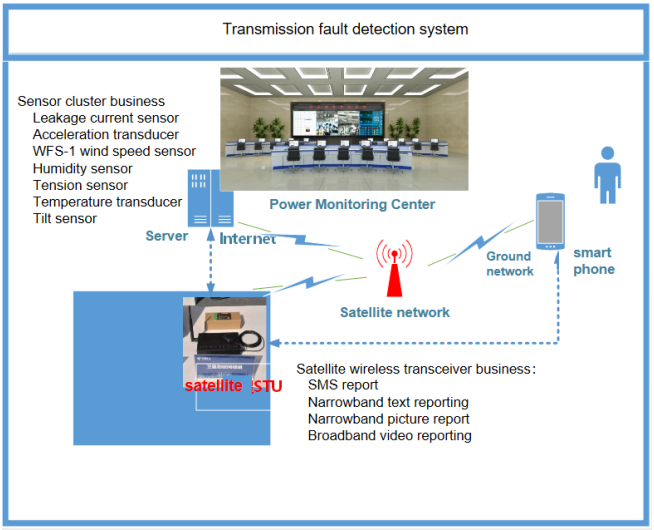


Figure 4. The logic diagram of the high-level scheme of the transmission fault monitoring system

The description of each part of the transmission fault monitoring system is as follows:

Satellite STU: It acts as an information transmission channel, and is particularly suitable for routine monitoring scenarios in areas without ground network coverage and disaster prevention and relief scenarios in areas without ground network coverage.

Power monitoring center: It is the existing monitoring center of the power system, not a newly deployed system platform, in order to save customers' capital investment budget to the greatest extent.

Responsible person's mobile phone: When an important failure occurs, the corresponding responsible person's personal mobile phone is notified by SMS to speed up the progress of failure handling and reduce accident losses. Of course, the level of important faults and the mobile phone number of the person in charge can be set before the implementation of the plan, which will not cause excessive SMS business and bring unnecessary pressure on the Responsible person

Sensor clusters: The main business capabilities realized in the entire scheme, that is, the types of transmission line status collection, depend on the configuration of the sensor clusters.

B. Distribution fault monitoring system

Transformer fault monitoring system, consisting of tower poles, sensor clusters, satellite STU, power monitoring center, and mobile phone of the responsible person, realize the status monitoring, platform docking, information reporting and fault early warning of distribution and transformation links.

The power distribution fault monitoring system is a solution to the fault monitoring and early warning requirements of physical nodes in the power distribution and transformation business that need to increase communication support. Conventional equipment on the ground network can partially address these needs, but cannot guarantee emergency communications during natural disasters. The terminal equipment STU used in the power distribution fault monitoring system is a combination of terrestrial network DTU, terrestrial network TTU, and satellite terminal technology, which can improve the reliability in the process of power distribution and transformation. Considering the cost of satellite service, it is recommended to construct and install it in important substations and distribution stations, but not in ordinary stations.

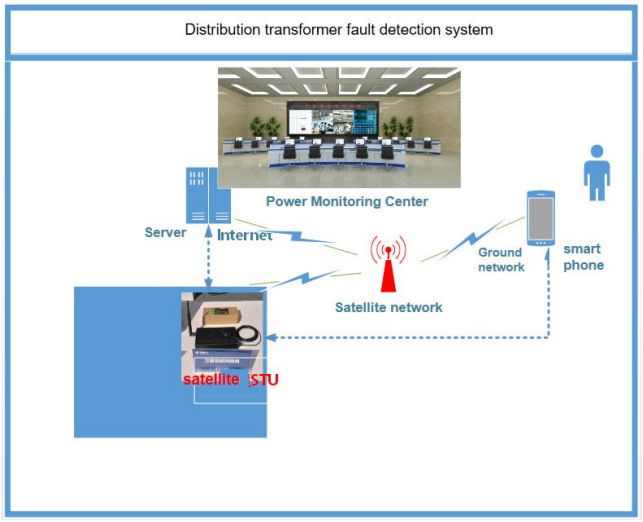


Figure 5. The logic diagram of the high-level scheme of the Distribution fault monitoring system

Satellite STU: It acts as an information transmission channel, and is particularly suitable for routine monitoring scenarios in areas without ground network coverage and disaster prevention and relief scenarios in areas without ground network coverage.

Power monitoring center: It is the existing monitoring center of the power system, not a newly deployed system platform, in order to save customers' capital investment budget to the greatest extent.

Responsible person's mobile phone: When an important failure occurs, the corresponding responsible person's personal mobile phone is notified by SMS to speed up the progress of failure handling and reduce accident losses. Of course, the level of important faults and the mobile phone number of the person in charge can be set before the implementation of the plan,

which will not cause excessive SMS business and bring unnecessary pressure on the Responsible person

The satellite mobile communication system can quickly form an emergency communication network in an emergency area without relying on the original communication network, enabling the field headquarters and the command center to quickly establish communication links, and provide a certain bandwidth transmission rate, while transmitting voice, images, and data And other information, supporting real-time transmission of multi-directional dynamic images, high-definition video conferences, etc. Therefore, the use of satellite mobile communication means can realize emergency communications support for emergency rescue and disaster relief in large-scale disasters and meet the requirements of "all-weather, whole-process, all-round" emergency communications support, which is the future development trend of emergency communications.

IV. CONCLUSION

In the event of a very catastrophic disaster such as a blizzard, hurricane, earthquake, mudslide, etc., public communication network facilities may be catastrophically damaged and paralyzed. At this time, the communication of the emergency report and rescue command is urgent, which is convenient for effectively monitoring the field power transmission and transformation facilities, and at the same time, it can achieve rapid and effective rescue in the event of a disaster. The monitoring program uses China's Tiantong satellite with independent intellectual property rights to establish a remote monitoring system for power transmission and transformation facilities based on the Internet of Things of Satellites. This system has high security and anti-interference, and can effectively solve remote monitoring of the wild environment. , Data communication, positioning and other problems, to facilitate better operation and maintenance of power transmission and transformation facilities, reduce the failure rate of facilities, improve accident repair capacity, rescue timeliness and power supply reliability.

REFERENCES

[1] Lv Ziping, Liang Peng, Chen Zhengjun. Development Status and Trends of Satellite Mobile Communications[J]. Satellite Application, 2016(01):48-55.

[2] Liu Yue. Review on the Development of GEO Mobile Communication Satellite Technology[J]. International Space, 2013(10):26-31.

[3] Xu Feng, Chen Peng. Development Progress and Trends of Foreign Satellite Mobile Communications[J]. Telecommunication Engineering, 2011, 51(06):156-161.

[4] He Shanbao. Inmarsat System and Its New Development[J].Space International, 2009(09):31-34.

[5] Andrew D S, Paul S. Utilizing the Globalstar Network for Satellite Communications in Low Earth Orbit[C].54th AIAA Aerospace Sciences Meeting, 2016:1-8.

[6] Huang Huashan. Application and Development of Satellite Mobile Communication[J].China High-Tech Enterprises, 2014(26):9-11.

[7] MA Chong, ZUO Peng, WANG Zhen-hua. Inmarsat System and its Application in Emergency[J]. Satellite Application, 2009(06):31-35.

[8] Zhu Lidong. Review on Development and New Technologies of Military Satellite Communications Aboard[J].Radio Communications Technology, 2016, 42(05):1-5.

[9] Lv Ziping, Zhang Gengxin, Liang Peng. The Development and Orientation of China's GEO and LEO Mobile Satellite Communications[J].Digital Communication World, 2016(07):26-28.

[10] Xiao Longlong, Liang Xiaojuan, Li Xin. The Development and Application of Satellite Mobile Communication System.

Improved Stereo Vision Robot Locating and Mapping Method

Yu Haige
School of computer science and engineering
Xi'an Technological University
Xi'an, Shaanxi, China
E-mail: 279084342@qq.com

Wei Yanxi
School of computer science and engineering
Xi'an Technological University
Xi'an, Shaanxi, China
E-mail: 407171251@qq.com

Yu Fan
School of computer science and engineering
Xi'an Technological University
Xi'an, Shaanxi, China
E-mail: yffshun@163.com

***Abstract*—Vision-based SLAM has an outstanding problem is not work when the camera fast motion, or camera operating environment characterized by scarce. Aiming at this problem, this paper proposes a SLAM method of IMU and vision fusion. This article uses a stereo camera to extract the image ORB feature points. During the camera movement, if the number of extracted feature points is less than a certain threshold and the camera movement cannot be estimated or the estimated camera rotation and translation is greater than a certain threshold, the camera pose is estimated by fusing IMU , Otherwise use feature points to estimate camera pose. This paper uses non-linear optimization methods to perform pose estimation of pure feature points and pose estimation of fused IMU, respectively. The experimental results show that binocular vision SLAM with IMU information can estimate the camera pose more accurately.**

Keyword-Robot; IMU; Stereo Vision; SLAM

I. INTRODUCTION

With the development of robot technology, more and more robots are approaching our lives, such as sweeping robots, shopping mall robots, etc. Mobile robots are the product of the cross fusion of various disciplines and technologies. Among them, SLAM(Simultaneous Localization and Mapping) is an important technology for mobile robots. SLAM means that the robot builds a map of the surrounding environment in real time based on sensor data without any prior knowledge, and infers its own positioning based on the map. From the 1980s to the present, more and more sensors are used in SLAM, from early sonar, to later 2D/3D lidar, to monocular, binocular, RGBD, ToF and other cameras. Compared with lidar, cameras used in vision SLAM have become the focus of current SLAM research due to their advantages such as low price, light weight, large amount of image information, and wide application range. Stereo cameras generally consist of two pinhole cameras placed horizontally. Compared to monocular vision's scale uncertainty and pure rotation problems, binocular cameras can directly calculate the pixel depth. At the same time, compared to RGB-D cameras, stereo cameras collect images

DOI: 10.21307/ijanmc-2019-070

directly from ambient light and can be used indoors and outdoors. Compared with lidar, the main disadvantage of the camera as a SLAM sensor is that when the camera moves too fast, the camera will blur images, and the camera will not work in a scene with insufficient environmental feature textures and few feature points.

Aiming at the problems of the above-mentioned visual SLAM system, this paper proposes an algorithm that fuses IMU and SLAM. Through the fusion of IMU, it can provide a good initial pose for the system. At the same time, during the camera movement process, it makes up for the shortcomings of visual SLAM, ensuring the accuracy of the camera pose estimation in the case of fast camera movement and lack of environmental texture.

II. RELATED WORKS

A. Camera coordinate system

Camera models generally have four coordinate systems: a pixel coordinate system, an image coordinate system, a world coordinate system, and a camera coordinate system. Figure 1:

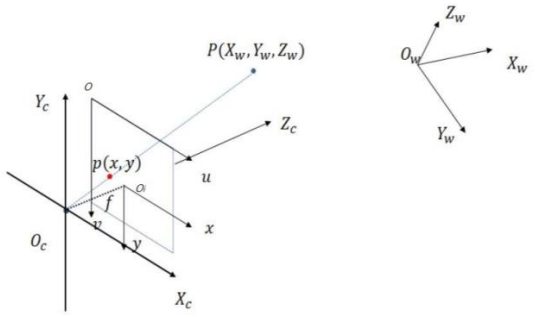


Figure 1. Camera related coordinate system

Among them, $O_w - X_w Y_w Z_w$ is the world coordinate system. The world coordinate system is the reference coordinate system in the visual SLAM system. The positions of the camera trajectory and map points are described based on this coordinate system. The unit is m.

$O_i - xy$ is the image coordinate system. The image coordinate system uses the intersection of the camera optical center and the image plane coordinate system as the origin. The unit is mm.

$O_c - X_c Y_c Z_c$ is the camera coordinate system. The camera coordinate system uses the camera optical center as the origin, and the directions parallel to the x-axis and y-axis of the image coordinate system are respectively taken as the X_c-axis and Y_c-axis, and the direction perpendicular to the image plane is the Z_c-axis. The unit is m.

$O-uv$ is the pixel coordinate system. The origin of the pixel coordinate system is generally the upper left corner of the image, with the u axis to the right parallel to the x axis, and the v axis to the y axis. The unit is pixel.

B. Camera projection model

The camera maps the coordinate points of the three-dimensional world to the two-dimensional image plane. This process is generally a pinhole model. Under the pinhole model, it is assumed that there is a spatial point P, and the coordinates of the point P are $[X, Y, Z]^T$. After the projection of the small hole O, the point P falls on the imaging plane $o-xy$, and the imaging point is p, The p-point coordinate is $[x, y, z]^T$. Let the distance from the imaging plane to the small hole be the focal length f. Therefore, according to the principle of triangle similarity, there are:

$$\frac{Z}{f} = \frac{X}{x} = \frac{Y}{y} \tag{1}$$

So we can get:

$$\begin{cases} x = f\dfrac{X}{Z} \\ y = f\dfrac{Y}{Z} \end{cases} \tag{2}$$

The difference between the pixel coordinate system and the imaging plane is a zoom and a translation of the origin. Suppose that the pixel coordinates are scaled α times on the u axis and β times on the v axis, and the origin is translated $\left[c_x, c_y\right]^T$, so we can get:

$$\begin{cases} u = \alpha x + c_x \\ v = \beta y + c_y \end{cases} \tag{3}$$

Equation (3) is substituted into equation (2) to get:

$$\begin{cases} u = f_x\dfrac{X}{Z} + c_x \\ v = f_y\dfrac{Y}{Z} + c_y \end{cases} \tag{4}$$

The unit of f is m and the unit of α and β is $pixel/m$, so the unit of f_x, and f_y is $pixel$. Written as a matrix:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \frac{1}{Z}\begin{pmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{pmatrix}\begin{pmatrix} X \\ Y \\ Z \end{pmatrix} \triangleq \frac{1}{Z}\mathbf{KP} \tag{5}$$

Among them, the matrix K is called the internal parameter matrix of the camera, and P is the coordinate representation of the space point in the camera coordinate system.

Let the coordinate P of the space point in the camera coordinate system correspond to the coordinate P_w in the world coordinate system, and use coordinate transformation to obtain:

$$Z\mathbf{P}_{uv} = Z\begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \mathbf{K}(\mathbf{R}\mathbf{P}_w + \mathbf{t}) = \mathbf{K}\mathbf{T}\mathbf{P}_w \tag{6}$$

Among them, T represents the pose of the camera relative to the world coordinate system, and can also be called the external parameter of the camera. In summary, the pinhole camera model uses the triangle similarity relationship to obtain the relationship between space points and pixels, which is a relatively ideal model. In practice, there will be errors in the manufacture and installation of optical lenses, which will affect the propagation of light during the imaging process and cause distortion in the images collected by the camera. Here we mainly consider radial distortion and tangential distortion.

Radial distortion is caused by the shape of the lens, and the distortion increases as the distance between the pixel and the center of the image increases. Therefore, a polynomial function can be used to describe the changes before and after the distortion, that is, the quadratic and higher-order polynomial functions related to the distance between the pixel and the center of the image can be used for correction. The polynomial of the coordinate change before and after the radial distortion correction is as follows:

$$\begin{cases} x_{corrected} = x(1 + k_1 \mathrm{r}^2 + k_2 \mathrm{r}^4 + k_3 \mathrm{r}^6) \\ y_{corrected} = y(1 + k_1 \mathrm{r}^2 + k_2 \mathrm{r}^4 + k_3 \mathrm{r}^6) \end{cases} \tag{7}$$

Among them, $\left[x, y\right]^T$ is the coordinates of the uncorrected points, and $\left[x_{corrected},\ y_{corrected}\right]^T$ is the coordinates of the points after the distortion is corrected. r is the distance from the point (x, y) to the origin. k_1, k_2 and k_3 are three radial distortion

parameters. Usually these three parameters can be obtained by the calibration step.

For tangential distortion, the reason is that the lens and the imaging plane cannot be strictly parallel during camera assembly. Tangential distortion can be corrected using two other parameters, p1 and p2:

$$\begin{cases} x_{corrected} = x + 2p_1xy + p_2(\mathrm{r}^2 + 2x^2) \\ y_{corrected} = y + 2p_2xy + p_1(\mathrm{r}^2 + 2y^2) \end{cases} \quad (8)$$

Considering the two types of distortion, we can find the correct position of a pixel in the pixel coordinate system through 5 distortion coefficients:

$$\begin{cases} x_{corrected} = x(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_1xy + p_2(r^2 + 2x^2) \\ y_{corrected} = y(1 + k_1r^2 + k_2r^4 + k_3r^6) + 2p_2xy + p_1(r^2 + 2y^2) \end{cases} \quad (9)$$

In summary, the parameters describing the camera model mainly include: in the camera's internal parameter matrix, and distortion correction parameters.

C. Stereo camera ranging principle

The binocular camera generally consists of two pinhole cameras placed horizontally, and the two cameras observe an object together. The aperture centers of both cameras are located on one axis, and the distance between the two is called the baseline b of the binocular camera. There is an existing space point P, which is an image in the left-eye camera and the right-eye camera, and is denoted as P_L, P_R. Due to the presence of the camera baseline, these two imaging positions are different. Remember that the coordinates of the imaging on the left and right sides are x_L, x_R, which can be seen from the similarity of the triangles:

$$\frac{z - f}{z} = \frac{b - u_L + u_R}{b} \quad (10)$$

We can get:

$$z = \frac{fb}{d} \quad (11)$$

The above model is an ideal model, which aims to explain the principle of measuring the actual three-dimensional point depth of the binocular camera. In practical applications, due to factors such as manufacturing and installation, it is difficult to achieve that the imaging planes of the binocular cameras are strictly on the same plane and the optical axes are strictly parallel. Therefore, before using a binocular camera for measurement, it should be calibrated to obtain the left and right camera internal parameters and the relative position relationship between the left and right cameras.

III. POSE ESTIMATION ALGORITHM

At present, the fusion method of monocular vision sensor and IMU can be divided into two types: loose coupling and tight coupling[1]. Loose coupling is based on the vision sensor and IMU as two separate modules, both of which can calculate the pose information, and then fused by EKF[2] and so on. Tight coupling refers to the non-linear optimization of vision and IMU data to obtain pose estimates. Because tight coupling can make full use of each sensor's data, this paper uses tight coupling to fuse vision and IMU data. Firstly, the purely visual feature point pose estimation method is used to estimate the camera pose. Then, during the camera movement, if the number of extracted feature points is less than a certain threshold value, the camera movement cannot be estimated or the estimated camera rotation and translation are greater than a certain threshold value, The camera pose is estimated by fusing the IMU, otherwise feature points are still used to estimate the camera pose.

A. Pose estimation using pure visual information

The ORB (Oriented Fast and rotated Brief) algorithm was proposed by Ethan Rublee et al. In 2011[3]. The ORB feature is composed of the FAST

feature and the BRIEF descriptor. It adds orientation and scale invariance to the FAST feature. Features are described using binary BRIEF descriptors. When performing feature matching, the descriptors between feature points and feature points are compared. The binocular camera can directly obtain the corresponding 3D position of the pixel under the known pixel matching of the left and right camera images. Therefore, the stereo camera-based SLAM system can use the known 3D point and its projection match in the current frame to obtain the current camera pose without the need to solve camera motion using epipolar geometry[4].

This paper first uses the method of EPnP[5] to solve the camera pose. The EPnP pose solution method can more effectively use the matching point information, and iteratively optimize the camera pose. EPnP is known as the coordinates $\left\{P_i^w, i=1,2,...,n\right\}$ of n space points in the world coordinate system and their corresponding coordinates $\left\{P_i^c, i=1,2,...,n\right\}$ in the image coordinate system to solve the rotation matrix R and translation vector t of the camera movement.Set four non-coplanar virtual control points in the world coordinate system, whose homogeneous sitting marks are: $\left\{C_i^w \mid i=1,2,3,4\right\}$ .The relationship between the world coordinates of the space points and the control points is as follows:

$$P_i^w = \sum_{j=1}^{4} \alpha_{ij} C_j^w, with \sum_{j=1}^{4} = \alpha_{ij} = 1 \tag{12}$$

Once thevirtual control point is determined and the premise that the four control points are not coplanar, $\left\{\alpha_{ij}, j=1,...,4\right\}$ is the only one determined.In the image coordinate system, the same weighting sum relationship exists:

$$P_i^c = \sum_{j=1}^{4} \alpha_{ij} C_j^c$$

Substituting equation (13) into the camera model gives:

$$\forall i, s_i \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = \mathbf{K} P_i^c = \mathbf{K} \sum_{j=1}^{4} \alpha C_j^c = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \sum_{j=1}^{4} \alpha_{ij} \begin{bmatrix} x_j^c \\ y_j^c \\ z_j^c \end{bmatrix} \tag{13}$$

The image coordinates u_i, v_i in Equation (13) are known, so:

$$s_i = \sum_{j=1}^{4} \alpha_{ij} z_j^c \tag{14}$$

From equations (13) and (14):

$$\begin{cases} \sum_{j=1}^{4} \alpha_{ij} f_x x_j^c + \alpha_{ij} (c_x - u_i) z_j^c = 0 \\ \sum_{j=1}^{4} \alpha_{ij} f_y y_j^c + \alpha_{ij} (c_y - v_i) z_j^c = 0 \end{cases} \tag{15}$$

In order to obtain the coordinates of the 2D point into the camera coordinate system, it is assumed that α_{ij} in the camera coordinate system is consistent with α_{ij} in the world coordinate system, that is, to find the rotation and translation of the four control points. Solve linear equations:

$$\mathbf{M}\mathbf{X} = 0 \tag{16}$$

Among them, M is a $2n \times 12$ matrix, and $\mathbf{X} = [C_1^{cT}, C_2^{cT}, C_3^{cT}, C_4^{cT}]$ is a vector composed of 12 unknowns to be solved.

$$\mathbf{X} = \sum_{i=1}^{N} \beta_i v_i \tag{17}$$

v_i is the right singular vector of M, and the corresponding singular value is 0. Solve the $\mathbf{M}^T\mathbf{M}$ eigen value and eigenvector. The eigenvector with eigenvalue of 0 is v_i. N is the dimension of the $\mathbf{M}^T\mathbf{M}$ space, and β_i is the coefficient to be determined.

Depending on the position of the reference point, the spatial dimension of the matrix $\mathbf{M}^T\mathbf{M}$ may take the values 1,2,3,4. According to the same distance between the control points in the world coordinate system and the camera coordinate system, six constraints can be obtained, and the pending coefficients can be solved.

When N = 1, according to the constraints:

$$\left\|\beta v^{[i]} - \beta v^{[j]}\right\|^2 = \left\|C_i^w - C_j^w\right\|^2 \tag{18}$$

and so：

$$\beta = \frac{\sum_{[i,j]\in[1,4]} \left\|v^{[i]} - v^{[j]}\right\| \bullet \left\|C_i^w - C_j^w\right\|}{\sum_{[i,j]\in[1,4]} \left\|v^{[i]} - v^{[j]}\right\|^2} \tag{19}$$

When N = 2:

$$\left\|\beta_1 v_1^{[i]} + \beta_2 v_2^{[i]} - (\beta_1 v_1^{[j]} + \beta_2 v_2^{[j]})\right\|^2 = \left\|C_i^w - C_j^w\right\|^2 \tag{20}$$

Since β_1 and β_2 only appear in the equation as quadratic terms, let $\boldsymbol{\beta} = [\beta_1^2, \beta_1\beta_2, \beta_2^2]^T$, and use the vector $\boldsymbol{\rho}$ to represent all $\left\|C_i^w - C_j^w\right\|^2$, thus obtaining the equation:

$$\mathbf{L}\boldsymbol{\beta} = \boldsymbol{\rho} \tag{21}$$

Where $\mathbf{L}$ is a 6×3 matrix composed of v_1 and v_2.

When N = 3, $\mathbf{L}$ is a 6×6 matrix.

In summary, the coordinate solution of the reference point in the camera coordinate system can be obtained as the initial value of the optimization, the optimization variable is $\boldsymbol{\beta} = [\beta_1, \beta_2, ..., \beta_N]^T$, and the objective function is:

$$Error(\boldsymbol{\beta}) = \sum_{(i,j)s.t.i<j} (\left\|C_i^c - C_j^c\right\|^2 - \left\|C_i^w - C_j^w\right\|^2) \tag{22}$$

Optimize $\boldsymbol{\beta}$ corresponding to the smallest dimension of the error, get the vector $\mathbf{X}$, and restore the coordinates of the control point in the camera coordinate system. At the same time, the coordinates of the reference point in the camera coordinate system are obtained according to the centroid coordinate coefficient. Finally, according to the coordinates of a set of point clouds in the two coordinate systems, the pose transformations of the two coordinate systems are obtained. The solution steps are as follows:

a) Find the center point:

$$p_c^c = \frac{\sum p_i^c}{n}, P_c^w = \frac{\sum p_i^w}{n} \tag{23}$$

b) To the center point:

$$q_i^c = p_i^c - p_c^c, q_i^w = p_i^w - p_c^w \tag{24}$$

c) Calculate the H matrix:

$$\mathbf{H} = \sum_{i=1}^{n} q_i^c q_i^{w^T} \tag{25}$$

d) SVD decomposition of H matrix:

$$\mathbf{H} = \mathbf{U}\Sigma\mathbf{V}^T \tag{26}$$

e) Calculate the rotation R:

$$\mathbf{R} = \mathbf{U}\mathbf{V}^T \tag{27}$$

If R <0, then R(2,.) =-R(2,0).

f) Calculate displacement t:

$$t = p_0^c - \mathbf{R} p_0^w \tag{28}$$

Taking the results of EPnP solution as initial values, the method of g2o was used to optimize the pose of the camera nonlinearly. Construct the least squares problem and find the best camera pose:

$$\xi^* = \arg\min_{\xi} \frac{1}{2} \sum_{i=1}^{n} \left\| u_i - \frac{1}{s_i} K \exp(\xi^\wedge) P_i \right\|_2^2 \tag{29}$$

Among them, u_i is the pixel coordinates of the projection point, K is the camera internal reference, ξ is the camera pose, and P_i is the space point coordinate.

B. Camera pose estimation method based on IMU

The measurement frequency of the IMU is often higher than the frequency at which the camera collects pictures. For example, the binocular camera used in this article has a frame rate of up to 60FPS and an IMU frequency of up to 500Hz. The difference in frequency between the two results in multiple IMU measurements between the two frames. Therefore, in order to ensure the information fusion of the two sensors, it is necessary to pre-integrate [6] the data of the IMU. That is, only the IMU information between the two image moments is integrated to obtain the relative pose value, and the integration result is saved for later joint optimization.The IMU-based camera pose estimation method mainly includes three coordinate systems: the world coordinate system, the IMU coordinate system, and the camera coordinate system. All pose and feature point coordinates are finally expressed in the world coordinate system. During the calculation process, this article will convert the state quantity in the camera coordinate system to the IMU coordinate system, and then to the world coordinate system.In this article, the letter W is used to represent the world coordinate system, the letter B is used to represent the IMU coordinate system, R_{WB} is used to represent the rotation matrix from the IMU coordinate system to the world coordinate system, and p_{WB} is used to represent the translation matrix from the IMU coordinate system to the world coordinate system.

The acceleration and angular velocity of the IMU are:

$$\begin{aligned} {}_B\tilde{\omega}_{WB}(t) &= {}_B\omega_{WB}(t) + b^g(t) + \eta^g(t) \\ {}_B\tilde{\mathrm{a}}_{WB}(t) &= R_{WB}^T(t)({}_W\mathrm{a}(t) - {}_W g) + b^a(t) + \eta^a(t) \end{aligned} \tag{30}$$

Among them, $b^a(t)$ and $b^g(t)$ represent the bias of the accelerometer and gyroscope respectively, $\eta^a(t)$ and $\eta^g(t)$ represent the noise of the accelerometer and gyroscope respectively, and ${}_W g$ represents the gravity vector in the world coordinate system.

The derivatives of rotation, velocity, and translation are expressed as:

$$\begin{aligned} \dot{R}_{WB} &= R_{WB}\,{}_B\omega_{WB}^\wedge \\ {}_W\dot{v}_{WB} &= {}_W a_{WB} \\ {}_W\dot{p}_{WB} &= {}_W v_{WB} \end{aligned} \tag{31}$$

The rotation, speed and translation in the world coordinate system can be obtained by the general integral formula:

$$\begin{aligned} R_{WB}(t+\Delta t) &= R_{WB}(t) Exp\left(\int_t^{t+\Delta t} {}_B\omega_{WB}(\tau) d\tau\right) \\ {}_W v(t+\Delta t) &= {}_W v(t) + \int_t^{t+\Delta t} {}_W\mathrm{a}(\tau) d\tau \\ {}_W p(t+\Delta t) &= {}_W p(t) + \int_t^{t+\Delta t} {}_W v(\tau) d\tau + \iint_t^{t+\Delta t} {}_W\mathrm{a}(\tau) d\tau^2 \end{aligned} \tag{32}$$

Use Equation (32) in discrete time for Euler integration:

$$\begin{aligned} R_{WB}(t+\Delta t) &= R_{WB}(t)Exp({}_B\omega_{WB}(t)\Delta t) \\ {}_W v(t+\Delta t) &= {}_W v(t) + {}_W \mathrm{a}(t)\Delta t \\ {}_W p(t+\Delta t) &= {}_W p(t) + {}_W v(t)\Delta t + \frac{1}{2}{}_W \mathrm{a}(t)\Delta t^2 \end{aligned} \tag{33}$$

The IMU model is obtained from equations (30) and (33):

$$\begin{aligned} R(t+\Delta t) &= R(t)Exp((\tilde{\omega}(t) - b^g(t) - \eta^{gd}(t)\Delta t) \\ v(t+\Delta t) &= v(t) + g\Delta t + R(t)(\tilde{\mathrm{a}}(t) - b^a(t) - \eta^{ad}(t))\Delta t \\ p(t+\Delta t) &= p(t) + v(t)\Delta t + \frac{1}{2}g\Delta t^2 + \frac{1}{2}R(t)(\tilde{\mathrm{a}}(t) - b^a(t) - \eta^{ad}(t))\Delta t^2 \end{aligned}$$

Suppose there are two image frames with time t_i and t_j, $t_j > t_i$. Therefore, the IMU's pre-integration observation model is:

$$\begin{aligned} \Delta\tilde{R}_{ij} &= R_i^T R_j Exp(\delta\phi_{ij}) \\ \Delta\tilde{v}_{ij} &= R_i^T(v_j - v_i - g\Delta t_{ij}) + \delta v_{ij} \\ \Delta p_{ij} &= R_i^T(p_j - p_i - v_i\Delta t_{ij} - \frac{1}{2}g\Delta t_{ij}^2) + \delta p_{ij} \end{aligned} \tag{34}$$

Among them, A, B, and C are the noise terms of the rotation amount, the pre-integrated speed noise term, and the pre-integrated translation noise term, respectively.

For the pose between two adjacent frames, this paper still uses a nonlinear optimization method to fuse IMU information and visual information. Among them, the state quantities that need to be optimized are:

$$\theta = \left\{R_{WB}^j, {}_W p_B^j, {}_W v_B^j, b_g^j, b_a^j\right\} \tag{35}$$

In equation (36), R_{WB}^j, v_{WB}^j, and p_{WB}^j are the rotation, velocity, and translation of the IMU coordinate system relative to the world coordinate system at time i, and the random walk bias of the gyroscope and accelerometer at time i, respectively.

Therefore, the optimal state quantity θ is solved by optimizing the visual reprojection error and the IMU measurement error:

$$\theta^* = \arg\min_{\theta}(\sum_k E_{proj(k,j)} + E_{IMU}(i,j)) \tag{36}$$

C. Experimental design

The upper computer of the experimental platform in this article is a laptop with Ubuntu 16.04 version, running memory is 8G, processor model is CORE i5 8250U, and the main frequency is 1.6GHz. The robot platform is a Dashgo D1 robot mobile platform that supports the ROS development system. The overall size is $\Phi 406\times 210$ and the diameter of the driving wheel is 125mm. The binocular camera sensor used is MYNT EYE D1000-IR-120/Color.

The experiments in this paper are mainly aimed at the positioning accuracy of the robot. The evaluation index is the RMSE (root-mean-square-error) of the robot position:

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{p}_i - p_i)^2} \tag{37}$$

Where $\hat{p}_i$ is the estimated robot position and p_i is the actual robot position.

Figure 2. Robot Straight Driving Positioning Experiment

In this paper, robot positioning experiments are performed in corridor environments with insignificant environmental characteristics and indoor environments with rich characteristics. In a corridor environment, a mobile robot is used to carry experimental equipment to travel at a constant speed of 10m in the positive direction of the camera, and then the positioning accuracy of pure vision and the positioning accuracy of vision fusion IMU are recorded separately. In a feature-rich indoor environment, a robot linear experiment was also performed to make the mobile robot move forward at a constant speed of 5m in the positive direction of the camera, but the speed was 2.5 times that of the previous experiment. Perform multiple experiments and record the results.

TABLE I. EXPERIMENTAL RESULT

Robot operating environment	Pure visual RMSE/m	Visual fusion IMU RMSE/m
Low-texture corridor environment	0.0746	0.02122
Feature-rich environment	0.1024	0.06502

From the experimental results, it can be seen that the stereo vision positioning error of the fusion IMU is less than the pure vision positioning error, which indicates that the visual positioning of the robot with the fusion IMU is more accurate than the vision-only positioning in low-texture environments and fast robot movements. degree.

IV. CONCLUSION

In this paper, the robot positioning technology in the robot system is researched, and a binocular vision fusion IMU-based robot positioning method is proposed. Compared with the pure vision robot localization method, the proposed method is more robust in low-textured environments and fast robot movements. The experimental results show that the visual positioning method integrated with IMU solves the defects of pure visual positioning to a certain extent and improves the positioning accuracy of the robot.

REFERENCE

[1] Agostino Martinelli. Closed-Form Solution of Visual-Inertial Structure from Motion[J]. International Journal of Computer Vision, 2014, 106(2):138-152.

[2] Smith R C, Cheeseman P. On the representation and estimation of spatial uncertainty[J]. International Journal of Robotics Research, 1986, 5(4): 56-68.

[3] Rublee E, Rabaud V, Konolige K, et al. ORB: An efficient alternative to SIFT or SURF[C]// 2011 International Conference on Computer Vision. IEEE, 2012.

[4] Gao Xiang, Zhang Tao. Fourteen lectures on visual SLAM [M]. Beijing: Publishing House of Electronics Industry, 2017.

[5] V. Lepetit, F. Moreno-Noguer, P. Fua. EPnP:An accurate o(n) solution to the pnp problem[J]. International Journal of Computer Vision,2008,81(2):155-166.

[6] Forster C, Carlone L, Dellaert F, et al. On-Manifold Preintegration for Real Time Visual--Inertial Odometry[J]. IEEE Transactions on Robotics, 2017, 33(1):1-21.

Design and Research of Future Network (IPV9) API

Xu Yinqiu
Shanghai Decimal Network Information Technology Co. Ltd.
E-mail: 8918616209@126.com

Xie Jianping
Shanghai Decimal Network Information Technology Co. Ltd.
E-mail: 13386036170@189.cn

***Abstract*—Socket is a way of process communication, that is used it to invoke some API function to realize the distribution network libraries in different host of data exchange between the relevant process. According to the TCP/IP protocol assigned to the network address of the local host, to communicate between the two processes, the host must know the other's location first, that is, the IP of the other host. At the same time, to get the port number, it is used to identify the local communication process; a local process in communication will occupy a port number, different process port number is different, so it must be assigned a port number that is not used before communication. A complete inter-network process consists of two processes and should use the same high-level protocol. IPV9 is the most important part of future network. This paper introduces the interface function and socket of IPV9, which lays a foundation for further network application programming.**

Keywords-IPV9; Interface; Socket; API

I. INTERFACE AND SOCKET

The transport layer implements end-to-end communication, so there are two terminals for each transport layer connection. What is the terminal of the transport layer connection? It is neither the host, nor the host's IP address, and not the application process, not the transport layer protocol port. The terminal to which the transport layer connects is called a socket. According to the definition of RFC793, the port number is spliced to the IP address to form a socket. A socket is actually a communication endpoint, an interface between an application and a network protocol. Each socket has a socket number, including the IP address of the host and a 16-bit host port number, such as (host IP address: port number).

In short, Socket is equals to (IP address: port number), which is represented by a decimal IP address followed by a port number, separated by a colon or comma. Each transport layer connection is uniquely identified by two terminals (that is, two sockets) at each end of the communication. For example, if the IPv4 address is 118.38.18.1 and the port number is 23, the resulting socket is (118.38.18.1:23), If the IPV9 address is 86[128[5]118.38.18.1 and the port number is 23, the resulting socket is (86[128[5] 18.38.18.1:23).

A socket can be thought of as a terminal in the communication connection between two network applications. During communication, a piece of information to be transmitted by one of the network applications is written into the Socket of its host, which sends the piece of information to the Socket of another host through the transmission medium of the network interface, so that the piece of information can be transmitted to other programs. Therefore, the data transfer between the two applications is done through the socket.

During network application design, IPv4 can be realized through the programming interface of TCP/IP provided by the system, since the core content of TCP/IP is encapsulated in the operating system.

All clients and servers of TCP-based socket programming begin with calling a socket, which

DOI: 10.21307/ijanmc-2019-071

returns a socket descriptor. The client then calls the connect function, while the server calls the bind (), listen (), and accept () functions. The socket is usually closed by using the standard close function, but it can be also used the shutdown function to close the socket. The Socket interaction flow is shown in figure 1.

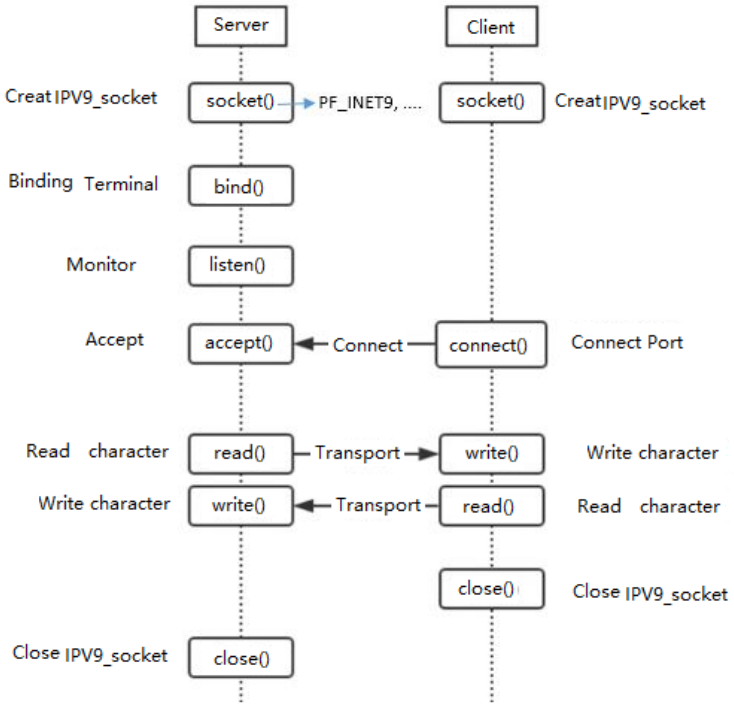


Figure 1. The Socket interaction flow

II. IPV9 SOCKET

In the Linux environment of IPv9, the core contents of TCP 9/IP9 are encapsulated in the operating system kernel. In order to support user development of application-oriented communication programs, most systems provide a set of application programming interfaces (API) based on TCP 9 or UDP 9, which are usually presented in the form of a set of functions, also known as sockets. These sockets are described below. This document is the IPv9 protocol experimental application development instructions, non-industry standard documents.

A. Socket

A socket is an abstraction layer through which applications can send or receive data and open, read, and close it as if it were a file.Sockets allow applications to plug I/O into the network and communicate with other applications in the network. This version of network sockets supports a combination of IPv4, IPv6, and IPv9 addresses and ports.

1) Head file

#include<sys/types.h>

#include<sys/socket.h>

2) Prototype

int socket(int domain, int type, int protocol);

3) Description

Socket: Creates and returns a communication sleeve interface handle.

The parameter domain describes the communication domain, that is, the select communication protocol family. These communication protocol families are defined in the header file

<sys/socket.h>. Currently supported protocol families are as follows:

- PF_UNIX,PF_LOCAL(Local communication protocol)
- PF_INETIPv4 (Protocol)
- PF_INET6(IPv6 Protocol)
- PF_INET9(IPv9 Protocol)
- PF_IPXNovell (Protocol)
- PF_NETLINK(Core user interface device)
- PF_X25ITU-T X.25 (Protocol)
- PF_AX25AX.25 (Protocol)
- PF_ATMPVC(Access to the original ATM PVCs)
- PF_APPLETALKAppletalk
- PF_PACKET(Low-level envelope interface)

The parameter type is used to describe the communication semantics. Currently defined types are as follows:

- SOCK_STREAM

It provides sequential, reliable, duplex, connection-based byte streams that can also support out-of-band data transfer.

- SOCK_DGRAM

It supports datagram (connectionless, unreliable messages of fixed maximum length).

- SOCK_SEQPACKET

It provides a sequential, reliable, duplex, connection-based data path for datagram of fixed maximum length.

- SOCK_RAW

It provides original network protocol access.

- SOCK_RDM

It provides a reliable datagram layer that does not guarantee order.

Some sets of interface types are not implemented on all protocol families, such as the SOCK SEQPACKET is not implemented in the AF_INET protocol family

The parameter protocol describes a special protocol for the socket interface. There is usually only one simple protocol that can support a particular set of interface types that contain a given family of protocols. Of course, sometimes when multiple protocols exist that must be specified with this parameter.

4) Returned value

-1 is returned value when an error occurs, and errno represents the error type value. Otherwise, the socket interface handle value is returned。

B. Bind ()

Bind () is a local address to a set of interfaces function. This function is suitable for unconnected datagram or stream class interfaces and is used before connect () or listen () calls. When a socket () is created, it exists in a namespace (address family) but is not named. The bind () function establishes a local binding (host address/port number) for the socket interface by assigning a local name to an unnamed socket interface.

1) Head file

#include<sys/types.h>

#include<sys/socket.h>

2) Prototype

Bind(intsockfd, structsockaddr *my_addr, socklen_taddrlen);

3) Description

Bind() provides the local address my_addr for the socket interface handle, the length of my_addr is the parameter addrlen, which is called set interface name assignment.

In general, a socket interface of type SOCK_STREAM must call bind() to assign a local address in order to connect and receive.

The structure of the assignment is also different for different protocol families. Such as for AF_INET is sockaddr_in and AF_INET9 is sockaddr_in9.

4) Returned value

Return 0 on success. The error returns -1, and errno represents the error type value.

C. Connect ()

Connect () is used to establish a connection to the specified socket。

1) Head file

#include <sys/types.h>

#include<sys/socket.h>

2) Prototype

Connect (intsockfd, conststructsockaddr *serv_addr, socklen_taddrlen);

3) Description

The handle sockfd must point to a socket interface. If the type of the socket interface is SOCK_DGRAM, the address represented by the parameter serv_addr is the default destination address of the datagram and the source address when the datagram is received. If the socket interface is of type SOCK_STREAM or SOCK_SEQPACKET, the call attempts to establish a connection to another socket interface. The other interface is described by the serv_addr parameter, which is the address of interface communication spaces, each of which interprets the serv_addr parameter.

Typically, connection-based protocols only successfully connect once; connectionless interfaces may connect multiple times to change sessions. A connectionless interface may also connect to an address whose family of protocols is AF_UNSPEC to cancel the session.

4) Returned value

Return 0 on success. The error returns -1, and errno represents the error type value.

D. Listen ()

It is used to create a socket interface and listen for the requested connection.

1) Head file

#include <sys/types.h>

#include<sys/socket.h>

2) Prototype

int listen(int s, int backlog);

3) Description

To confirm the connection, the socket is called to create a socket interface, and the listen () describes the willing to confirm the connection and the length limit of the connection queue before calling accept to confirm the connection. The listen () call only works on the socket interfaces of types SOCK_STREAM and SOCK_SEQPACKET.

The parameter backlog defines the maximum length of the unconnected queue.

4) Returned value

Return 0 on success. The error returns -1, and errno represents the error type value.

E. Accept ()

It is used to create a socket interface and monitoring for the requested connection.

1) Head file

#include <sys/types.h>

#include<sys/socket.h>

2) Prototype

int accept(int s, structsockaddr *addr, socklen_t *addrlen);

3) Description

The accept function can be used based on the socket interface type of the connection (SOCK_STREAM, SOCK_SEQPACKET, and SOCK_RDM). It selects the first connection request in

the unconnected queue, creates a new connected socket interface similar to the parameter s, and then assigns a handle to that socket interface and returns. The newly created socket interface is no longer in the listening state, and the source socket interface s is not affected by the call.

4) Returned value

The error returns -1, and errno represents the error type value. Successfully returns a non-negative integer, representing the handle to the socket interface.

F. Select ()

It is used for monitoring three socket interfaces.

1) Head file

#include <sys/time.h>

#include <sys/types.h>

#include <unistd.h>

2) Prototype

int select(int n, fd_set *readfds, fd_set *writefds, fd_set *exceptfds,

structtimeval*timeout);

FD_CLR(intfd, fd_set *set);

FD_ISSET(intfd, fd_set *set);

FD_SET(intfd, fd_set *set);

FD_ZERO(fd_set *set);

3) Description

Select () allows to monitoringthe three socket interface at the same time: readfds, writefds and exceptfds.

The socket interface in the Readfds will be listened for acceptable characters; the socket interface in writefds will be monitored to see if data can be sent immediately. The socket interface in exceptfds will be monitored for exceptions.

Four macros are defined to manipulate the set of socket interfaces: FD_ZERO will empty a set; FD_SET and FD_CLR add or remove a handle from the set. FD_ISSET is used to test whether a handle is in the set.

The parameter n should be equal to the value of the highest file descriptor plus 1.

The timeout parameter defines the maximum interval for the select call to block until it returns. It can be zero, so that select returns directly. If the timeout structure pointer is NULL, select will block the indeterminate time.

4) Returned value

On success, return the socket interface handle contained in the socket interface set, returns 0 if no change occurs after the maximum interval. -1 is returned on error, and errno represents the error type value

G. Recv (), Recvfrom(), Recvmsg()

1) Head file

#include <sys/types.h>

#include<sys/socket.h>

2) Prototype

intrecv(int s, void *buf, size_tlen, int flags);

Intrecvfrom(int s, void *buf, size_tlen, int flags, structsockaddr *from,

socklen_t *fromlen);

Intrecvmsg(int s, structmsghdr *msg, int flags);

3) Description

The Recvfrom() and recvmsg() calls are used to receive information from a socket interface, regardless of whether the socket interface is connection-oriented. If the from () parameter is not NULL, then the socket interfaces is not connection-oriented and the source address of the message is assigned to it. The fromlen parameter starts with the data buffer size of the parameter from and returns the buffer size of the actual storage address in the parameter from.

A Recv () call is usually used in a connected socket interface, this is equivalent to the case where the parameter from is NULL when recvfrom is called.

If the data message is successfully received, the return value is the length of the data message. If the length of the data message exceeds the length of the data buffer, the excess is discarded, depending on the type of socket interface used to receive the message.

If the socket interface does not receive the information, it will always wait for the information unless the socket interface is non-blocking. When the socket interface is non-blocking, the return value is -1 and the errno value is EAGAIN.

The Recvmsg call the MSGHDR structure, defined in the header file <sys/socket.h>.

4) Returned value

The length of the received data is returned on success, -1 is returned on error, and errno represents the value of the error type.

H. Send(), Sendto(), Sendmsg()

1) Head file

#include <sys/types.h>

#include<sys/socket.h>

2) (2) Prototype

Intsend(int s, const void *msg, size_tlen, int flags);

Intsendto(int s, const void *msg, size_tlen, int flags, conststructsockaddr *to,

socklen_ttolen);

Intsendmsg(int s, conststructmsghdr *msg, int flags);

3) Description

The Send (), sendto, and sendmsg calls are used to transfer information to other interfaces. The Send () call applies only to the connection-oriented socket interface, while the sendto and sendmsg calls apply to all situations.

The destination address is set by the parameter to, its length is the parameter tolen, and the length of the message is represented by the parameter len. If the length of the message is too large to be sent all at once by the low-level protocol, -1 is returned, and errno is set to EMSGSIZE.

If the length of the send message is greater than the length of the socket interface send buffer, the send call will normally block unless the socket interface is set to a non-blocking mode. In non-blocking mode -1 is returned and errno is set to EAGAIN. The Select call can determine whether more data can be sent.

The structure MSGHDR is defined in the header file <sys/socket.h>.

4) Returned value

The length of the sent data is returned on success, -1 is returned on error, and errno represents the value of the error type.

I. Ioctl()

1) Head file

#include <sys/ioctl.h>

2) Prototype

intioctl(int d, intrequest, ...);

3) Description

Ioctl( ) calls operate on the parameters of the underlying device. Parameter d is the file handle, and the parameter request determines the type and size of the back parameters. See the <sys/ioctl.h> for the macro definition used to describe the parameter request.

4) Returned value

0 is returned on success, -1 is returned on error, and errno represents the error type value.

J. Getsockopt()，setsockopt()

1) Head file

#include <sys/types.h>

#include<sys/socket.h>

2) Prototype

Intgetsockopt(int s, intlevel, intoptname, void *optval, socklen_t *optlen);

Intsetsockopt(int s, int level, intoptname, const void *optval, socklen_toptlen);

3) Description

The Getsockopt() and setsockopt() calls can operate on the options of the socket interface. Options exist at multiple protocol levels, but are always represented at the highest socket interface level. When socket interface options are setting, it must be specify the level name and option name. For the socket interface level option, the level is called SOL_SOCKET. For other levels of protocol, other protocol control Numbers are provided, such as the TCP protocol, and the level name must be the TCP series.

The parameters optval and optlen are used when setsockopt calls access option values. For the getsockopt calls, they are buffers that return the request option value; the optlen parameter starts with the size of the buffer optval, and returns with the buffer size of the actual return value. If no option value can be returned, the parameter optval is set to NULL.

The optname and option parameters are sent to the appropriate core protocol module for interpretation without explanation. In the header file <sys/socket.h>, there is a detailed definition of the socket interface level and option structure, and the option formats and names for different protocol levels vary greatly.

Most interface-level options take an integer value as the parameter optval, and for setsockopt calls, the parameter must be non-zero to support Boolean options, or zero to disable.

In the design of IPv9 stream label, the following call can be used:

```
int on = 1；
struct in9_flowlabel_req freq；
structin9_addrdst_addr；
memcpy(&(freq.flr_dst)，&dst_addr，32)；
freq.flr_label = htonl(0x0000000f)；
freq.flr_action = IPV9_FL_A_GET；
freq.flr_share = IPV9_FL_S_EXCL；
freq.flr_flags = IPV9_FL_F_CREATE；
freq.flr_expires = 0；
freq.flr_linger = 0；
freq.__flr_pad = 0；
setsockopt(s, IPPROTO_IPV9, IPV9_FLOWINFO_SEND, &on, sizeof(int))；
setsockopt(s, IPPROTO_IPV9, IPV9_FLOWINFO, &on, sizeof(int))；
setsockopt(s, IPPROTO_IPV9, IPV9_FLOWLABEL_MGR, &freq,
sizeof(structin9_flowlabel_req));
```

The above code sets the stream label of socket s to 0000f, where the destination address of the stream label is defined in dst_addr.

Structure in9 flowlabelreq is defined as follows

```
struct in9_flowlabel_req{
    struct in9_addrflr_dst；
    __u32    flr_label；
    __u8     flr_action；
    __u8     flr_share；
    __u16    flr_flags；
    __u16    flr_expires；
    __u16    flr_linger；
```

```
        __u32    __flr_pad;
    };
```

4) Returned value

0 is returned on success, -1 is returned on error, and errno represents the error type value.

III. IPV9 DEVELOPMENT INSTRUCTION

1) Development environment

Centos7 operating system with Linux operating environment with IPv9 kernel;

VMware virtual machine image:: Centos7_ IPv9_ dev_vm.

The compiled program copy in Centos7_IPv9_dev_vm virtual machine image, it can run normally, provides the virtual machine application development and compilation environment, C language headers file and IPv9_Linux kernel.

2) IPv9 network application development directory:

/develop9

Development document directory:

/develop9/docs

The demo directory:

/develop9/test9

demo README

/develop9/test9/README

3) Test9 program

The test9 program mainly changes the socket family program file.

cd /develop9/test9

make

4) Demo operation

#Configure the IPv9 address

ifconfig9 eth1 add 32768[86[21[4]10001

#Start the IPv9 server program: /test9_tcpserver

#Start the IPv9 client program: /test9_tcpcli 32768[86[21[4]10001

Verify the caught: tcpdump -s 0 -i any -w t.cap, or wireshark with ipv9 plugin open t.cap.

IV. CONCLUSION

This paper introduces the commonly used socket able and interface functions, including creating a socket, binding function, link function, monitoring function and accept function, read the function and writing function, etc., each function is connected the header files, prototyping, description, and the return value, these are the basis of network programming, mastering these functions, which plays a major role for application development.

REFERENCES

[1] https://zh.wikipedia.org/wiki/Berkeley%E5%A5%97%E6%8E%A5%E5%AD%97, Wikipedia: Berkeley sockets 2011-02-18, (Goodheart 1994, p. 11), (Goodheart 1994, p. 17)

[2] Cisco Networking Academy Program, CCNA 1 and 2 Companion Guide Revised

[3] Third Edition, P.480, ISBN 1-58713-150-1

[4] Jack Wallen (2019-01-22). "An Introduction to the ss Command".

[5] V. S. Bagad, I. A. Dhotre (2008), Computer Networks (5th revised edition, 2010 ed.), Technical Publications Pune, p. 52

[6] Ian Griffiths for IanG on Tap. 12 August, 2004. Raw Sockets Gone in XP

[7] "raw(7): IPv4 raw sockets - Linux man page". die.net.

[8] "Raw IP Networking FAQ". faqs.org.

[9] www-306.ibm.com - AnyNet Guide to Sockets over SNA

[10] books.google.com - UNIX Network Programming: The sockets networking API

[11] books.google.com - Designing BSD Rootkits: An Introduction to Kernel Hacking

[12] historyofcomputercommunications.info - Book: 9.8 TCP/IP and XNS 1981 - 1983

[13] mit.edu - The Desktop Computer as a Network Participant.pdf 1985

A Study on the Sinter Brazing Joint of Powder Metal Components

Pujan Tripathi
Department of Engineering Technology
Missouri Western State University
4525 Downs Drive
St. Joseph, MO 64507, USA

George Yang
Department of Engineering Technology
Missouri Western State University
4525 Downs Drive
St. Joseph, MO 64507, USA
E-mail: yang@missouriwestern.edu

***Abstract*—The research focuses on the development of a new joint method, the sinter brazing of powder metal components. Various kinds of powder metallurgy composition were tested in the sinter brazing joint, mainly to study the operating conditions, strength of the joint. Although each of these processes brings about different results in metal, all of them involve three basic steps: heating, soaking, and cooling. Heat treatment such as heating, soaking, cooling, hardening, tampering has been investigated.**

Keywords-Powder Metallurgy; Sintering; Brazing; Metal Joining

I. INTRODUCTION

Powder metallurgy is a processing method where green parts are compacted using dies and get sintered. Sintering offers equivalent strength as a cast iron and superior design flexibility and produces Near-Net-Shaped (NNS) parts at lower costs and it reduces the need for the machining process. Now, sinter brazing is an established joining process for Powder Metal components, and it is often used in the production of automotive applications. A successfully brazed joint relies widely on the interaction between brazing alloy, the joining surfaces, and sintering atmosphere conditions. The purpose of testing brazing materials is to test the capabilities of the brazing filler materials in order to produce sinter brazed components, processed through sinter brazing to heat treatment with air quench and to compare different brazing pastes to determine their quality and strength to the powder metal components.

Product designs within the powder metal industries utilize joining techniques to assemble a component from different compacted pieces. This enables PM industries to provide cost-effective complex parts for various applications compared to traditional fabrication practices. Sinter-brazing is one of the methods for joining the parts easily and efficiently. Brazing mechanism can be complex; however, it has the potential to reduce additional processing steps in a manufacturing scenario. The process can be achieved within one step instead of the traditional two steps. It also has economic advantages [1]. The metallic bond formed between parent mental surfaces has an adequate strength to achieve the high-performance standards according to the requirements.

Interconnected porosity generates significant capillary forces that rapidly pull braze away from the connection interface. Moreover, the pore network becomes like a conduit that fills the bulk of the part with filler materials which results in a joint deficiency. Therefore, porosity is a considerable challenge in the sinter-brazing process. Using Copper (Cu) to infiltrate before brazing can be a potential solution to fill the pore network but is an expensive process. Another option is compressing to densities > 7. g/cm3 [4] but has a risk to get reduced strength than expected in the connection interface.

Burgess Norton Mfg. Co. (Geneva Illinois) is making powder metal products since 1903 and operating seven facilities all over the world. 1000101-MPIF FLC:4405 is a widely used filler alloy for infiltration during the sinter-brazing process by Burgess Norton. This alloy limits the loss of material in the pore network and creates a strong bond between the parent metal surfaces. This can be achieved by increasing the liquid temperature to increase the surface tension, thus reducing the flow ability to prevent braze alloy to penetrate further. However, the process is not entirely smooth and influenced by many external parameters. Despite having a high success ratio, it has occasional failure due to excessive infiltration. Significant work has been done to understand this challenge within manufacturing. One of the methods to counter excessive infiltration is to

add some iron powder to the braze pre-mix to influence the onset of solidification temperature [5,6].

Brazing has many established best practices [1,6,7]. Performing each step of the sintering cycle within the controlled environment is one critical factor. To have an effective relubrication without shooting, a little oxidizing is required within the preheat zone. Although, some surface contaminants can negatively influence wetting and hinder braze from flowing through a gap. The excessive oxidizing atmosphere can cause inefficiency in reducing the brazing material; a greenish tint on the surface indicates excessive oxidation[1]. Also, furnace temperature should be adjusted for different loads. Furthermore, maintaining consistent temperature is essential to regulate desired brazing metal flow. Slow heating rates can segregate braze from ler melting constituents and alloy with an iron after re-solidification, impeding the remaining brazing metal from flowing into the pore network [2].

For successful brazing, wetting parent material surfacing with liquid braze is one of the key factors. It is important to dissolve the surface oxides by utilizing a flux on brazing and parental material to prevent braze to flow into the pore network within the joint [8]. However, in that process, glassy residues of metal oxides are left behind which can be adherent in nature. Thus, it is recommended to incorporate blind holes or enclosed cavities within the design. Regardless, if not performed in a vacuum chamber, adding flux is essential in the sinter-brazing process.

Traditionally, sinter-brazing was performed with 1000101-MPIF FLC:4405and sintering in a furnace cycle process. But to challenge the capabilities of sinter brazing materials and its components, another test took a place and the research done with the powder 1000111-MPIF FLC-4808 which showed the significant results in the sinter brazing procedure.

II. MATERIALS AND METHODS

The braze alloys that are used in the study are, Brazing powder 1000101-MPIF FLC:4405 and 1000111-MPIF FLC-4808. These are the two most used powder of Burgess Norton Mfg. Co. (Geneva Illinois). The sinter brazing paste is used SCM-Sinter Braze Grade:C-458 and SCM-Sinter Braze Grade: EXP6778-99.

TABLE I. DESCRIPTION OF TWO TESTING METAL POWDERS

TYPE	GRADE	Description
Pre-alloyedSteel	FL-4405	Low alloy steel with pre alloyed manganese, molybdenum and nickel content for better hardenability.
Hybrid-alloySteel	FLN2-4405, FLN4-4400, FLN4-4405, FLN6-4405, FLNC-4405,	Low alloy steel with pre alloyed molybdenum and admixed nickel and copper for better compressibility.

For the procedure weight of the powder used 18.5 to 20.00 grams. Selected TSI is 33 to 35. And the pressure used to test the shear strength is 39,197lbs to 43, 761lbs.for the testing density needed was 6.7 to 6.9 g/cm^3. Brazing slugs weighted length of 1.2mm, width 0.5mm and approximate Height of 0.2mm. Brazing material used a sinter brazing paste and in which the weight of the paste used around 0.15 to 0.20grams.

TABLE II. MOLYBDENUM- NICKEL-STEEL- PRE ALLOYED- AS SINTERED

	Chemistry					Typical properties					
GRADE	C	Mn	Mo	Ni	Cu	Density g/cm^3	Tensile Strength ksi (MPa)	Yield Strength ksi (MPa)	Elongation %	Unnotched Impact Energy ft-lb (J)	Apparent Hardness HRB
FL-4405	0.4-0.7	0.05-0.30	0.75-0.95	-	-	6.70 7.10	52 (360) 66 (460)	42 (290) 52 (360)	1	6 (8) 16 (22)	60 73

TABLE III. MOLYBDENUM-NICKEL STEEL – PRE-ALLOYED –HEAT-TREATED

	Chemistry						Typical properties				
GRADE	C	Mn	Mo	Ni	Cu	Density g/cm³	Tensile Strength ksi (MPa)	Yield Strength ksi (MPa)	Elongation %	Unnotched Impact Energy ft-lb (J)	Apparent Hardness HRC
FL-4405-HT	0.4-0.7	0.05-0.30	0.75-0.95	-	-	6.70 7.10	110 (760) 160 (1100)	~UTS	<1	5.5 (7) 9 (12)	24 34

Reference: https://www.ssisintered.com/materials/low-alloy-molybdenum-nickel-steels.[10]

III. METHOD

Make slugs of powder Brazing powder 1000101-MPIF FLC:4405 and 1000111-MPIF FLC-4808, the basic requirement for making slugs is to get a density of 6.7 to 6.9g/cm^3. Once the density and slugs are made use sinter brazing paste on both the different products. Once the brazing material is applied the brazed parts are ready for sintering. Sintering is effective when the process reduces the porosity and enhances properties such as strength, electrical conductivity, transparency and thermal conductivity; yet, in other cases, it may be useful to increase its strength but keep its gas absorbency constant as in filters or catalysts. During the firing process, atomic diffusion drives powder surface elimination in different stages, starting from the formation of necks between powders to the final elimination of small pores at the end of the process. In a furnace, test brazed parts are heated for 16 hours cycle and at 1600c temperature. Once the sintering is done slugs were being tested for the shear strength. Once the product is sintered then there are two possibilities for the heat treatment, there might be a chance that the product might get melt, or its bond gets stronger. But there are high chances that the product gets melted. After the sintering the heat treatment took place, Heat treatment is any one of several controlled heating and cooling operations used to bring about the desired change in the physical properties of a metal. Its purpose is to improve the structural and physical properties for some particular use or for future work of the metal. There are five basic heat-treating processes: hardening, case hardening, annealing, normalizing, and tempering. Although each of these processes brings about different results in metal, all of them involve three basic steps: heating, soaking, and cooling. Heat treatment steps are as followed, heating, soaking, cooling, hardening, tampering. Following the heat treatment process was Air quenching carburizing process. Specification of this process, A Cycle – Air Cooled, Endo Gas Carburizing, Deep Freeze Tempering – 150 ℃ – 230 ℃

1st Zone - 1600
2nd Zone – 1600
3rd Zone – 1600
4th Zone – 1550
5th Zone – 1550

Figure 1. Temperature Range of Sintering

The following figure illustrates the method to use the Universal Tester machine to separate the two brazed components, to calculate shear strength.

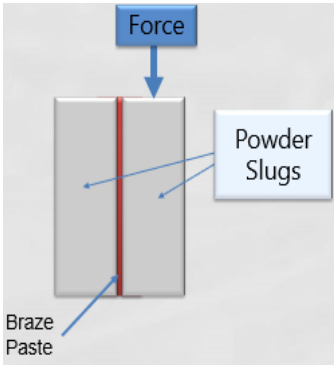


Figure 2. Tensile Testing Illustration

This device is measuring the rupture strength. Shear strength is calculated by the following procedure:

- Drill a 0.25" (6.4mm) diameter hole by 0.125" (3.2mm) deep through the center of braze
- Place a 0.25" Dowel pin in the hole and push down with tensile tester until part splits
- Record reading = Breaking Load

IV. RESULTS AND CONCLUSION

The research took place at Burgess Norton Mfg. [9]. Co. The test showed that the brazing paste SCM-Sinter

Braze Grade: EXP6778-99 showed a significant increase in strength in comparison to the SCM-Sinter Braze Grade:C-458 paste. The EXP6778-99 paste was also not affected by the Heat Treat operation (typically weakens). This was a significant finding because no other sinter brazing process are currently Heat-Treating parts after they have been brazed due to the weakening of the bond.

TABLE IV. SHEAR STRENGTH OF DIFFERENT JOINTS

Shear Strength (psi) Summary		
Powder	**714080-99 RoHS Compliant**	**EXP6778-99**
1000101	2643.4	5018.3
1000101HT	2405.6	6230.5
1000111	3998.5	7118.6

Reference: Burgess Norton Mfg. Co. [9]

REFERENCES

[1] Gabrielov, C. Wilson, and J. Hamill, "Sinter-Brazing Automotive Powertrain Components," International Journal of Powder Metallurgy, Vol. 37, 7, 2001, pp 46-48.

[2] M. Frank, "Advanced Sinter brazed P/M Subassemblies in Torque Transfer Systems: Reduction Hubs and Carriers," SAE, 2002.

[3] M. Onoda, R. Kameda, and T. Koiso, "Application of Sinter-Brazing," Metal Powder Report, November 1983, pp 632 – 634.

[4] P. Beiss, "Finishing processes in powder metallurgy," Powder Metallurgy, 1989, 32, 4, pp 277-284.

[5] M. Stromgren and O. Andersson, "Brazing of P/M Parts by Sintering," Technical Report - PM 89-6.

[6] W. Knopp, Brazing Alloy Composition, United States Patent 3,717,442 Feb. 20, 1973.

[7] R. Peaslee, "How to Identify Brazing Failures," Advanced Materials & Processes, December 2003, p 35.

[8] AWS Committee, Brazing Manual, 1963, American Welding Society, New York, NY.

[9] Burgess Norton Mfg. Co. https://www.ssisintered.com/materials/low-alloy-molybdenum-nickel-steels.

Application of the Source Encryption Algorithm Model in the Power Industry

Jiao Longbing
Shaanxi Huabiao Network Technology Co., Ltd.
Xi'an, 610041, China
E-mail: james9894@163.com

***Abstract*—With the development of Internet technology and Internet of Things technology, the Internet of Everything has become a hot topic, and in March 2019, the National Grid for the first time clarified the definition of the Pan-In Power Internet of Things, pointing out that the company's most urgent and important task is to accelerate the construction of the Pan-In Power Internet of Things. The security of Data Transfer on-line at any time is particularly important, in order to ensure the security of data, in the process of data transmission, data needs to be encrypted. This paper expounds a model of the information source data encryption algorithm, analyzes the encryption algorithm and the encryption method, and then provides a reference basis for the data transmission data security of power system.**

Keywords-Data Communication; Encryption Algorithm Model; Encryption Technology

I. INTRODUCTION

Driven by intelligence and informatization, ubiquitous electric Internet of things is just at the right time. The construction of power Internet of things puts forward higher requirements for data management and information management. At present, the state grid system is connected to more than 500 million terminal devices, and with the construction of electric Internet of things and the surge, there will be a huge amount of data. Data is an important asset, data privacy protection, the construction of data security grading system, based on different security levels to determine the open rights of data, to ensure the efficiency of business execution and smooth management.

Internet of things technology is developing rapidly, but the corresponding infrastructure and security protection capabilities do not adapt to it. Network security is the biggest hidden danger of the power Internet of things. On the one hand, non-ip communication protocol is often used to transmit data in the Internet of things, which lacks effective security measures. On the other hand, the increasingly intelligent and professional means of network attack have brought new problems to network security protection, leading to frequent network security incidents in the field of power grid in recent years.

Therefore, strengthening the security risk control and management of intelligent and informationized power Internet of things will be a key point of China's ubiquitous power Internet of things construction.

This paper will focus on the introduction of a source encryption algorithm model, hoping to provide a digital security model reference for the application of digital business in the power industry.

II. INTRODUCTION TO ENCRYPTION TECHNOLOGY

As an important part of network security, data encryption technology plays a very important role in the network. It involves the confidentiality, authentication, non-repudiation and integrity of data. Key is the key of data encryption, which controls the

DOI: 10.21307/ijanmc-2019-073

implementation of encryption and decryption algorithms. According to the different keys, the encryption technology is divided into symmetric encryption technology, asymmetric encryption technology, mixed encryption technology.

A. *Symmetric encryption*

Symmetric encryption means that the encryption key can be inferred from the decryption key, and the decryption key can also be inferred from the encryption key. In most symmetric algorithms, the encryption key and the decryption key are the same. For this algorithm, its key (secret key) usually needs messenger or secret channel to transmit, and it is difficult to transmit and manage the key. In this case, the secret preservation of the key determines the security of the algorithm. RC4, chaos algorithm, DES, IDEA, RCZ algorithm are typical representative of symmetric key encryption system. Because both parties have the same key, symmetric encryption technology is easy to implement and fast, so it is widely used in communication and storage data encryption and decryption. The security of symmetric encryption depends on the key, so the secret of the key is very important to the security of communication. The symmetric encryption process is shown in the figure 1.

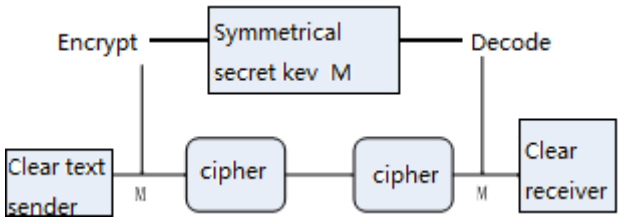


Figure 1. Symmetric encryption flow chart

B. *Asymmetric encryption*

This technique can also be called public key cryptography. The encryption key (public key) can be made public, that is, it can be obtained by strangers and used to encrypt the information, but the information can only be decrypted with the corresponding decryption key (private key). Compared with symmetric encryption algorithms, asymmetric encryption algorithms usually require two keys: public key and Private key. When data is encrypted with a key, if it is encrypted with a public key, it can only be decrypted with the corresponding private key. Instead, it is decrypted with the corresponding public key. The advantage of public key cryptography is that it can adapt to the open requirements of the network, but the speed is relatively slow, not suitable for encrypting files. The asymmetric encryption process is shown in the figure 2.

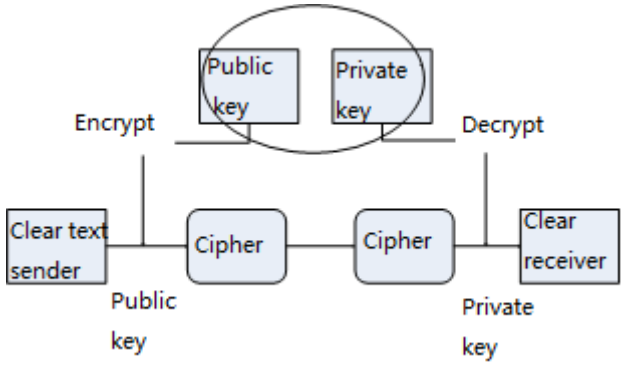


Figure 2. Asymmetric encryption flowchart

C. *Hybrid encryption*

Hybrid encryption is not a single encryption technology, but a combination of the above two data encryption technology combined product. The communication process of the communication parties is divided into two parts. The parties first use asymmetric encryption technology to transmit the symmetric key used in the communication, and then use symmetric encryption technology to encrypt and transmit the file.

III. TEMPORAL AND SPATIAL VARIABLES PARTICIPATE IN THE SOURCE ENCRYPTION MODEL

A. *Time variable definition:*

1) Year, month, day, hour and time variables

TABLE I. TIME VARIOMETER

		1	2	3	4	5	6	7	8	9	10	11	12
		A	B	C	D	E	F	G	H	I	J	K	L
1	a	Aa	Ba	Ca	Da	Ea	Fa	Ga	Ha	Ia	Ja	ka	La
2	b	Ab	Bb	Cb	Db	Eb	Fb	Gb	Hb	Ib	Jb	kb	Lb
3	c	Ac	Bc	Cc	Dc	Ec	Fc	Gc	Hc	Ic	Jc	kc	Lc
4	d	Ad	Bd	Cd	Dd	Ed	Fd	Gd	Hd	Id	Jd	kd	Ld
5	e	Ae	Be	Ce	De	Ee	Fe	Ge	He	Ie	Je	ke	Le
6	f	Af	Bf	Cf	Df	Ef	Ff	Gf	Hf	If	Jf	kf	Lf
7	g	Ag	Bg	Cg	Dg	Eg	Fg	Gg	Hg	Ig	Jg	kg	Lg
8	h	Ah	Bh	Ch	Dh	Eh	Fh	Gh	Hh	Ih	Jh	kh	Lh
9	i	Ai	Bi	Ci	Di	Ei	Fi	Gi	Hi	Ii	Ji	ki	Li
10	j	Aj	Bj	Cj	Dj	Ej	Fj	Gj	Hj	Ij	Jj	kj	Lj

Year, month, day and hour variables are added to the partial key format of data information source, as shown in figure format, 12 bits and 10 bits are combined to form a unique time variable. Year, month, day, and hour cycle with 60 and map to Gregorian calendar time.

2) Increase position variable

The position variable is assigned to map the location of the source to the latitude and longitude coordinates. As the key bit.

TABLE II. LOCATION VARIABLES

4 southeast	9 south	2 southwest
3 east	5 center	7 west
8 northeast	1 north	6 northwest

3) The source data is rearranged and encrypted according to location, time and solar term variables

The data format conversion of an encryption depends on a unique time point, and the time point of data encryption determines that the data conversion mode in the figure below is unique.

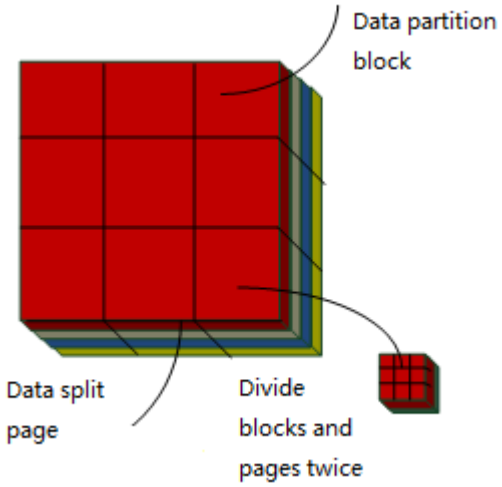


Figure 3. Data conversion mode

Determine the starting time of the data encryption variable, time variable form the basis for packet rotation changes, packets after be confirmed time and location, the data layer, and then each layer of blocks, each block in the above order to secondary polarization, and then according to the law of time each block and the PAGE will be rotating, and the rule of each PAGE rotation, eventually forming the smallest unit, each unit to block position number.

After the above sequence of data format and order is completely disrupted, according to the unconventional format of the order, is to refer to the passage of time and position rotation.

After such data arrangement, data, pictures, videos and other files, even if they are intercepted, cannot obtain the key encryption variables of time, solar term and location, and cannot obtain useful key information even if they are enumerated.

4) Law of rotation arrangement of source data format

After the data is divided into different pages, it is rotated and changed according to the hierarchy of layers. For example, the bottom layer is the rotation mode, the third layer is the feig reordering, the second layer is the rotation according to the third layer of feig, the first layer is the rotation according to the second

layer of rotation. To this data page is partitioned and rearranged. Then the quadratic element is carried out, and the data block of the first part is redistributed into 16 blocks. After three dimensions, three dimensions and four dimensions are differentiated to the smallest data unit bit.

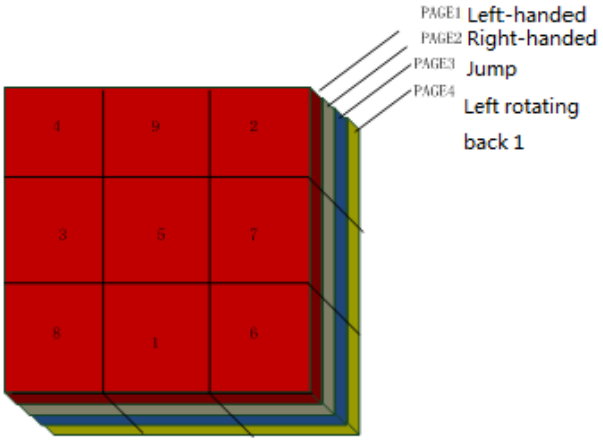


Figure 4. Rotation arrangement rule

5) Data decryption process model

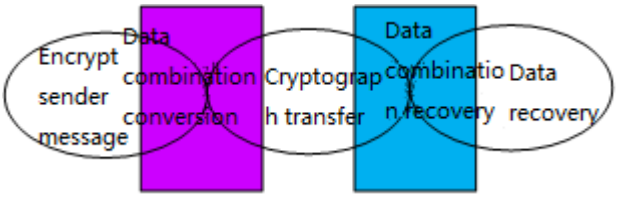


Figure 5. Data decryption model

a) The sender rearranges the blocks and pages of the data format according to the time variable and the position marker bit, and the quadratic element and cubic element are segmented and arranged according to the same method to form the minimum bit of scatter random data.

b) Transfer process: the file transfer process is transmitted with three-point random minimum bit data, and the time at this time is a shift. After the time shift, the information received by the receiver should be arranged according to the change of time variable for time key mapping.

c) C. Data receiver, the receiver is minimum scatter data, the receiver must key is according to the sender to the time, location, variables and receiving time to time, location, and, according to the minimum level to the smallest unit according to position the label to the reverse calculation into pieces, and then pushing block reverse page to the model, and then push the page to the original data model. The process record of backward calculation is the decryption key.

The above data encryption model is the data encryption model established based on the cubic dimension secret calculation model, which can be used to split and switch the top secret data multiple times. In addition, it is not only applicable to data, but also applicable to audio and image data.

6) The programming language realizes the model construction idea of the encryption model

- Data type definition

```
using System;
using System.Collections.Generic;
using System.Linq;
using System.Text;
namespace FateJudge.Core.QM
{
    public class   QM Data
    {
        public static int[] ShunZhuan = new int[] { 1, 8, 3, 4, 9, 2, 7, 6 };
        public static int[] NiZhuan = new int[] { 1,6,7,2,9,4,3,8};
        public static String[] jushu=new string[]{
            "174","285","396",
            "852","963","174",
            "396","417","528",
            "417","528","639",
```

```
            "936","825","714",
            "258","147","936",
            "714","693","582",
            "693","582","471"};
        public static String QYYW="
    public static String[] RPQM = new string[] { " };
        public static String[] tianpanxingyuanwei = new string[] { };
        public static String[] SPQM = new string[] {  };
        public static String[] renpanxingzhuanxu = new string[] {}; public static String[] tianpanxingzhuanxu = new string[] { }; public static String[] shenpanzhuanxu = new string[] { };/
    }
}
```

- Data rotation

```
public int getRPG (string yinyan, int zhishigong, string zhishi, string ren)
        {
            int ret = 0;
            int n = getStringArrayIndex(QMData.rpxzx, ren);// - getStringArrayIndex(QMData. rpxzx, zS) + 8) % 8;
            int n1 = 0;
            n1 = (getIntArrayIndex(QMData.ShunZhuan, zhishigong) + n) % 8;
            ret = QMData.ShunZhuan[n1];
            return ret;
        }
        public int getIntArrayIndex(int[] sa, int n)
        {
            int ret = 0;
            for (int i = 0; i < sa.Length; i++)
                if (sa[i].Equals(n))
                {
                    ret = i;
                    break;
                }
            }
            return ret;
        }
    public string getPostRenPan(string zhishi, int zhishigong, int p)
            {
                string ret = string.Empty;
                int n1 = 0;
                int n2 = 0;
                int n3 = 0;
                n1 = (getIntArrayIndex(QMData.ShunZhuan, p) - getIntArrayIndex(QMD.ShunZhuan, zhishigong) + 8) % 8;
                n2 = getStringArrayIndex(Qi=M=Data. rpxzx, zhishi);
                n3 = (8 + n2 + n1) % 8;
                ret = QiMenData. rpxzx [n3];
                return ret;
            }
```

IV. CONCLUSION

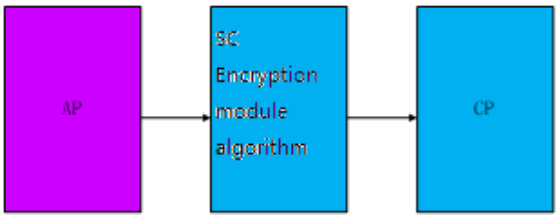


Figure 6. Data encryption model

Power data is the lifeblood of a country's economy and people's life. With the great development of Internet of things, Internet and ubiquitous electric Internet of things, the security of data connection is particularly important. Through the practice of the above data encryption model, the source information can be encrypted, which is different from several symmetric encryption algorithms and asymmetric encryption algorithms. This kind of model calculation can be encapsulated into IC or T card to form an encryption module, which is widely used in power inspection terminals, smart electricity meters and power big data cleaning applications, to increase the security means of data transmission, and to provide some enlightenment and reference for the data transmission encryption methods of power companies.

REFERENCE

[1] Song Lei, Luo Qiliang, Luo Yi, Tu Guangyu. Encryption scheme of real-time data communication in power system [J]. Power system automation, 2004 (07)

[2] Liu gang, liang ye et al. Realization and application of digital certificate technology in power secondary system [J]. Power grid technology,2006,10

[3] Xingyuan Wang,Hongyu Zhao,Le Feng,Xiaolin Ye,Hao Zhang. High-sensitivity image encryption algorithm with random diffusion based on dynamic-coupled map lattices[J]. Optics and Lasers in Engineering,2019,122.

[4] Chen changqing. Research on computer network communication intrusion prevention method based on data encryption technology [J]. Information and computer (theoretical edition),2019(14):171-172.

[5] DOE/Oak Ridge National Laboratory; ORNL to take on nine power grid modernization projects as part of DOE award[J]. NewsRx Health & Science,2019.

[6] Chen zhiguang. Analysis and design of power project management system of state grid fuzhou power supply company [D]. Yunnan university,2016.

On the RFID Information Query Technology Based on IPV9

Li Guiping
The School of Computer Science and Engineering
Xi'an Technological University, Xi'an, China
E-mail: 15693685@qq.com

Xue Lei
Shandong University of Science and Technology
223 Daizong Street, Tai'an City
Shandong Province, 271000
E-mail: Shirleyxue06@163.com

***Abstract*—Since coding format of RF label is inconsistent with the protocol format of the information server's network, a design scheme of network architecture is proposed to achieve the connectivity between the decimal network based on IPV9 and the Internet network based on IPV4/IPV6. And also, two ways of information's Query and connectivity based on D-ONS and direct routing between the decimal network is devised by using the expert modules. The experiment results shows that both the two ways can provide efficient service of RFID Information Query .**

Keywords-IPV9 ; RFID; Domain Transformation; Information Query

I. INTRODUCTION

With the development of radio frequency technology, product-related information can be quickly obtained by the readers if RF tags are assigned to each product. Recently, people have began to adopt IPV6 and even IPV9 protocol formats for RF tag coding since IPV4 address space has been used up. If the server storing product-related information is stored in an IPV4 or IPV6 network, how to query RF information across the network should be solved. This paper mainly studies the RF information query process of interconnection between decimal network and Internet network and the domain name conversion rules obtained by decimal network query server.

II. IPV9

IPV9, short for decimal network and the new generation of Internet, is the result of China's independent innovation of core technologies which has IPV9 protocol, IPV9 addressing, digital domain name system and other core technologies are adopted with original and independent intellectual property rights. Full digital text is used to represent the IP address. The address space is larger than that of IPV4 or IPV6. The 1st-41th level of the address space is expressed as binary 256 bits, while using the decimal 256 bits to express the 42th.

III. RFID

Radio Frequency Identification(RFID) is a non-contact automatic recognition technology. It communicates with other object using reflected power. It can automatically identify target objects and obtain relevant data through radio frequency signals, which can quickly track items and exchange data. And also, the identify ability need no one to participate in. A typical RFID system consists of electronic tag, reader (including antenna) and application system.

Further, electronic tags are the data carrier of RFID system which consists of a label antenna and a label chip. It can receive the reader's electromagnetic field modulation signal and return the response signal to achieve the label identification code and memory data read or write operations. The reader is used to receive host commands and send the data stored in the sensor back to the host by the wired or wireless way. It contains a controller and an antenna. If the reading distance is long, the antenna will stand alone. The terminal computer of the application system interacting with the RFID system transmits the work instructions issued by the MIS application system. It makes electronic tags and readers be coordination through middleware, processes all data collected by RFID system, and carries out calculation, storage and data transmission. The process can be described as Fig.1and Fig.2.

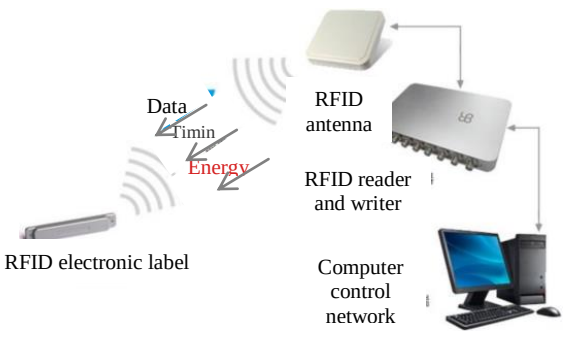


Figure 1. The requery process of information based on RFID

DOI: 10.21307/ijanmc-2019-074

Figure 2. Electronic tags

The operating principle of RFID system is to receive the radio frequency signal emitted by the reader when an item with an electronic tag enters the radiation range of the reader antenna. The passive tag sends data stored in the tag chip using the energy generated by the induced current, while active electronic tags can send data stored in the tag chip, actively. Generally, readers are equipped with middleware of certain functions. Hence, it can read data, decode and directly carry out simple data processing, then send to the application system. The application system judges the legitimacy of electronic tags according to the logic operation, and carries out corresponding processing and control for different Settings, thus realizing the basic functions of RFID system.

IV. NETWORK ARCHITECTURE BASED ON IPV9 RFID INFORMATION QUERY TECHNOLOGY

RFID information query technology can provide people with the function of inquiring information related to products through the RFID tag. The information related to RFID tags is stored on the information server and is generally maintained by the manufacturer of the product. In view of the actual use of the Internet, it is necessary to design a network architecture for interconnection between the decimal network and the Internet, which meets certain conditions. On this basis, RFID tag information query service is implemented. The overall network architecture design scheme is as follows.

A. Overall design of network architecture

The architecture of RFID tag information query service on Internet is based on IPV4 and IPV6 protocols. Routing adopts IPV4 and IPV6 protocols, and resource positioning is completed by SNS and PSNS. The architecture of electronic tag information location query service based on decimal network is a network architecture based on IPV9 protocol, including the following two ways.

(1) Using the routing to locate the information server directly. The route uses the IPV9 protocol without the DNS resolver.

(2) Adopting the parsing service of the application layer with U-code Resolution Server. DONS uses host domain name resolution to provide IPV4, IPV6 and IPV9 addresses, while using IPV4, IPV6 and IPV9 protocols as routing protocols .Resource positioning is done by D-ONS.

The network architecture of RFID tag information query service mainly includes the decimal network information query service technology system and the Internet information query service technology system. Specifically, the decimal network architecture includes middleware, information server, D-ONS server, DDNS and IPV9 direct router. Internet architecture includes SNS server, PSNS server, information server, .cn root DNS server, DDNS server and.cn root. DNS server through the domain name resolution forwarding digital domain name and English domain name connectivity.

B. The key module of network architecture -- expert module

The expert module is the middleware used between the decimal network and the Internet to realize the interconnection between the two, and the data exchange format between the two is XML. It includes the following interfaces.

- RF information query of Decimal network and architecture of discovery technology input product and service digital identifiers to query the Internet RF information, through the expert module, discovery technology architecture request to return the product and service specific information storage address or URI.
- RF information query of Internet and architecture of discovery technology input product and service digital identifiers to query the Decimal network RF information, through the expert module, discovery technology architecture request to return the product and service specific information storage address or URI.
- RF information query of Decimal network and architecture of discovery technology input product and service digital identifiers to query the Internet RF information, through the expert module, discovery technology architecture request to return the product and service specific information.

- RF information query of Internet and architecture of discovery technology input product and service digital identifiers to query the Decimal network RF information, through the expert module, discovery technology architecture request to return the product and service specific information.

V. RESEARCH ON INFORMATION QUERY PROCESS

Based on the above network architecture, the implementation of RFID tag information query service based on IPV9 can be designed in two ways: D-ONS based decimal network and Internet network information query and direct routing mode of exchange visits between decimal network information query service system and Internet information query service system.

A. Exchange query process Between D-ons-based decimal network and Internet network information query service system

When the related product information is stored in the Internet information server and label coding format is using able format, the decimal network based on D-ONS accessing the information query process of Internet mainly involves the following several key modules: decimal network query server, decimal network, expert module, information server middleware, the SNS and PSNS server. The access relationship between these modules is shown in figure 3.

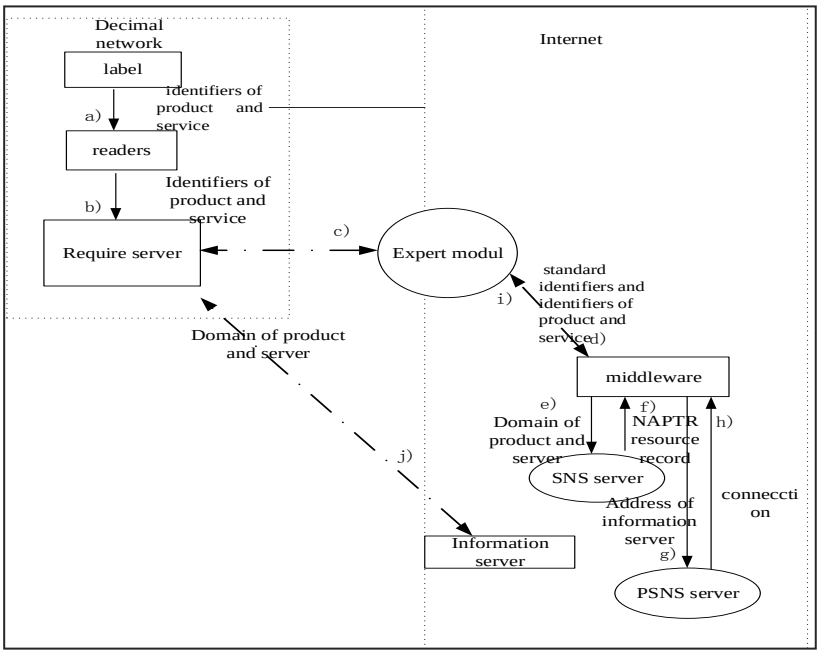


Figure 3. The process of accessing the Internet on a D-ONS- based decimal network.

1) The access process can be described as follows

a) Using RFID readers to read IPV9 identifiers and product and service identifiers from electronic tags.

b) Submitting the read identifiers and product and service identifiers to the query server in the decimal network.

c) Calling decimal network query server and Internet interface expert modules to access the Internet.

d) The Internet interface of the expert module accesses the middleware of the Internet architecture through standard identifiers and identifiers of product and service .

e) The service middleware converts the standard identifier to the domain name format and sends it to the SNS server to request the resolution service.

f) The SNS server returns the conversion rules of PSID domain name by the form of regular expressions to the service middleware.

g) Service middleware issues query request to PSNS server based on PSID domain name.

h) PSNS returns the NAPTR record containing the product and service information service or PSDS address.

i) The service middleware returns the results to the expert module, whose decimal network interface returns service of the product and information or NAPTR records of PSDS address to the query server.

j) Querying the server request and finally getting the product information returned by the information server.

When product-related information is stored in the information server of the decimal network, and the label encoding format is IPV4 or IPV6 format, the D-ONS based decimal network needs to be accessed through the Internet network. The access process involves the key modules: Internet query server, expert module, D-ons server, information server. The access process is shown in figure 4.

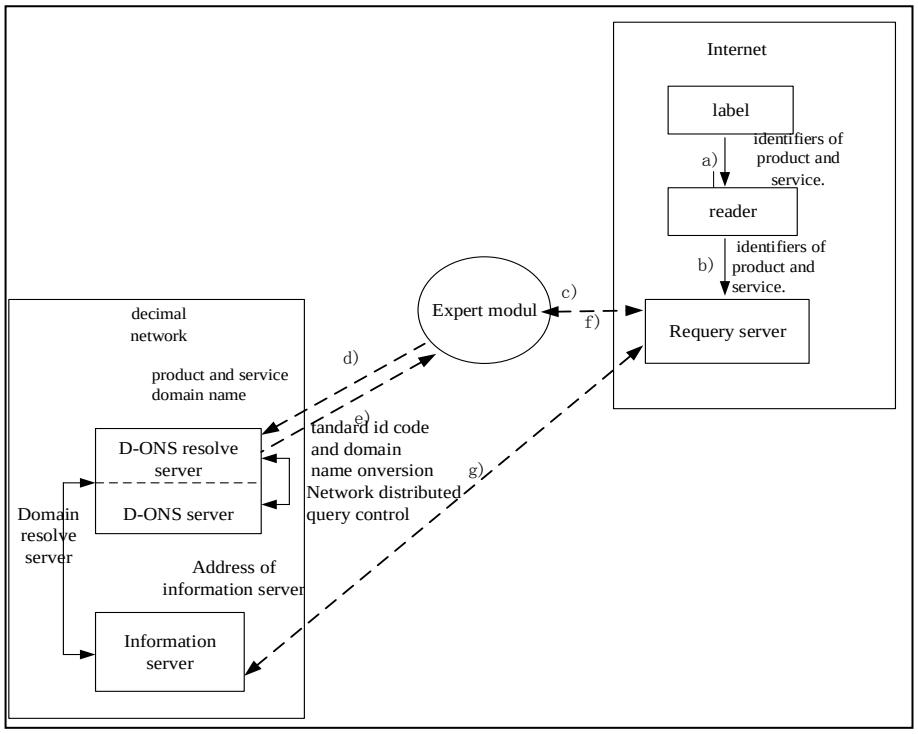


Figure 4. The process of Internet accessing to a decimal network based on the d-ons protocol.

2) The access process can be described as follows

a) Using Rf readers to read product and service identifiers from electronic tags encoded in IPV4 and IPV6 formats.

b) Submitting the identifier and identifiers of product and service to the query server.

c) Calling by the query server the expert module's decimal network interface between the Internet network and the decimal network.

d) The decimal network interface of the expert module sends a request to the D-ONS server of the decimal network architecture to inquire the information of the server domain name where the product information is stored through the identifier and product and service.

e) The D-ONS server receives the request and returns information of the product and service domain name.

f) The information server of decimal network queries the Internet server for product information.

g) The query server returns product information.

B. Directly routing the decimal network information query service and the Internet information query service system exchange query process

The process of mutual visits between the system of decimal network information query service with direct routing and the Internet information query service system are shown in figure 5 and figure 6. It involve the query server, expert module, middleware, information server, SNS server and PSNS server.

The interconnection of IPV4, IPV6 and IPV9 can realize the mutual visits between IPV4, IPV6 and IPV9 through protocol conversion. Specifically, a protocol conversion server needs to be set up, and all data packets are converted into specified protocols to satisfy the data communication between different protocols.

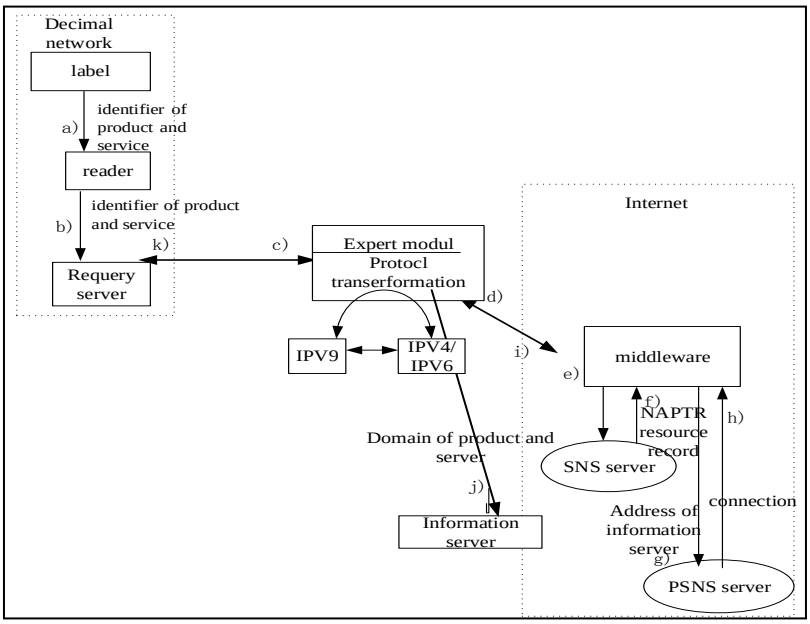


Figure 5. The accessing process of Direct routing of decimal network to the Internet

1) The process of searching the product information stored in the Internet and encoding format used IPV9 can be described as follows.

RF reader reads IPV9 identifier and the identifier of product and service from electronic tag.

Submitting identifiers of product and service to the query server on a decimal network.

a) The query server calls the expert module's Internet interface.

b) The Internet interface of the expert module accesses the middleware of the Internet architecture through identifiers of product and service.

c) Internet middleware delivers product and service domain names to the SNS server.

d) SNS server returns the middle results to the middleware .

e) The middleware returns the results to the expert module.

f) The expert module requests product information from the information server according to the address information inquired.

g) The expert module returns the product information to the query server of decimal network to complete the process of product information query.

In the direct routing mode, if the product's RF tags are encoded in IPV4 and IPV6 protocols, and the product-related information is stored in the server of the decimal network, the query process needs to involve IPV9 router, information server, expert module and query server.

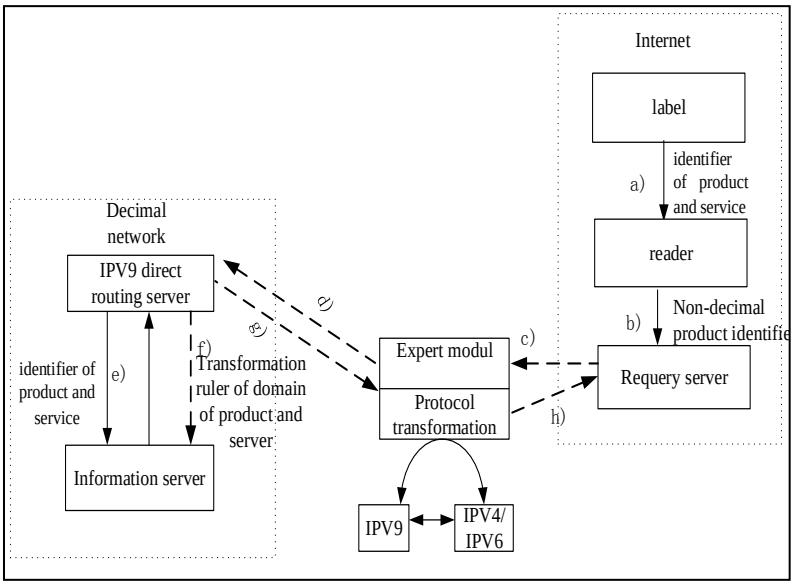


Figure 6. The access process from Internet to decimal network under the direct routing

2) The access process can be described as follows.

a) RF readers read IPV4 or IPV6 identifiers and identifiers of product and service from electronic tags.

b) Submitting identifiers of product and service to the query server over the Internet network.

c) The query server calls the expert module's decimal network interface.

d) The expert module of decimal network interface accesses the IPV9 router for the decimal network architecture through identifiers of product and service .

e) The PV9 router converts the product and service digital identifiers to the IP address of the product information server.

f) The IPV9 router accesses the information server and requests product information.

g) The IPV9 router returns product information to the Internet user via the expert module.

In the above process, the expert module between the two network systems realizes the translation and conversion of the data formats of the two systems. The protocol conversion module can translate the IP address, message and header.

VI. CONCLUSION

This paper proposes two kinds of information query exchange services between decimal network and Internet to solve the problem that the encoding format of radio frequency tag is not uniform with the network protocol format of product information server. Experimental results show that both methods can provide efficient rf information query service.

REFERENCES

[1] SJ/T CCCCC-CCCC Digital identity format specification for information processing products and services.

[2] SJ/T DDDDD-DDDD Rf-based domain name specification for products and services

[3] IETF RFC 1034 Domain names-concepts and facilities

[4] IETF RFC 1035 Domain names-implementation and specification

[5] IETF RFC 1347 TCP and UDP with Bigger Addresses

[6] IETF RFC 1561 ISO Use of ISO CLNP in TUBA Environments

[7] IETF RFC 1606 A Historical Perspective On The Usage Of IP Version 9

[8] IETF RFC 1607 A VIEW FROM THE 21ST CENTURY

[9] IETF RFC 1700 - Assigned Numbers）

A Solution to Incompatibility Between IPv6 and IPv4 Communications

(IPv10 Draft Specification)

Khaled Omar Ibrahim Omar
6th of October City, Giza, Egypt
Passport ID no.: A19954283, Phone: +2 01003620284
E-mail: eng.khaled.omar@hotmail.com

***Abstract*—It has been about one year since last hearing anything about the Internet Protocol v10 (IPv10) proposal while this week it's now available in draft form. While IPv6 isn't widely-adopted around the globe yet, IPv10 is already in development and helps to address some of the woes of IPv6. IPv10 is designed to allow IPv6 addresses to communicate to/from IPv4 addresses. IPv10 hopes to speed the adoption to IPv6 addressing by making it more backwards compatible with IPv4 with allowing the two Internet Protocol standards to better coexist.**

From the draft specification: "It solves the issue of allowing IPv6 only hosts to communicate to IPv4 only hosts and vice versa in a simple and very efficient way, especially when the communication is done using both direct IP addresses and when using hostnames between IPv10 hosts, as there is no need for protocol translations or getting the DNS involved in the communication process more than its normal address resolution function. IPv10 allows hosts from two IP versions (IPv4 and IPv6) to be able to communicate, and this can be accomplished by having an IPv10 packet containing a mixture of IPv4 and IPv6 addresses in the same IP packet header."

Keywords-IPv4; IPv6; IPv10

IPv10 is the proposed name as IPv10 = IPv4 + IPv6. The draft specification can be found at IETF.org and will expire in March 2018.

I. SPECIFICATION

Internet Protocol version 10 (IPv10)

Draft-omar-IPv10-06 Khaled Omar

Internet-Draft The Road

Intended status: Standards Track

Expires: March 2, 2018 September 2, 2017

A. Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsolete by other documents at any time.

It is in appropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on March 2, 2018.

DOI: 10.21307/ijanmc-2019-075

B. Copyright Notice

This document specifies version 10 of the Internet Protocol (IPv10), sometimes referred to as IP Mixture (IPmix).

IP version 10 (IPv10) is a new version of the Internet Protocol, designed to allow IP version 6 [RFC-2460] to communicate to IP version 4 (IPv4) [RFC-791] and vice versa. Internet is the global wide network used for communication between hosts connected to it. These connected hosts (PCs, servers, routers, mobile devices, etc.) must have a global unique address to be able to communicate through the Internet and these unique addresses are defined in the Internet Protocol (IP).

II. THE FIRST VERSION OF THE INTERNET PROTOCOL OF IPV4.

When IPv4 was developed in 1975, it was not expected that the number of connected hosts to the Internet reach a very huge number of hosts, more than the IPv4 address space; also it was aimed to be used for experimental purposes in the beginning. IPv4 is (32-bits) address that allowing approximately 4.3 billion unique IP addresses.

A few years ago, with the massive increase of connected hosts to the Internet, IPv4 addresses started to run out. Three short-term solutions (CIDR, Private addressing, and NAT) were introduced in the mid-1990s but even with using these solutions, the IPv4 address space ran out in February, 2011 as announced by IANA, The announcement of depletion of the IPv4 address space by the RIRs is as follows:

* April, 2011: APNIC announcement.
* September, 2012: RIPE NCC announcement.
* June, 2014: LACNIC announcement.
* September, 2015: ARIN announcement.

A long term solution (IPv6) was introduced to increase the address space used by the Internet Protocol and this was defined in the Internet Protocol version 6 (IPv6).IPv6 was developed in 1998 by the Internet Engineering Task Force (IETF).

III. IPV6

IPv6 is (128-bits) address and can support a huge number of unique IP addresses that is approximately equals to 2^128 unique addresses. So, the need for IPv6 became a vital issue to be able to support the massive increase of connected hosts to the Internet after the IPv4 address space exhaustion. The migration from IPv4 to IPv6 became a necessary thing, but unfortunately, it would take decades for this full migration to be accomplished.

19 years have passed since IPv6 was developed, but no full migration happened till now and this would cause the Internet to be divided into two parts, as IPv4 still dominating on the Internet traffic (85% as measured by Google in April, 2017) and new Internet hosts will be assigned IPv6-only addresses and be able to communicate with 15% only of the Internet services and apps. So, the need for solutions for the IPv4 and IPv6 coexistence became an important issue in the migration process as we cannot wake up in the morning and find all IPv4 hosts are migrated to be IPv6 hosts, especially, as most enterprises have not do this migration for creating a full IPv6 implementation.

Also, the request for using IPv6 addresses in addition to the existing IPv4 addresses (IPv4/IPv6 Dual Stacks) in all enterprise networks have not achieve a large implementation that can make IPv6 the most dominated IP in the Internet as many people believe that they will not have benefits from just having a larger IP address bits and IPv4 satisfies their needs, also, not all enterprises devices support IPv6 and also many people are afraid of the service outage that can be caused due to this migration.

The recent solutions for IPv4 and IPv6 coexistence are:

Native dual stack (IPv4 and IPv6); Dual-stack Lite; NAT64; 464xlat; MAP. (other technologies also exist, like lw6over4; they may have more specific use cases). IPv4/IPv6 Dual Stack, allows both IPv4 and IPv6 to coexist by using both IPv4 and IPv6 addresses for all hosts at the same time, but this solution does not allows IPv4 hosts to communicate to IPv6 hosts and vice versa. Also, after the depletion of the IPv4 address space, new Internet hosts will not be able to use IPv4/IPv6

A. Dual Stacks.

Tunneling, allows IPv6 hosts to communicate to each other through an IPv4 network, but still does not allows IPv4 hosts to communicate to IPv6 hosts and vice versa.

B. NAT-PT,

It allows IPv6 hosts to communicate to IPv4 hosts with only using hostnames and getting DNS involved in the communication process, but this solution was inefficient because it does not allows communication using direct IP addresses, also the need for so much protocol translations of the source and destination IP addresses made the solution complex and not applicable that's why it was moved to the Historic status in the RFC 2766. Also, NAT64 requires so much protocol translations and statically configured bindings, and also getting a DNS64 involved in the communication process.

IV. INTERNET PROTOCOL VERSION 10 (IPV10)

IPv10 is the solution presented in this Internet draft. It solves the issue of allowing IPv6 only hosts to communicate to IPv4 only hosts and vice versa in a simple and very efficient way, especially when the communication is done using both direct IP addresses and when using hostnames between IPv10 hosts, as there is no need for protocol translations or getting the DNS involved in the communication process more than its normal address resolution function.

IPv10 allows hosts from two IP versions (IPv4 and IPv6) to be able to communicate, and this can be accomplished by having an IPv10 packet containing a mixture of IPv4 and IPv6 addresses in the same IP packet header. From here the name of IPv10 arises, as the IP packet can contain (IPv6 + IPv4 /IPv4 + IPv6) addresses in the same layer 3 packet header.

V. THE FOUR TYPES OF COMMUNICATION

A. IPv10: IPv6 Host to IPv4 Host.

1) IPv10 Packet

```
        |          128-bit          |          128-bit          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Data|    Source IPv6 Address     | 0000..0   ASN
MAC   Destination IPv4 Address |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Sending IPv10 host TCP/IP Configuration:

IP Address: IPv6 Address

Prefix Length: /length

Default Gateway: IPv6 Address (Optional)

DNS Addresses: IPv6/IPv4 Address

2) Example of IPv10 Operation:

R1 & R2 have both IPv4/IPv6 routing enabled are as figure1.

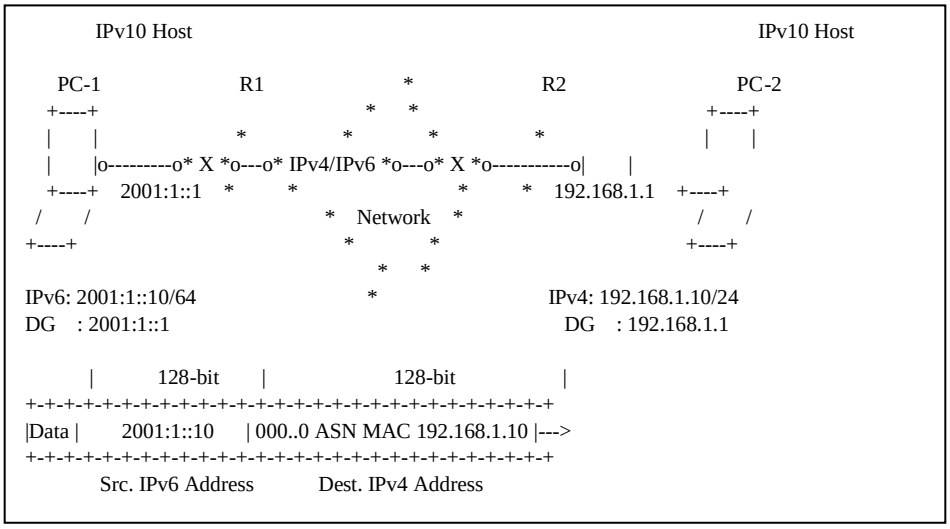


Figure 1. IPv6 host to IPv4 host

B. IPv10: IPv4 Host to IPv6 Host.

1) IPv10 Packet

| 128-bit | 128-bit |

+-+

| Data| 000..0 ASN MAC Source IPv4 Address | Destination IPv6 Address |

+-+

2) Sending IPv10 host TCP/IP Configuration:

IP Address: IPv4 Address

Subnet Mask: /mask

Default Gateway: IPv4 Address

DNS Addresses: IPv4/IPv6 Address

3) Example of IPv10 Operation

R1 & R2 have both IPv4/IPv6 routing enabled are as figure2.

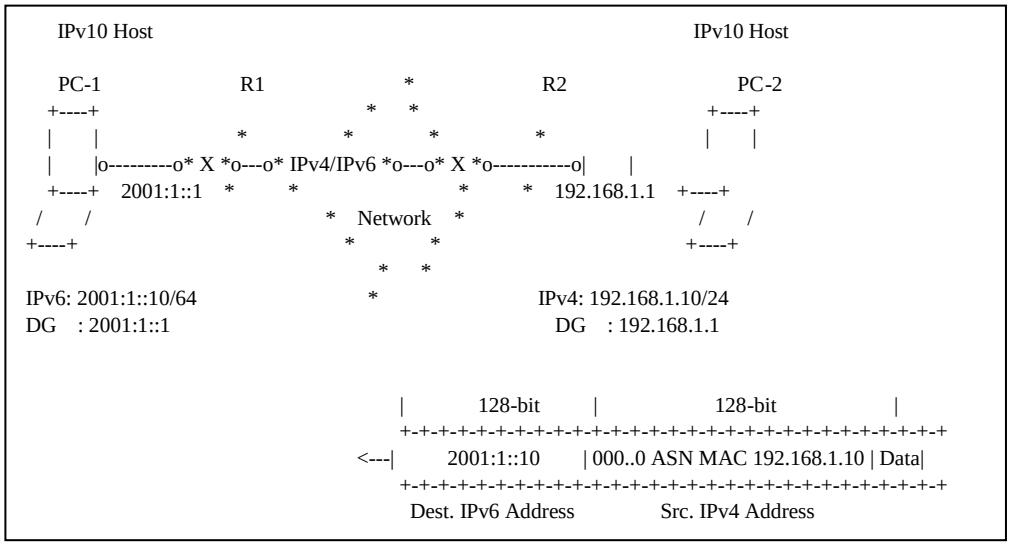


Figure 2. IPv4 host to IPv6 host

C. IPv6 Host to IPv6 Host.

1) IPv10 Packet

| 128-bit | 128-bit |
+-+

| Data| Source IPv6 Address | Destination IPv6 Address |
+-+

Sending IPv10 host TCP/IP Configuration:

IP Address: IPv6 Address

Prefix Length: /Length

Default Gateway: IPv6 Address (Optional)

DNS Addresses: IPv6/IPv4 Address

2) Example of IPv10 Operation

R1 & R2 have both IPv4/IPv6 routing enabled are as figure3.

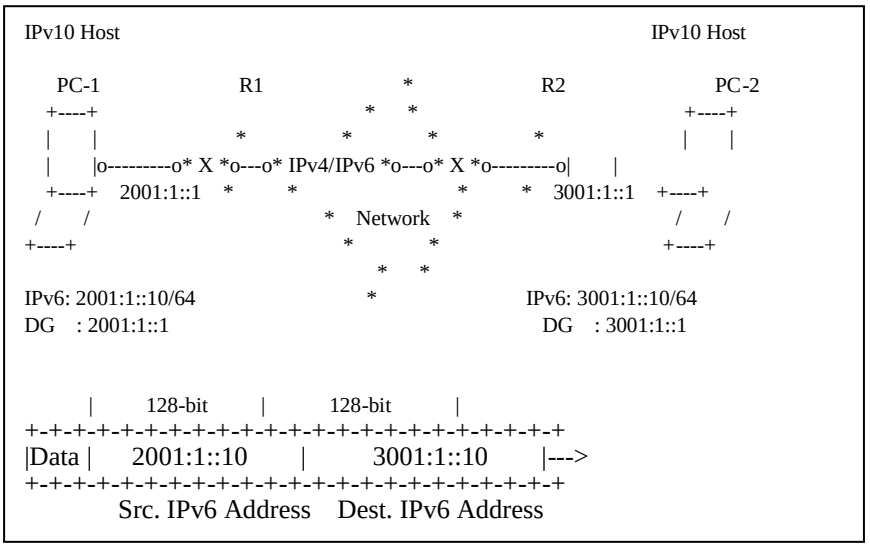


Figure 3. IPv6 host to IPv6 host

D. IPv4 Host to IPv4 Host.

1) IPv10 Packet

| \ 128-bit | 128-bit |
+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

| Data| 000..0 ASN MAC Source IPv4 Address | 000..0 ASN MAC Destination IPv4 Address |
+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+

- Sending IPv10 host TCP/IP Configuration:

IP Address: IPv4 Address

Subnet Mask: /Mask

Default Gateway: IPv4 Address

DNS Addresses: IPv6/IPv4 Address

2) Example of IPv10 Operation:

R1 & R2 have both IPv4/IPv6 routing enabled are as figure 4.

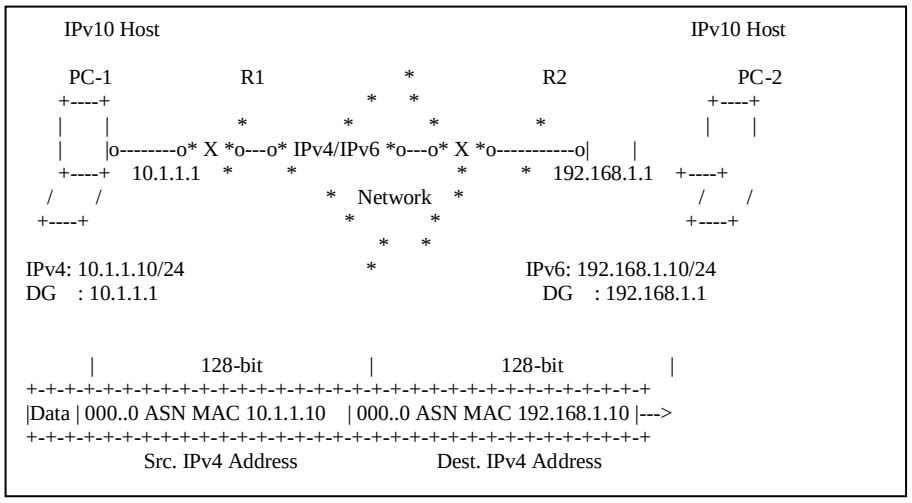


Figure 4. IPv4 host to IPv4 host

Important Notes: IPv4 and IPv6 routing must be enabled on all routers, so when a router receives an IPv10 packet, it should use the appropriate routing table based on the destination address within the IPv10 packet.

That means, if the received IPv10 packet contains an IPv4 address in the destination address field, the router should use the IPv4 routing table to make a routing decision, and if the received IPv10 packet contains an IPv6 address in the destination address field, the router should use the IPv6 routing table to make a routing decision.

All Internet connected hosts must be IPv10 hosts to be able to communicate regardless the used IP version, and the IPv10 deployment process can be accomplished by ALL technology companies developing OSs for hosts networking and security devices.

VI. IPv10 Packet Header Format

The following figure 5 shows the IPv10 packet header which is almost the same as the IPv6 packet header.

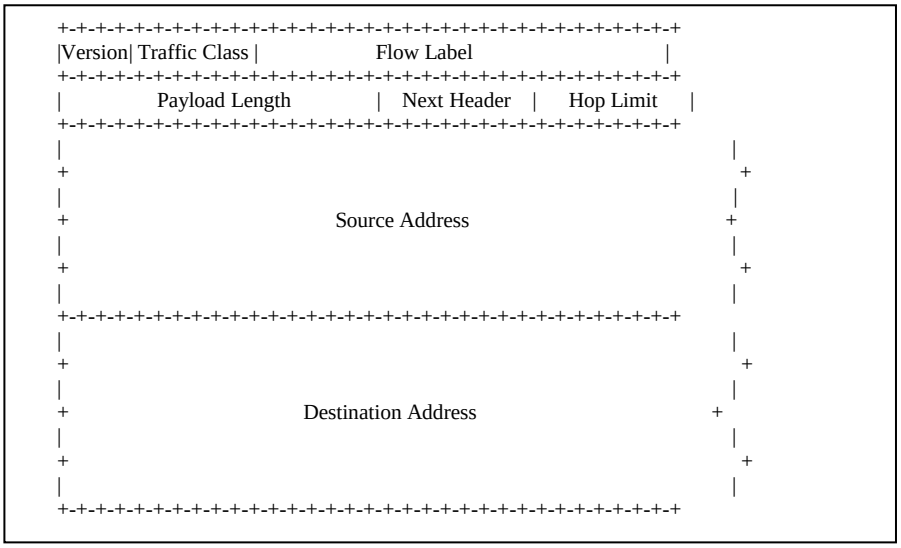


Figure 5. IPv10 packet header

1) Version

4-bit Internet Protocol version number.

- 0100 : IPv4 Packet

(Src. and dest. are IPv4).

- 0110 : IPv6 Packet

(Src. and dest. are IPv6).

- 1010 : IPv10 Packet

(Src. and dest. are IPv4/IPv6).

2) Traffic Class

8-bit traffic class field.

3) Flow Label

20-bit flow label.

4) Payload Length

16-bit unsigned integer. Length of the payload, i.e., the rest of the packet following this IP header, in octets. (Note that any extension headers [section 4] present are considered part of the payload, i.e., included in the length count.)

5) Next Header

8-bit selector. Identifies the type of header immediately following the IP header.

6) Hop Limit

8-bit unsigned integer. Decremented by 1 by each node that forwards the packet. The packet is discarded if Hop Limit is decremented to zero.

7) Source Address

128-bit address of the originator of the packet.

```
|      32-bit     |  16-bit |  48-bit |      32-bit       |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    IPv6 Address       |    OR     | 00000.....0  |
ASN   |    MAC    | IPv4 Address  |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       128-bit         |                |
128-bit          |
```

8) Destination Address

128-bit address of the intended recipient of the packet (possibly not the ultimate recipient, if a Routing header is present).

```
|      32-bit     |  16-bit |  48-bit |      32-bit       |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|    IPv6 Address       |    OR     | 00000.....0  |
ASN   |    MAC    | IPv4 Address  |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|       128-bit         |                |       128-bit      |
```

VII. Advantages of IPv10

- Introduces an efficient way of communication between IPv6 hosts and IPv4 hosts.
- Allows IPv4 only hosts to exist and communicate with IPv6 only hosts even after the depletion of the IPv4 address space.
- Adds flexibility when making a query sent to the DNS for hostname resolution as IPv4 and IPv6 hosts can communicate with IPv4 or IPv6 DNS servers and the DNS can reply with any record it has (either an IPv6 record Host AAAA record or an IPv4 record Host A record).
- There is no need to think about migration as both IPv4 and IPv6 hosts can coexist and communicate to each other which will allow the usage of the address space of both IPv4 and IPv6 making the available number of connected hosts be bigger.
- IPv10 support on "all" Internet connected hosts can be deployed in a very short time by technology companies developing OSs (for hosts and networking devices, and there will be no dependence on enterprise users and it is just a software development process in the NIC

cards of all hosts to allow encapsulating both IPv4 and IPv6 in the same IP packet header.

- Offers the four types of communication between hosts:

  - IPv6 hosts to IPv4 hosts (6 to 4).

  - IPv4 hosts to IPv6 hosts (4 to 6).

  - IPv6 hosts to IPv6 hosts (6 to 6).

  - IPv4 hosts to IPv4 hosts (4 to 4).

VIII. SECURITY CONSIDERATIONS

The security features of IPv10 are described in the Security Architecture for the Internet Protocol [RFC-2401].

IANA Considerations: IANA must reserve version number 10 for the 4-bit Version Field in the Layer 3 packet header for the IPv10 packet.

IX. FULL COPYRIGHT STATEMENT

ACKNOWLEDGMENTS

The author would like to thank S. Krishnan, W. Haddad, C. Huitema, T. Manderson, JC. Zuniga, A. Sullivan, , K. Thomann, M. Abrahamsson, S. Bortzmeyer, J. Linkova, T. Herbert and Lee H. for the useful inputs and discussions about IPv10.

REFERENCES

[1] [RFC-2401] Stephen E. Deering and Robert M. Hinden, "IPv6 Specification", RFC 2460, December 1998.